Guiding the choice of analytic approach for economic evaluations of oncology treatments

Rachid Rafia

BSc, MSc.

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield,

Faculty of medicine, Dentistry and Health,

School of Health and Related Research,

Health Economics and Decision Science.

March 2021

AKNOWLEDGMENTS

I would like to thank a number of individuals for their help and support during these long and tumultuous three and a half years. Firstly, considerable thanks go to my supervisors Professor Paul Tappenden, Professor Mark Strong, Dr Nick Latimer and Dr Jean Hamilton for their ongoing advice, support, incredible patience and steering throughout the course of this study. Without their support and flexibility, I would not have been able to complete my study.

I would also like to thank Sue Harnan, Alice Bessey, Dr Aline Navega Biz, Tushar Srivastava, Professor Ben Van Hout and Dr Roberta Ara for offering their help in running my simulation study, general support throughout my study, but also for the inconvenience that I caused them. I would also like to thank Tushar Srivastava for reading the whole thesis and for providing useful comments.

I would also like to thank Professor Alan Brennan and Dr Kate Ren for commenting on my confirmation review document and my other colleagues and friends for helping me to get through it.

I would also like to thank all the experts that I contacted when conducting my review of methods.

This thesis is based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this thesis.

FUNDING SOURCES

This PhD thesis has been prepared as part of a Doctoral Research Fellowship funded by the National Institute for Health Research (NIHR) (reference number 2016-09-092). The views and opinions expressed therein are those of the author and do not necessarily reflect those of the NIHR.

DECLARATION

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (*www.sheffield.ac.uk/ssid/unfair-means*). This work has not been previously presented for an award at this, or any other, university.

ABSTRACT

Different modelling approaches are used to address the same decision problem but can lead to different estimates of life years gained and quality-adjusted life years. Three common methods are used in health economics: the partitioned (PSM), the state-transition (STM) and more recently the multi-state model (MSM). Novel methods were also identified to jointly model progression and survival using a copula to jointly model survival outcomes and MSMs with transitions estimated simultaneously. Differences in model predictions may have the propensity to change the conclusions of an economic analysis and the decisions made on the basis of such analyses.

A simulation study was conducted to identify whether one approach is consistently superior to others under particular circumstances, or in general. The simulation study suggests that no single method is satisfactory in all circumstances and that approaches cannot be selected based on observed data characteristics alone. Case studies using real trial data also indicated that different assumptions could be made when modelling treatment effects, that PSMs and STMs may be inaccurate to varying degrees when estimating incremental outcomes and that neither is bias-free.

This thesis demonstrated that it is not possible to determine with certainty a priori which approach to select, based only on the observed characteristics of the available data; thus, analysts and decision-makers need be careful when relying on predictions from a single approach. Recommendations are formulated to improve the transparency of health economic analyses and increase decision-makers' confidence in the use of those models. Because it is unknown whether ICERs generated using a single analytic approach are adequate, in some cases, decision-making should consider ICERs from a range of alternative approaches to account for structural uncertainty. This thesis also highlights the importance of clinical input in selecting the most appropriate approach for the extrapolation of survival data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AFT	Accelerated failure time
AG	Assessment Group
AIC	Akaike Information Criterion
AUC	Area under the curve
BC	Breast cancer
BIC	Bayesian Information Criterion
CBA	Cost-benefit analysis
CC	Colorectal cancer
CCA	Cost-consequences analysis
CDF	Cumulative density function
CEA	Cost-effectiveness analysis
CUA	Cost-utility analysis
CI	Confidence interval
CLL	Chronic lymphocytic leukaemia
CR	Complete response
DFS	Disease-free survival
DSU	Decision Support Unit
EFS	Event-free survival
EMA	European Medicines Agency
ERG	Evidence Review Group
FDA	Food and Drug Administration
FL	Follicular lymphoma
HE	Health economics
HR	Hazard ratio
HRQoL	Health-related quality of life
HST	Highly Specialised Technology
HTA	Health Technology Assessment
ICER	Incremental Cost-Effectiveness Ratio
IPCW	Inverse Probability of Censoring Weighting
KM	Kaplan-Meier
MC	Monte Carlo
MCL	Mantle cell lymphoma
MCMC	Markov Chain Monte Carlo methods
MF	Myelofibrosis
MLE	Maximum likelihood estimation
MM	Multiple myeloma
MSE	Mean Squared Error
MSM	Multi-state model
MTA	Multiple Technology Appraisal
NICE	National Institute for Health and Care Excellence
NPMLE	Non-parametric maximum likelihood estimation
NSCL	Non-small cell lung cancer
ONS	Office for National Statistics
OC	Ovarian cancer
OS	Overall survival
PSM	Partitioned survival model
PC	Prostate cancer
PD	Progressive disease
PDF	Probability density function

PF	Progression-free
PFS	Progression-free survival
PH	Proportional hazard
PICO	Population, intervention, comparator, outcome
PPS	Post-progression survival
PR	Partial response
PrePS	Pre-progression survival
PSA	Probabilitic sensitivity analysis
PSM	Partitioned survival model
Q-TWIST	Quality-adjusted time without symptoms or toxicity
QALY	Quality-adjusted life years
RCC	Renal cell carcinoma
RCT	Randomised controlled trial
RFS	Recurrence-free survival
RMSE	Root mean squared error
SE	Standard error
SMR	Standardised mortality ratio
STA	Single Technology Appraisal
STM	State-transition model
TA	Technology Appraisal
TTD	Time to treatment discontinuation
TTP	Time to progression
UK	United Kingdom

1 CHAPTER I: IMPORTANCE OF CONSISTENT DECISION-MAKING

The aim of this thesis is to guide the choice of modelling approach for estimating health state sojourn time for anticancer therapies conditional on the nature of data available and to guide decision-making based on these models. In particular, this study addresses the following question – *"Is it possible to identify when particular analytical approaches may perform better than others, subject to the nature of the data available?"* In this chapter, I explain the motivations underpinning the thesis.

1.1 Motivation for this thesis

In the UK, drug reimbursement decisions take into account multiple factors, including the safety, clinical effectiveness and cost-effectiveness of the technology under assessment (whether the drug represents good value for money for the health service).[1, 2]

The determination of whether a particular health technology is cost-effective is informed by an economic evaluation. Broadly speaking, economic evaluation can be defined as the comparative analysis of two or more competing options in terms of their costs and consequences.[3] Economic evaluation may take numerous forms (e.g. cost-effectiveness analysis [CEA], cost-utility analysis [CUA], cost-benefit analysis [CBA] and cost-consequences analysis [CCA]). In England, CUA is frequently used for the economic analysis of health care technologies. Whilst it is sometimes possible to undertake an economic evaluation based on a single study (e.g. a randomised controlled trial [RCT]), more typically, mathematical models are required to predict long-term outcomes and costs for all relevant decision options.

For drugs, a decision rule based on the incremental cost-effectiveness ratio (ICER) is employed to help inform resource allocation decision-making. The ICER represents the ratio of the incremental costs to incremental health benefits for the new technology versus current practice and provides a basis for decision-makers to consider the balance of the additional value of a health technology against the opportunity costs associated with curtailing existing treatments and services to fund the new technology. In England, the National Institute for Health and Care Excellence (NICE) uses quality-adjusted life years (QALYs) gained as the common measure for health benefits; this is a measure of both the quantity and quality of life experienced by patients. In principle, this allows consistent reimbursement recommendations to be made across all disease areas, thereby promoting the efficient allocation of healthcare resources. For the majority of drug technologies, NICE typically adopts a decision-making threshold range of £20,000 to £30,000 per QALY gained, although technologies with higher ICERs may be recommended under particular circumstances, for example, where the intervention satisfies NICE's End of Life criteria or where it meets the criteria for being considered as a Highly Specialised Technology (HST).[2, 4]

In order to allow a 'fair' comparison between treatments, it is essential that economic evaluations adopt a robust and consistent approach. When seeking market access/reimbursement in the UK, pharmaceutical companies and sponsors are required to submit an economic evaluation within the NICE Single Technology Appraisal (STA) process. To achieve some degree of standardisation, NICE has published a methods guide to support its Technology Appraisal (TA) process.[2] However, this guidance is not a technical document and covers only NICE's broad principles for undertaking Health Technology Assessment (HTA).

Owing to ongoing and rapid drug development, often with the same product having multiple licensed indications, oncology represents an area in which a large number of technologies require appraisal before they can be routinely commissioned within the NHS. Overall survival (OS) is a key endpoint in oncology in the advanced/metastatic setting, defined in a trial setting as the time from randomisation (or study entry for non-randomised studies) to death from any cause. OS is an objective endpoint and is generally straightforward to measure and record. However, long follow-up durations are typically required in order for a new technology to demonstrate a survival advantage; such evidence is frequently lacking. Progression-free survival (PFS), which is usually defined as the time from randomisation to progression or death (whichever occurs first) has therefore been suggested as a potential surrogate for OS in advanced cancer; this usually requires much shorter follow-up than OS. In addition, the United States Food and Drug Administration (FDA)[5] and the European Medicines Agency (EMA),[6] recently accepted that prolonged PFS and disease-free survival (DFS) could be considered relevant measures of clinical benefit, and therefore could be used as primary outcome, provided that the magnitude of the treatment effect outweighs safety concerns. In this case, OS should be reported as a secondary outcome. The use of PFS as a primary endpoint (or surrogate for OS) is not solely attributable to its practicality. It should be noted that although PFS is sometimes regarded as a valid surrogate for OS, a review conducted by the decision support unit (DSU)[7] suggests that the level of evidence available supporting a relationship between PFS and OS varies considerably by cancer type and is not always consistent, even within one specific cancer type. Consequently, instances in which OS data are immature due to not being the primary endpoint, or are contaminated by treatment switching and the use of subsequent therapies, are becoming increasingly common.[8] Trials may also be terminated earlier when a significant difference in PFS is achieved.

Health economic models of anticancer therapies in the advanced/metastatic setting typically share the same structure, articulated around three health states; (i) progression-free (PF); (ii) progressed disease (PD), and; (iii) death.[9] Notwithstanding other health effects which might be included in the model (e.g. health losses due to adverse events), the time spent in each of the alive health states (sojourn time) is usually weighted according to the average level of health-related quality of life (HRQoL) associated with each state to estimate overall QALYs gains.

Recent methodological research around modelling for anticancer therapies has largely focused on survival extrapolation, adjusting OS to account for treatment switching and whole disease modelling.[10-15] Whilst this represents an important step, there remain research gaps with respect to selecting the appropriate analytic approach to estimate health state sojourn time, and by extension, to estimate QALYs.

A number of approaches, described in Chapter 2, are available and are commonly used in health economics when estimating the health states sojourn time for anticancer treatments; these include the general partitioned survival approach and the general state-transition approach (further detail is available in Chapter 2). It should be noted that for each method, variations exist and a number of other choices need to be made. These general approaches require different data inputs and involve the use of different assumptions.

Each approach is subject to inherent limitations and their appropriateness is likely to vary subject to the amount and nature of data available, as well as other factors related to the decision problem such as whether there is a need to deviate from the trial (inclusion of a stopping rule for instance), or whether the downstream treatment pathway should be explicitly modelled.

Whilst taxonomies have been developed to help analysts determine the most appropriate modelling technique (e.g. Brennan *et al*[16], Barton et al[17]), these are generally broad and do not focus on issues which are specific to the modelling of anticancer therapies. In particular, the PSM approach is now widely used in oncology but does not fit neatly onto these taxonomies that assume that IPD are available and transitions can be estimated. These taxonomies should be updated to reflect modelling approaches which estimate health state occupancy directly without modelling the underlying disease process.

At present, the choice of analytical approach to estimate the health state sojourn time for anticancer therapies in the advanced/metastatic setting is largely based on what the modeller considers to be appropriate for the decision problem at hand and/or references to models developed to inform previous appraisals in the given disease area.[9] The availability for individual patient level data also drives the choice of approach. There is no explicit framework to help the modeller decide whether one particular approach may be more appropriate in particular cases, for instance when data are less mature or when the available data indicate some degree of dependence between outcomes (e.g. if longer time of progression appears to be associated with longer time to death). The choice of model can influence the conclusion of an analysis.

As a consequence, despite the use of the generally accepted model structure for anticancer therapies, different modelling approaches are used inconsistently to estimate health state sojourn times and resulting QALY gains within NICE appraisals of cancer technologies. In addition to factors related to

the decision problem (inclusion of a stopping rule, modelling the pathway...) that influence the preferred method, other factors could inform the choice of analytic approach. A recent review conducted by Woods *et al* (2017)[9, 18] found that the choice of analytic approach was rarely justified with respect to the data, but was instead commonly justified through reference to previous appraisals as precedent. The review also highlighted some general inconsistencies in analysts' understanding of the approaches.

The identification of new approaches and the formal exploration of when one particular approach may fare better than others is therefore valuable and may lead to more consistent decisions.

1.2 Examples of health technology appraisals where the choice of approaches had a considerable impact on decision-making.

In this section, I describe briefly three examples of NICE appraisals in which the company's choice of analytic approach was contested by the Assessment Group (AG)/Evidence Review Group (ERG). The key approaches employed and described in this section are the partitioned survival model (PSM) and the state-transition model (STM). These are described in more detail in Chapter 2. The stated rationale provided by the AG/ERG for preferring an alternative approach is described. The validity of some of these arguments are discussed in Chapter 2.

- **Example 1: NICE TA257 - Assessment of the cost effectiveness of lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone receptor positive breast cancer (BC) that over-expresses HER2**

In NICE TA257 in BC,[19] the two pharmaceutical companies each submitted an economic evaluation (using a PSM structure), which used OS as the primary source of survival information and extrapolated OS from the trials using standard parametric curves from the Generalised F family of survival functions. The AG involved in this appraisal used an alternative modelling approach (the STM), calculating the expected OS from the time in the PF and PD states, resulting in a large variation in projected OS, compared with the projection from the two companies. The Kaplan-Meier (KMs) plots for PFS and OS and the numbers of PFS and OS events were not reported in the documentation for this appraisal as they were marked as confidential. The AG argued that their decision to model OS by combining PFS and PPS using a STM structure was justified because (i) OS is a result of the combination of patient experience in two distinct phases (the risk of death is lower in patients who have not progressed and greater in patients who have progressed) and therefore it was believed by the AG that standard parametric statistical models cannot accurately represent an outcome measure such as OS; (ii)

modelling OS over a long time horizon can result in large cumulative errors; (iii) post-progression survival (PPS) is more stable and there is greater confidence when extrapolating over a long time horizon, with narrower confidence intervals, and (iv) modelling PFS and OS independently could lead to some anomalies with negative estimates of PPS when both outcomes are projected independently from each other. It should be noted that these arguments reflect the view of the AG involved in this appraisal, some of which I consider to be debatable.

Although the differences in ICERs generated by the companies and the AG are not entirely attributable to the alternative modelling approaches adopted (PSM vs. STM), the AG's estimate of the ICER was approximately three times as high as that reported by the companies. Estimates of life years gained (LYGs) and QALYs gained are shown in Table 1. When looking at the LYs (discounted), one company estimated the life years to be 3.40 years for lapatinib+letrozole (LAP+LET) and 2.82 years for LET alone, leading to an incremental survival gain of 0.58 years. The AG, using an alternative structure estimated the LYs (discounted) to be 2.69 and 2.55 respectively, leading to an incremental LYs of 0.14; this is approximately one quarter the size of the gain estimated by the company. In contrast, the same company estimated the LYs to be 3.05 years for trastuzumab+anastrozole (TRA+ANA) and 2.66 years for ANA, leading to an incremental LY of 0.39; compared with 2.70 and 2.22 estimated by the AG, leading to an incremental LY of 0.48. The incremental LY estimated by the AG was therefore higher compared with those estimated by one of the company (competitor). The estimate for QALYs were also very different between the companies and the AG approach.

Table 1 : Comparison of LYs and QALYs estimated by the companies and AG in TA257

	TRA+ANA	LAP+LET	ANA	LET
LYGs				
Company A (PSM)	3.05	3.40	2.66	2.82
Company B (PSM)	NR	NR	NR	NR
AG (STM)	2.70	2.69	2.22	2.55
QALYs				
Company A (PSM)	2.14	2.39	1.79	1.92
Company B (PSM)	1.87	1.71	1.29	1.29
AG (STM)	1.69	1.57	1.24	1.46

Abbreviations: ANA: anastrozole; LAP: lapatinib; LET: letrozole; NR: not reported; TRA: trastuzumab

- **Example 2: NICE TA472 - Assessment of the cost-effectiveness of obinutuzumab with bendamustine for treating follicular lymphoma (FL) refractory to rituximab**

A more recent example is TA472 in FL.[20] The company submitted what is usually referred to as a cohort-based semi-Markov model (a type of state transition model in which event rates are conditional on the time at which patients enter an intermediate model state, such as PD). In this type of model, OS is not taken directly from the trial but is instead estimated as a function of sojourn time in the PF and PD states. The company justified this approach on the basis of the immaturity of OS in the trial and the indolent nature of the condition. The probability of dying after progression was taken from the pooled PPS data across both trial arms.

Despite the data for OS being more immature for the intervention arm, the ERG considered that directly modelling OS was more appropriate for the following reasons: (i) the evidence used to inform the PPS could equally be considered immature and subject to uncertainty (as the same number of events are observed - only the denominator changes); (ii) evidence to inform PFS for the intervention was also immature (and therefore basing predictions on two immature endpoints for the intervention could introduce inaccuracy), and (iii) discrepancies between the model's predictions and the observed data for OS (especially the for the control arm) in the company's model.

Consequently, the ERG explored the use of a PSM whereby models were fitted directly to the trial OS data, with some assumptions on when the hazard of death would be the same between treatment arms. ICERs were not reported and were marked as confidential. LYGs and QALYs reported by both the company and ERG are shown in Table 2. Whilst estimated survival was similar for the intervention arm between the ERG and the company (5.80 vs 5.73 LYGs), differences were much larger for the control arm (4.27 vs. 5.30 LYGs). The ERG further noted that when plotting the predicted vs. observed OS, the approach taken by the ERG led to similar predictions to the company for the intervention arm, and both approaches provided a reasonable fit to the observed data. In contrast, for the comparator arm, the ERG commented that the company's approach provided a poor fit to the observed OS, but that the ERG approach (PSM structure) provided a much better fit to the observed data, with the differences in survival between the comparator and intervention arms being much less pronounced. It should be noted that the comment from the ERG is unsurprising given that OS in the PSM is fitted directly to the trial data, and therefore a good visual fit to the observed data is generally expected.

In TA472, most of the differences were attributable to the choice of analytical approach rather than the approach to parametric extrapolation

Table 2 : LYs and QALYs reported by the company and the ERG in TA472

	O-benda	benda	Incremental
LYGs			
Company A (STM)	5.80	4.27	1.54
AG (PSM)	5.74	5.30	0.44
QALYs			
Company A (STM)	4.23	2.92	1.32
AG (PSM)	4.09	3.48	0.62

Abbreviations: benda: bendamustine; O: obinutuzumab

- **Example 3: NICE TA381 - Assessment of the cost-effectiveness of olaparib for the maintenance treatment of relapsed, platinum-sensitive, BRCA mutation-positive ovarian, fallopian tube and peritoneal cancer after response to second-line or subsequent platinum-based chemotherapy**

A third example of different analytic approaches used between the company and the ERG is evident in TA381 (olaparib for patients with ovarian cancer[21]). The company used a semi-Markov model whereby patients move through a series of states, with OS modelled as a function of the time spent in these health states. The ERG had a number of concerns and argued that a PSM was more appropriate because of challenges in modelling the pathway, the presence of treatment switching in the trial and discrepancies between observed and model-predicted OS. The company estimated QALYs to be 2.58 for the intervention and 1.69 for the comparator arm, leading to an incremental gain of 0.90 QALYs. The ERG's most optimistic analysis, across any combination of parametric survival models, suggested an incremental gain of 0.52 QALYs. The ERG's analysis suggested that the ICER for olaparib was considerably higher than the company's estimate.

These three examples highlight that alternative models can be used to address the same decision problem, but that each of these analytic approaches are associated with limitations. These different approaches can lead to very different estimate in terms of LYGs and QALYs. It is unknown whether one approach is consistently better than the other approach under particular circumstances, if not under none or all cases, and therefore the choice is often the responsibility of the analyst. As simply shown in the three examples, the differences between the approaches have the propensity to change the conclusions of an economic analysis and the decisions made on the basis of such analyses.

In TA381, most of the differences were attributable to the choice of analytical approach rather than the approach to parametric extrapolation.

1.3 Research questions

The aim of this thesis is to guide the choice of analytic approach to estimate health state sojourn time for modelling anticancer therapies conditional on the nature of data available and to guide decision-making based on these models. In particular, this study addresses the following question – "*Is it possible to identify when particular analytical approaches may perform better than others subject to the nature of the data available (e.g. under different levels of censoring, dependence and follow-up*)?"

In order to address this question, the following sub-questions will be explored:

(i) How is health state sojourn time currently estimated in health economic models of anticancer therapies?
(ii) Are the simplifications made in health economic models of anticancer therapies appropriate?
(iii) How could the dependence structure between progression and survival outcomes be included when estimating the health state sojourn time?
(iv) How do the identified approaches perform in terms of prediction subject to the nature of the data available e.g. under different levels of censoring, dependence between the time to progression and death following progression and follow-up?
(v) Is it appropriate to rely on predictions obtained from a single analytical approach to estimate health state sojourn time?

These research questions were developed by me and refined by his supervisory team following use of the different analytical approaches, both as part of his previous role as an ERG for ScHARR-TAG and work conducted with pharmaceutical companies

The focus in this thesis is when a model is developed based on data from a RCT only, without explicit use of external evidence to replace outcomes from the trial, i.e. transitions are estimated within the trial (and not taken from external sources). It should however be noted that external evidence could be used to inform the long-term plausibility of predictions. However, this is not the focus in this thesis. In addition, the focus in this thesis is when individual-patient level data (IPD) are available and when cohort models are appropriate (e.g. interaction between individuals does not need to be modelled), and therefore both the PSM and STM could be used. In some cases, the choice of approach could be driven due to IPD not being available, or the decision problem at hand.

1.4 Thesis structure

This thesis is structured in five parts and is comprised of ten chapters. The links between chapters and how the different chapters influence each other is shown in Figure 1. Chapter 1 is the introduction to the thesis.

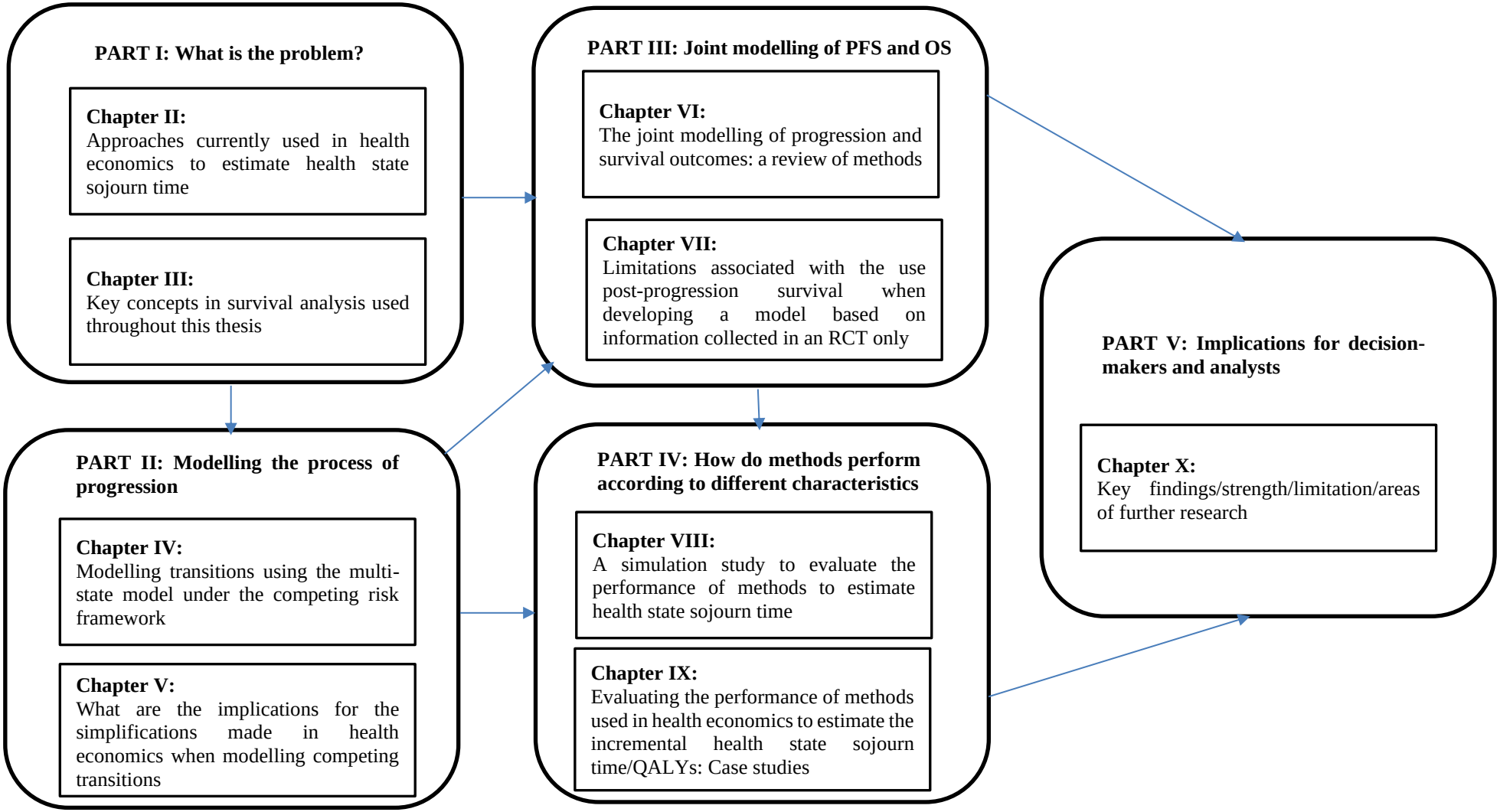
Part I is composed of two chapters. The first chapter (Chapter 2) describes the current approaches used to estimate health state sojourn time in economic models of anticancer therapies, namely the PSM and the STM approaches. In this chapter, I describe the key strengths and limitations of the approaches, and highlight whether they have been systematically compared. Chapter 3 provides a theoretical background on key concepts in survival analysis. In this chapter, I briefly describe the general concept of survival analysis, censoring, the survivor, hazard and cumulative hazard functions, the Kaplan-Meier and Nelson-Aalen estimates, the Cox regression model and the concept of competing risks. These concepts will be briefly explained through reference to a dataset in breast cancer. These two chapters set the scene and provide the general background to the key terms and concepts used throughout the thesis.

Part II of the thesis is also comprised of two chapters. In Chapter 4, I provide further detail on the estimation of health state sojourn time using multi-state models (MSMs) – a type of state transition model which combines transitions under a competing risk framework. This chapter focusses on the implementation of the MSM using two packages available in R, the `msm` package and the `mstate` package. I describe the key differences between the approaches and how transitions are combined under a competing risk framework. In Chapter 5, I highlight the simplifications made in health economic (HE) models of anticancer therapies to estimate health state sojourn time using the STM approach and demonstrate how this compares with the MSM approach. I then discuss the key assumptions and implications for the simplifications typically made in health economics when modelling competing transitions between health states. The findings from Chapters 4 and 5 impact on all subsequent chapters.

Part III concerns the joint modelling of progression and survival outcomes and consists of: (a) a systematic review of methods, conducted across a range of disciplines, to explore the available methods used to jointly model progression and survival outcomes when estimating the health state sojourn time (Chapter 6), and (b) a discussion of limitations associated with the use of PPS when developing a model based on information collected in an RCT only (Chapter 7).

Chapters 6 and 7 influence all of the subsequent chapters. In particular, the review (Chapter 6) identifies the potential approaches that could be used or combined with existing approaches which are currently used in health economics to jointly model PFS and OS. Challenges associated with searching the methodological literature are addressed through the use of iterative searching.

Figure 1 : Schematic representation of the thesis structure



Chapter 7 describes some of the limitations associated with the use of PPS when developing a model based on information collected in an RCT only. In this chapter, I use real datasets to illustrate the potential biases associated with the use of PPS data estimated only in a subset of patients who have progressed. I then discuss the merits and limitations of an approach that has been considered to adjust PPS (i.e. making the time to death following progression conditional on the time to progression) and show in real datasets whether this approach consistently improve predictions.

In Part IV, alongside common approaches used in health economics (Chapter 2-5), the range of methodologies identified in Chapter 6 is applied to a series of simulated datasets (Chapter 8) and real datasets in Chapter 9. In Chapter 8 (simulation study), hypothetical trial data are generated covering a wide range of possible scenarios relating to different data characteristics. Methods described in Chapter 2-6 are then applied to test their appropriateness and performance subject to the nature of the data available (dependence between the time to progression and time to death following progression, level of censoring and duration of follow-up). Only those methods for jointly modelling PFS and OS identified from the review in Chapter 6 which could be adopted easily in health economics (i.e. had already been programmed in a suitable package) were considered in the simulation study. In Chapter 8, selected methods are applied to simulated single trial arms only to examine their performance in estimating health state sojourn time and QALYs. This single trial arms approach was adopted in order to avoid the potential for spurious conclusions arising from apparently appropriate incremental outcomes despite the presence of a poor model fit in both treatment groups. Furthermore, when estimating life years and QALYs in the intervention group, different modelling assumptions could be made (using hazard ratio, pooling data...) as less information is available about the treatment effect, increasing challenges when interpreting results.

Because a single trial arms approach is used in Chapter 8, for completeness, in Chapter 9, an exploratory analysis is conducted whereby selected methods (the PSM and STM) are applied to two trial arms to examine their performance in estimating the incremental LYGs and QALYs between competing options, to support and confirm findings from Chapter 8. In Chapter 9, methods are applied to a series of case-studies involving real data from trials in gastric cancer.

Part V (Chapter 10) presents the conclusions of the thesis, with key recommendations and a summary of findings, a description of the limitations, strengths, and areas for further research.

PART I: WHAT IS THE PROBLEM?

2 CHAPTER II: APPROACHES CURRENTLY USED IN HEALTH ECONOMICS TO ESTIMATE HEALTH STATE SOJOURN TIME

2.1 Chapter overview

This chapter highlights the key differences between the partitioned survival model (PSM) and state-transition model (STM) approaches, as well as the strengths and limitations that are often perceived with these approaches. The PSM and STM are described in Section 2.3 and 2.4, respectively. In Section 2.5, I discuss whether these approaches have been compared systematically.

2.2 Introduction

Two broad analytic approaches are currently used in health economics (HE) to estimate health state sojourn time and QALYs when modelling anticancer treatments;

- The partitioned survival approach whereby the OS and PFS curves are used as the primary sources of time-to-event information, with the sojourn time in the PD state derived as the difference between the cumulative survivor functions for OS and PFS, and;
- The state-transition approach, whereby OS is estimated indirectly as a function of the time to progression (TTP), the time to pre-progression death (PrePS) and the time to death following progression (also referred as PPS). The term state-transition is used here to describe the general process by which patients move through a series of mutually exclusive and jointly exhaustive model health states. Different terminologies are commonly used in the literature to describe the general state-transition process, including: compartmental model; illness-death model, and progressive three state model. There are two key variations of the state-transition approach which will be the focus in this thesis:
    - **The multi-state model (MSM)** whereby the competing transitions (transitions from PF to PD or death) are explicitly modelled and combined under a competing risk framework. Therefore, the term MSM is used throughout this document to refer to the STM whereby transitions are combined under a competing risk framework,
    - **The STM whereby simplifications are made with respect to the competing transitions (referred hereafter as the Simplified STM).** In this variation of the STM, which is the most commonly used in health economic analysis (as demonstrated in Chapter 4.5), PFS, the cumulative incidence of the competing events (progression and pre-progression death) is used directly to represent the combined transitions to progression or death from the progression-free health state, with additional assumption made to separate the events.

It should be noted that whilst the focus of this thesis is around the commonly used three state advanced/metastatic cancer structure, additional health states could be included, for example to reflect multiple progression events associated with sequences of treatments. Issues around the choice of analytic approach are largely similar irrespective of the number of health states included in the model. Key characteristics, strengths and limitations of the PSM and the STM approaches in the case of the typical three state cancer model are summarised below in Sections 2.3 and Section 2.4, respectively.

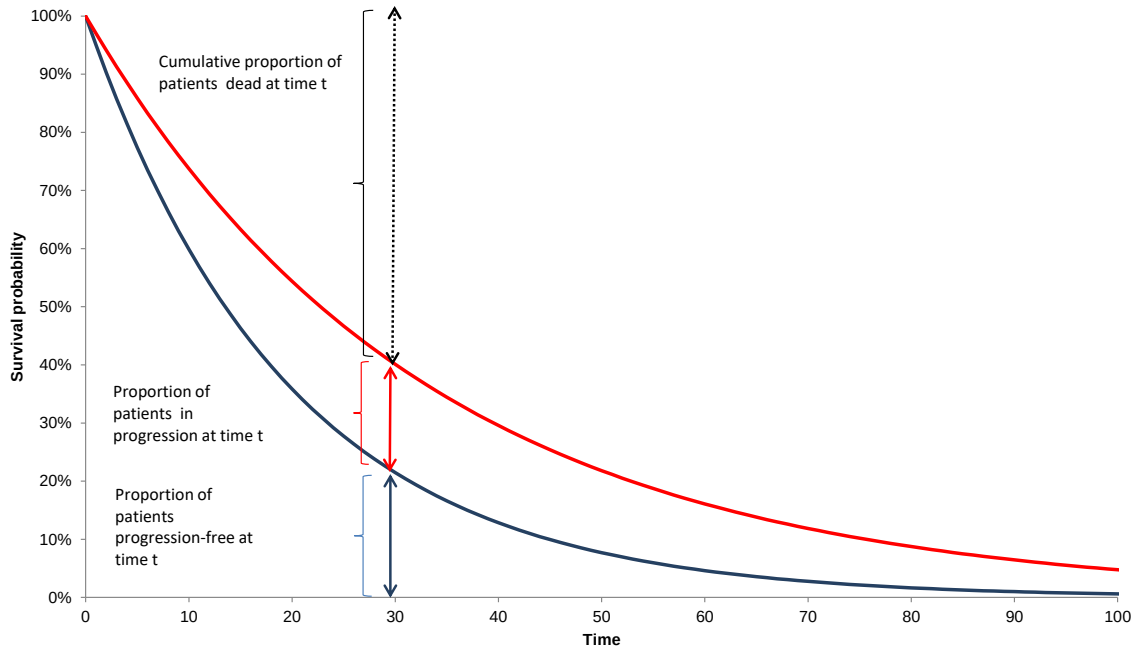
Furthermore, the focus in this thesis is when IPD is available, and therefore both the PSM and STM could be used. In some cases, the choice between approach could be driven due to IPD not being available, or the decision problem at hand.

2.3 The partitioned survival approach

As highlighted in a recent review conducted by Woods *et al* (2017), the PSM is probably the most commonly used modelling method applied in the economic analysis of advanced/metastatic anticancer treatments.[9] Under this approach, PFS and OS are modelled as two independent processes using data on these outcomes. This approach is intuitively attractive, as it relies directly on the PFS and OS data from the trial with no (or minimal) assumptions about any potential relationship between PFS and OS.

The PSM originates from the Quality-Adjusted Time Without Symptoms or Toxicity (Q-TWIST) approach,[9] which was developed to incorporate quality of life information into survival analysis. The typical implementation of PSMs involves partitioning overall survival time into time with/without progression and applying utilities to each alive health state to estimate QALYs gained. Considering the typical three-state model (PF, PD and death) used in the economic analysis of advanced/metastatic cancer therapies, the time-to-event curves for OS and PFS are used as primary sources of time-to-event information to estimate health state occupancy over time. Except in the rare cases in which survival data are not subject to censoring or Kaplan-Meier functions are complete, PFS and OS are typically extrapolated using parametric functions. In the PSM, the probability of being in the PF health state at any time *t* is given by the cumulative probability of PFS; the probability of being alive at any time *t* is given by the cumulative probability of OS, and; the probability of being in the PD state at time *t* is given by the difference between the OS and PFS functions. Ignoring discounting, the mean sojourn time in the PFS state is estimated by integrating the PFS survival function, whilst the mean sojourn time in PD is estimated as the difference between the area under the curve for the PFS and OS functions (see Figure 2).

Figure 2 : Estimation of health state occupancy in partitioned survival models



A key characteristic of the PSM approach, compared with the STM process (described later in Section 2.4), is that it deals with state occupancy directly rather than estimating transitions between health states. Therefore, within the PSM, transitions are not explicitly modelled as the proportions of individuals residing in each model health state at each time-point are estimated directly from the cumulative PFS and OS survivor functions.

Whilst the simplicity of the PSM approach is often considered attractive, it is associated with several limitations; these are described below. It should be noted that some of the arguments that have been made against the PSM approach reflect different point of views, hence some of these may contradict each other or may be debateable:

- **Limitation 1: Difficulties in representing the hazard of death using a single parametric function**. It has been argued that, in some circumstances, OS cannot be represented by a single standard parametric survivor function.[19] This is because OS is a function of the patient experience in two distinct phases in which the hazard rates would be expected to exhibit different dynamics (i.e. in PFS the patient is likely to have reduced risk of death, after progression, the risks are more likely to revert to higher levels of uncontrolled disease progression). There may also be heterogeneity between groups and outcomes may also be different between groups. The use of flexible parametric models, mixture-cure models or response-based models may help to address this problem.

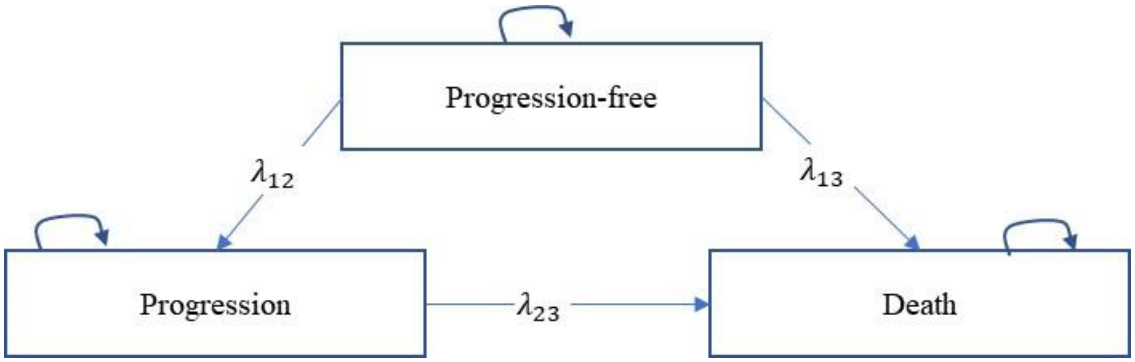
- **Limitation 2: Inability to model the underlying process.** Related to Limitation 1, the PSM does not involve modelling the underlying disease process and therefore, the selection of an appropriate survival function to extrapolate long-term outcomes may be challenging when there is little information on the expected long-term effects of the intervention.
- **Limitation 3: Difficulties in extrapolating outcomes when observed data are immature**. Related to Limitation 2, producing a reliable extrapolation is difficult when trial follow-up is limited and/or when the degree of censoring is substantial. This is particularly important given that OS data are often immature.[22] Within the PSM, extrapolation of the survivor functions is often based on short-term trends observed in the trial. These trends are then assumed to hold throughout the extrapolation period.[9] This may not be appropriate especially when extrapolating over a long time period. It has been argued that the extrapolation of OS could prove less reliable and uncertain compared with the extrapolation using intermediate endpoints (such as modelling OS using PFS and PPS). However, as described in Chapter 7, this is debateable, and there are a number of issues with the extrapolation of PPS.
- **Limitation 4: Problems associated with ignoring the dependence structure between PFS and OS**. The dependence structure between PFS and OS is typically not included in health economics whereby survival models are fitted separately to the available data on PFS and OS. However, PFS and OS events overlap with one another (PFS is a composite endpoint including progression and death occurring prior to progression) and events are structurally dependent in the sense that death cannot be followed by progression.[9] The lack of consideration of the structural dependence between PFS and OS may in some instances lead to anomalies, whereby the cumulative probability of PFS exceeds the cumulative probability of OS at certain time-points. The independent modelling of PFS and OS could also lead to potentially implausible scenarios whereby the cumulative PFS probability reaches zero quickly (short tailed) but the OS function reaches a plateau (long tailed). Ignoring the dependence structure between the outcomes leads to a disconnect between PFS and OS. Whilst there are methods available in other fields to incorporate the dependence structure between joint survival endpoints (such as modelling PFS and OS under a semi-competing risk framework), these are rarely used in health economic models.
- **Limitation 5: Lack of transparency around the underlying process.** Related to Limitation 2, the PSM deals with health state occupancy. Therefore, transitions are not explicitly modelled. Therefore, it may be more difficult to understand the underlying process of progression or what is implied about the relative treatment effects of a new therapy (compared with the STM, which involves the explicit modelling of the underlying disease process).

2.4 The state-transition model approach

Within the STM, transitions between health states are explicitly modelled and individuals move between these health states over time. This is different to the PSM, whereby state occupancy is modelled directly. In the STM, OS is estimated indirectly conditional on the transition intensities (rates). The relationship between OS and the intermediate health states is made explicit and requires quantification.

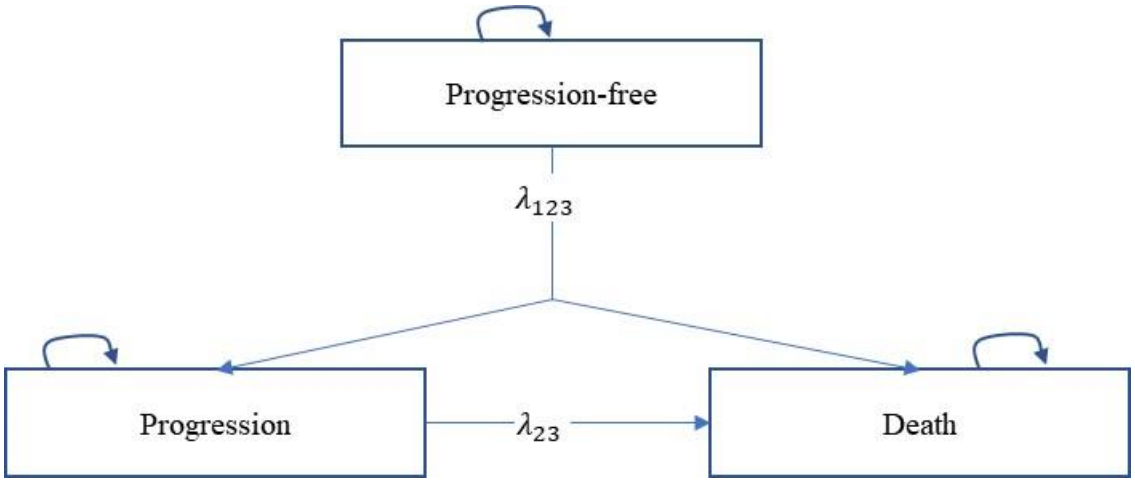
With the multi-state model, transition intensities (rates) for progression (λ_{12}), pre-progression mortality (λ_{13}) and death following progression (λ_{23}) are explicitly modelled as show in Figure 3.

Figure 3 : Conceptual representation of the multi-state model



The STM and PSM as currently implemented in health economics are often conceptually represented using Figure 3. This not correct as the competing transitions are not modelled separately in the version of the STM as implemented in health economics. A better conceptual representation is Figure 4 as PFS is typically used directly with progression and pre-progression death events split. Representing the PSM using Figure 3 (as usually done) is also not appropriate as transitions are not modelled within the PSM. A better conceptual representation is the one shown above in Figure 2.

Figure 4 : Conceptual representation of the STM as implemented in health economics.



When transitions are constant (i.e. event rates are assumed to be exponentially distributed), STM is often called as a (time-homogenous) Markov model. A key property of the Markov model is that it is "memoryless", that is, transitions from one state to another depend solely on the patient's current state and are independent of their previous transitions.[23] When the transition intensities vary with time (for instance, piecewise exponential or other time-varying models), STM is called as time-varying (or non-time homogeneous) Markov model. The term semi-Markov is often used when one or more transition intensities depend on the time spent in an intermediate health state or calendar time;[9] hence, event risk in an intermediate health state is conditional on when the patient entered that state. It should be noted that the term Markov model is often used inappropriately in health economics to describe the general state-transition process, despite most models not being true 'Markov models' in that transition intensities are not constant.[9]

The STM approach is attractive as it can provide greater flexibility compared with the PSM. In particular, the STM approach allows a more conceptually valid representation of the modelling of OS by considering the differences in event hazards between different group of patients who progressed and did not progress (addressing Limitation 1 of the PSM) but also modelling the general underlying natural history process (addressing some of Limitations 2 and 5 of the PSM). Secondly, using data from the intermediate endpoints (TTP/PFS and PPS) could reduce the uncertainty associated with the extrapolation as more events are recorded in people participating to the transition (addressing Limitation 3 of the PSM). Whilst not commonly done, it can also be easier to introduce the dependence structure when modelling PFS and OS (addressing Limitation 4 of the PSM) by making PPS conditionally dependent on TTP (see Chapter 7), although this can also be done with the PSM under a semi-competing risk framework (see Chapter 6.9). Finally, the STM may be argued to be more transparent (addressing Limitation 5 of the PSM) compared with the PSM as the relationship between health states is explicitly stated and transitions between these health states are modelled. This is particularly attractive when there is the need to use external data, for instance, when data on OS are immature or lacking. However, the STM approach is also subject to limitations, with some of the advantages described above also being argued to be disadvantages. As with the PSM, the limitations highlighted below represent some of the arguments that have been made against the STM, and therefore, reflect different points of view (some of which may be arguable or may contradict themselves):

- **Limitation 1: Greater number of assumptions required**. Additional assumptions are required with the STM due to the additional number of parameters (transitions). For example, PFS is assumed to be a good surrogate for OS, which may not always be appropriate. Other assumptions could be that the time to death following progression is same between arms. Assumptions need be made on how to separate progression and pre-progression death events. Assumptions are made explicit in the STM, whilst for PSMs, these assumptions are implicit.

Whilst the explicit statement of assumptions may increase transparency, the number of assumptions required in the STM, may make the model less transparent.

- **Limitation 2: The greater number of variables could increase uncertainty when the extrapolation is already uncertain.** The plausibility of the OS projection is dependent on the robustness and long-term extrapolation of every transition. Combining uncertain transitions (uncertainty associated with extrapolation) increases uncertainty in the resulting OS. This was highlighted by the ERG in TA472 (described in Chapter 1.2), where the estimation of OS was reliant on the extrapolation of immature data for PFS and PPS. In the PSM, OS is a results of the extrapolation of a single event (death).
- **Limitation 3: The use of a post-randomisation outcome (PPS).** There could be some selection biases when estimating the transition from the progressed health state to death (PPS), as this is a post-randomisation measure. For example, for PPS, this becomes problematic if limited numbers of patients have progressed, and/or if those progressing early are expected to have different prognoses to those who have not progressed at the end of the clinical study. This issue is described in further detail in Chapter 7. Furthermore, whilst it is often argued that using PPS helps with the extrapolation, the same number of events (or fewer) are included in the estimation of this transition compared with OS, with the only difference relating to the number of patients included in the denominator (N progressed rather than N randomised). Therefore, whilst the estimate for PPS may be good for those who participated to this transition, there is no information on whether this is appropriate for people who did experience this transition in the trial. Therefore, using PPS may not necessarily make the estimation of OS more reliable, compared with directly fitting a model to OS data (this is discussed in Chapter 7). Another related limitation is that patients progress later are more likely to be censored in the PPS dataset. This would suggest that there is time-dependent bias in the form of informative censoring. Therefore, the STM may only partly address the limitation with the PSM in representing the true OS.
- **Limitation 4: Variation in implementation.** Different approaches/assumptions can also be used when implementing STMs. As described previously, transitions from progression-free to progression and pre-progression mortality could be modelled separately under a competing risk framework (also referred as the MSM). Alternatively, PFS, that is, the cumulative incidence of progression and pre-progression mortality, could be used together with assumptions on the contribution of progression and pre-progression mortality to the overall number of PFS events (referred to here as the Simplified STM approach commonly used in HE). These differences in implementation could potentially lead to inconsistencies in predictions, as demonstrated in Chapter 4.5 and 5.4.

- **Limitation 5: Impact of interval censoring:** PFS may be subject to interval censoring, which may affect its robustness and therefore impact the estimation for OS in the STM.

- **Limitation 6: more difficult to estimate transition in the absence of patient level data**. Although not impossible, it is challenging to implement a STM when IPD are not available (as PFS and OS are correlated).

2.5 Have the approaches been compared systematically?

There is increasing evidence, or at least recognition, that different analytic approaches may lead to different estimates of modelled health gains, and therefore different cost-effectiveness results. These inconsistencies have been highlighted in a number of NICE TAs (described in Section 1.2) including NICE TA257[19] (lapatinib and trastuzumab in combination with an aromatase inhibitor for breast cancer) NICE TA472[20] (obinutuzumab plus bendamustine for follicular lymphoma), and NICE TA381 (olaparib for ovarian cancer).[21] These differences have also been highlighted by other researchers including for example, Williams *et al* (2017),[24] Briggs *et al* (2015),[25] Smare *et al*,[26] Batteson *et al*,[27] and Cranmer *et al*.[28]

Whilst these studies have compared estimates in LYGs and QALYs using the PSM and the STM (including MSM) approaches,[19, 24, 25] they did not compare the different approaches systematically and comparisons were typically limited to single case studies. This makes it difficult to determine whether one approach is consistently superior to another. For instance, in the examples described in Chapter 1.2, the use an alternative approach was justified by the AG/ERG based on their perception of limitations with the approach used by the company.

Furthermore, the comparison in some of these studies could sometimes be deemed unfair due to the different assumption used.[24, 27, 28] As a consequence, it is difficult to understand whether the differences observed are attributable to differences in model assumptions or implementation. For instance, Williams *et al* (2017) compared the PSM with two variations of the STM; the MSM (under a competing risk framework) using the R package `mstate` and the Simplified STM as implemented in a NICE cancer appraisal (TA174) for the first-line treatment of chronic lymphocytic leukaemia (CLL).[24] The authors showed some differences in results between the two variations of the STM (the MSM and the Simplified STM) and the PSM. However, different assumptions were made between the MSM and the Simplified STM, which makes drawing any comparison problematic. Transitions were assumed to be constant in the Simplified STM (as this was the assumption used in TA174). In contrast, transitions were assumed to be time-varying within the MSM. However, the discussion section of the paper acknowledges that under the same assumptions, predictions were closer between approaches in their case-study.

Similarly, in the comparison of the STM and PSM in some of the studies highlighted above, the choice of parametric extrapolation was sometimes questionable[26, 27] and therefore it is difficult to understand whether the appropriateness of an approach is primarily driven by the type of approach or the underlying assumptions (for example, the choice of a parametric survival models used to extrapolate outcomes).

It is possible that one approach may be more appropriate than another depending on the nature of the data available. Whilst these individual comparisons provide some useful knowledge on the performance and limitations of particular methods, model choice is not based on knowledge of the performance of a particular method given the available data, because that information is not available. Perhaps more importantly, whilst some comparisons between approaches are available (single case studies), the assessment of the performance of a method is often limited to the fit of the method to the observed period of a trial. This does not provide any information on whether the long-term extrapolation with one approach is more suitable than another. This is because data are not observed following the end of the trial. However, it should be acknowledged that some studies are available and compared the prediction to earlier cut-off with results from the trial reported at later cut-off, but data were still not complete.[23] It should be noted that a number of other studies are available in the literature and that the examples described in Chapter 1.2, in addition to the additional studies referenced in this section only provide a small selection of studies comparing the different approaches. However, whilst the list provided in this section is far from exhaustive, all the studies identified/known are limited to single case studies, sometimes, with arguable assumptions to allow a fair comparison. No study providing a systematic comparison of approaches is known or has been identified following a rapid search of the literature.

Although no formal systematic review has been conducted, no study providing a systematic comparison of approaches has been identified following a rapid search of the literature (non-systematic search in web-based search engines) or was known by the student, his supervisory team or experts consulted during this thesis. It is reasonable to assume that if such study existed and was available/published, this would be known and commonly cited.

Consequently, a systematic comparison of the different approaches is required to understand whether a particular approach may be more appropriate and perhaps superior to predict PFS and OS (over a lifetime horizon) according to the characteristics of the data.

In the next chapter (Chapter 3), I briefly describe the general concept of survival analysis, censoring, the survivor, hazard and cumulative hazard functions, the Kaplan-Meier and Nelson-Aalen estimates, the Cox regression model and the concept of competing risks. These concepts will be explained through reference to a dataset in breast cancer and are important as they reflect key terms and concepts used throughout the remainder of this thesis.

3 CHAPTER III: KEY CONCEPTS IN SURVIVAL ANALYSIS USED THROUGHOUT THIS THESIS

3.1 Chapter overview

This chapter aims to provide some theoretical background and a brief description of some of the key concepts used in survival analysis that will be used throughout this thesis. These concepts will be necessary to understand some of the approaches described hereafter. Further description and details on survival analysis are available in a number of textbooks including Collett D. Modelling Survival Data in Medical Research, Third Edition. *Textbook - Chapman & Hall/CRC Texts in Statistical Science* 2014).[29] In this chapter, key concepts will be demonstrated through reference to data for the comparator arm from a real trial in breast cancer (CALGB 40502).[30, 31]

The datasets used is described in Section 3.2. The concepts of survival analysis and censoring are described in Sections 0 and 3.4, respectively. The concepts of the survivor, hazard and cumulative hazard functions are described in Section 3.5. Non-parametric estimates of the survivor function (the Kaplan-Meier and Nelson-Aalen estimate) are described in Section 3.6. In Section 3.7 and Section 3.8, I describe the Cox proportional hazards model and the concept of parametric models. In Section 3.9, I briefly describe how parameters are estimated using the maximum likelihood function. Finally, in Section 3.10 I briefly describe the concept of competing risks.

3.2 Description of the datasets used throughout this thesis

A number of datasets are used in this thesis to either illustrate the key concepts in survival analysis (Chapter 3), the implementation of key methods (Chapters 4 and 5), the impact of censoring on the estimation of post-progression survival (Chapter 7) or are used as in case-studies to assess the performance of methods.

Datasets used in this thesis include:

- Individual patient-level data for the control arm from a published randomised Phase III trial of weekly paclitaxel compared to weekly nanoparticle albumin bound nab-paclitaxel or Ixabepilone with or without bevacizumab as first-line therapy for locally recurrent or metastatic breast cancer (CALGB 40502).[31] The breast cancer dataset included 275 patients, of whom 217 had a recorded PFS event (195 progression events and 22 deaths prior to progression).[32] There were 137 death events overall. This dataset was available through the Project Data Sphere.

- Individual patient-level data for the control arm from a published randomised Phase III randomised, placebo-controlled study of docetaxel in combination with zibotentan in patients with metastatic castration-resistant prostate cancer. The prostate cancer dataset included 470 patients, of whom 401 had a recorded PFS event (340 progression events and 61 deaths prior to progression).[32] There were 255 death events overall. This dataset was available through the Project Data Sphere.
- Individual patient-level data from a published randomised Phase III trial of darbepoetin alpha in previously untreated extensive-stage small-cell lung cancer treated with platinum plus etoposide. The lung cancer dataset included 479 patients, of whom 440 had a recorded PFS event (313 progression events and 127 deaths prior to progression).[33] There were 397 death events overall. This dataset was available through the Project Data Sphere.
- A dataset by the GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) group used in a previously meta-analysis to assess the surrogacy between PFS and OS, including a collection of 20 cancer trials in advanced/recurrent gastric cancer publicly available in the R package *surrosurv* (gastadv dataset). The gastric cancer dataset included information on 4,069 individuals treated with different treatments in gastric cancer, of whom 3,820 had a recorded PFS event (2914 progression events and 906 deaths prior to progression). There were 3,635 death events overall.

Table 3 : Summary of characteristics of datasets used throughout this thesis

	Breast	Prostate	Lung	Gastric
Available from:	Project Data Sphere	Project Data Sphere	Project Data Sphere	*Surrosurv* (gastadv dataset)
N	275	470	479	4,069
Number PFS events recorded	217	401	440	3,820
• Progressive event	195	340	313	2,914
• Death prior progression	22	61	127	906
Total number of deaths	137	255	397	3,635
Used in Chapter	Chapter 3-5	Chapter 5	Chapter 5	Chapter 7, 9.

The Kaplan-Meier (KM) estimates (this concept is described in Section 3.6) for PFS, OS, TTP, prePS and PPS are shown in Figure 5, for the breast cancer dataset, Figure 6 for the prostate cancer dataset, Figure 7 for the lung cancer dataset and Figure 8 for the Gastric dataset (collection of trials).

The BC dataset is used in (1) Chapter 3 to illustrate the key concept in survival, (2) Chapter 4 when illustrating the implementation of the multi-state model and (3) Chapter 5 when illustrating the implementation of the STM. Although the three datasets (breast, prostate and lung) are used within this thesis, the breast cancer dataset is used when illustrating the implementation of key approaches and describing key concepts in survival analysis as the breast cancer had a lower number of recorded PFS and OS events

The Lung and Prostate datasets are only used in Chapter 5 when assessing assumptions regarding the estimation of the transition for leaving the progression-free health state (combined transition).

The Gastric dataset is used in Chapter 7 when assessing the implication for using PPS in a trial and in Chapter 9 in case-studies.

Figure 5 : Kaplan Meier Breast cancer dataset

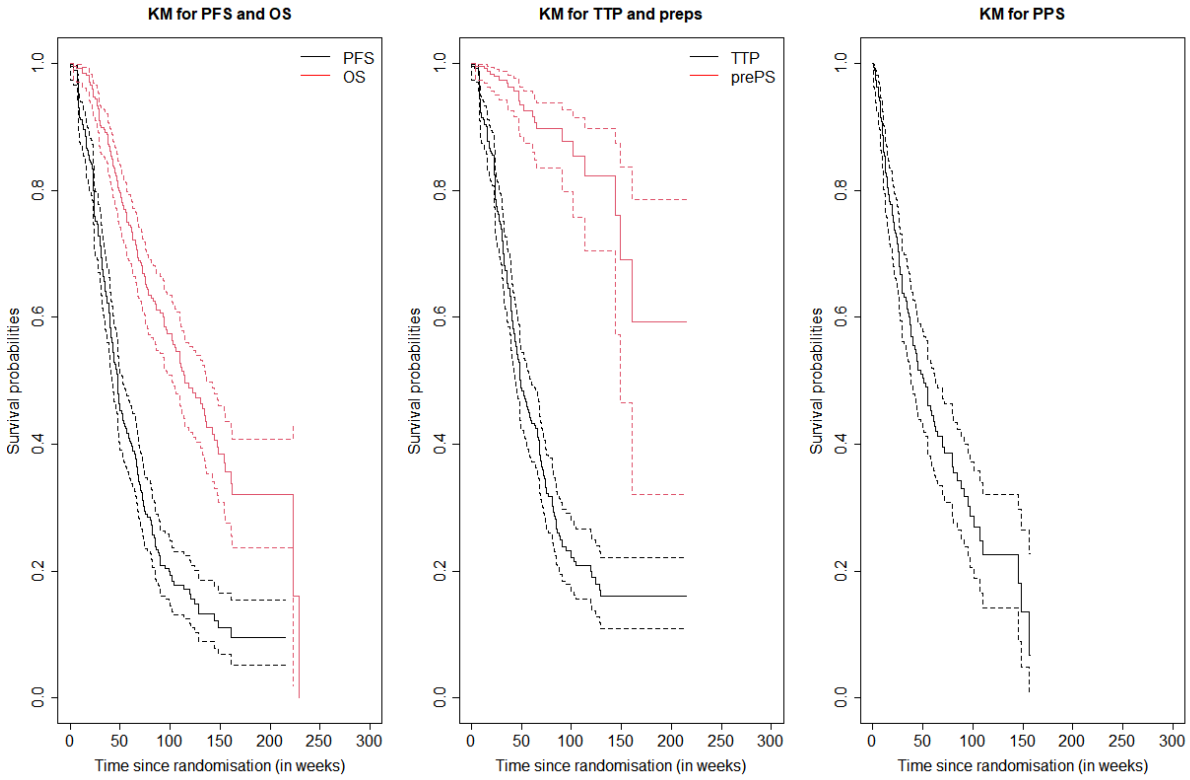


Figure 6 : Kaplan Meier Prostate cancer dataset

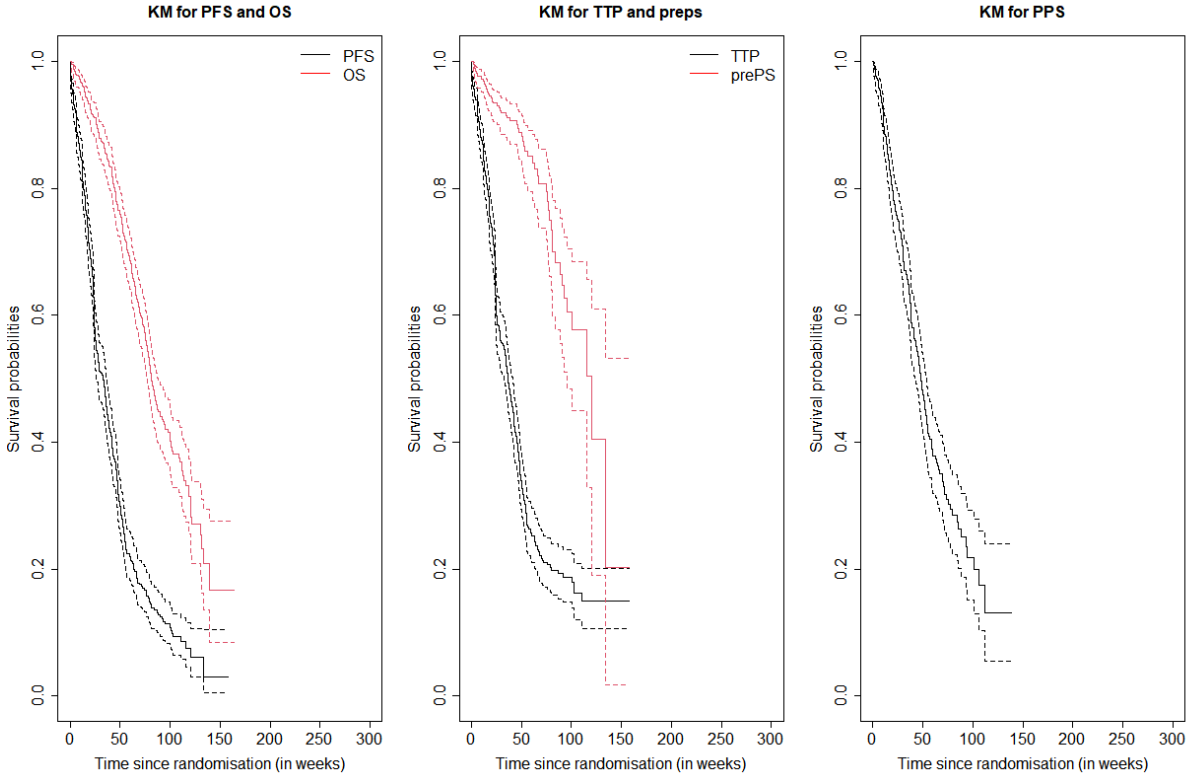


Figure 7 : Kaplan Meier Lung cancer dataset

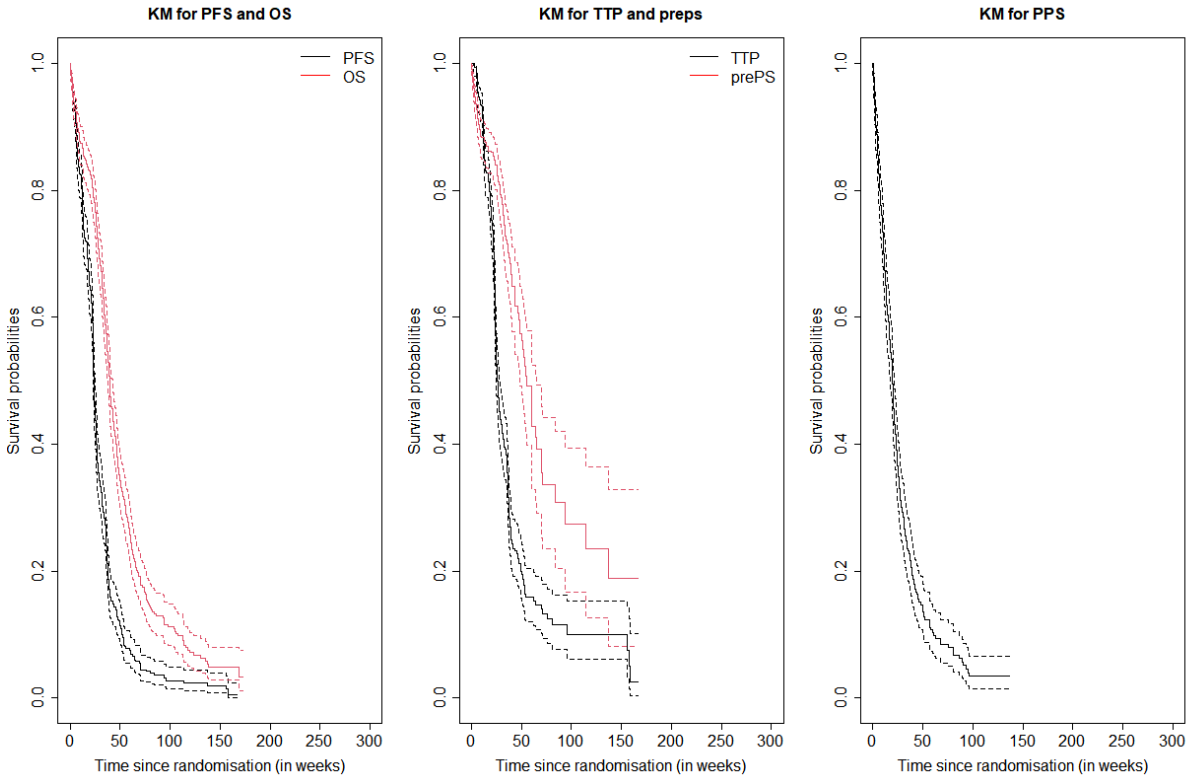
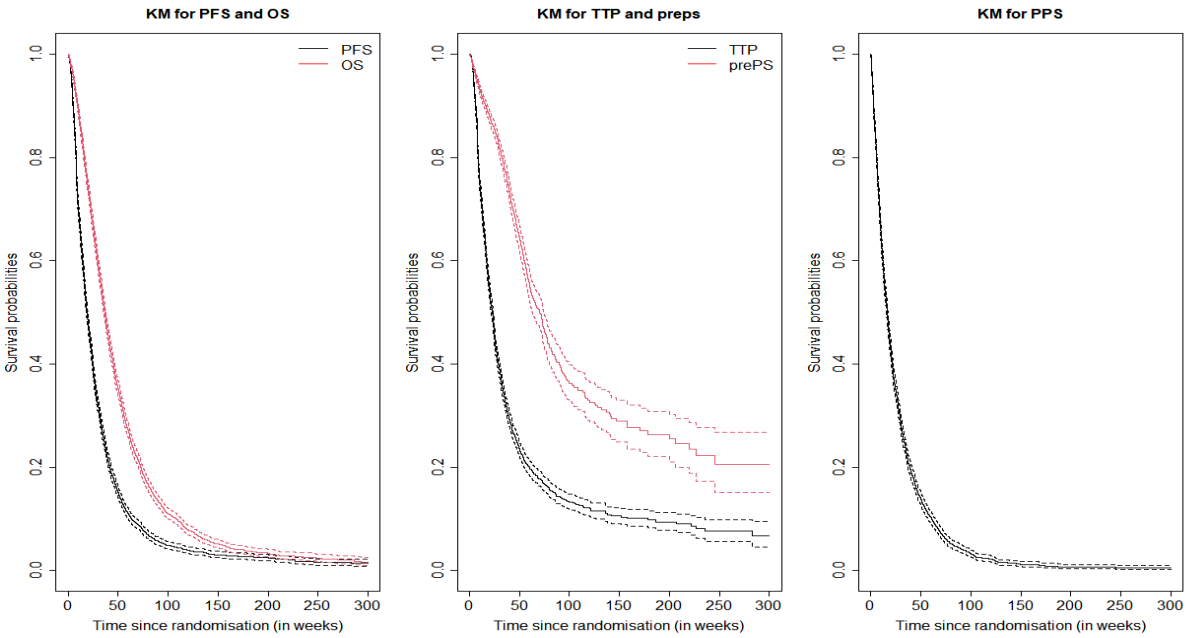


Figure 8 : Kaplan Meier Gastric cancer datasets (collection of trials)



3.3 What is survival analysis

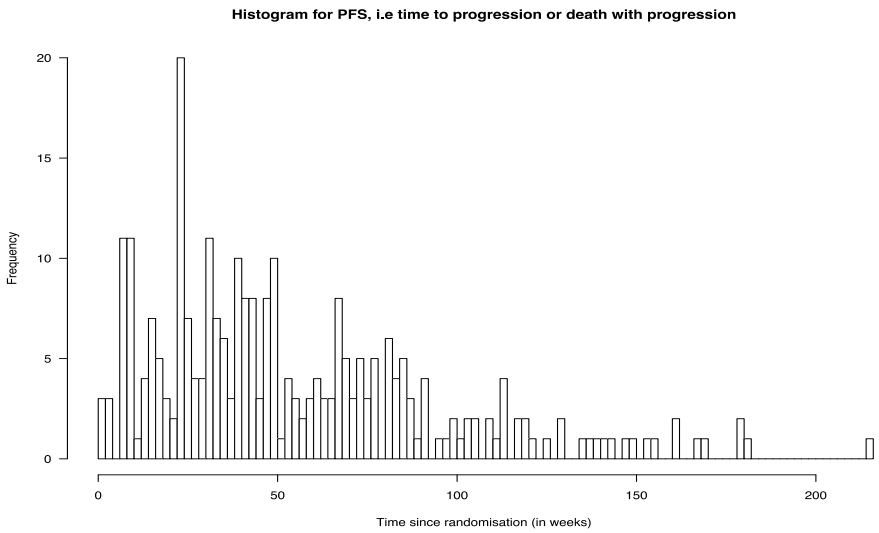
Survival analysis is described by Collett (2015)[29] as "*the analysis of data in the form of times from a well-defined time origin until the occurrence of some particular event or end-point*".

In our context of trials of anticancer treatments;

- The time origin is typically the time of randomisation/recruitment. The time origin could also be the time at which a particular treatment is initiated or the time at which a particular prior event occurs.
- The events/endpoints typically include:
    - Death – measured in terms of overall survival (OS)
    - Progression - measured in terms of time to progression (TTP) or progression-free survival (PFS). The latter includes death occurring prior to progression.
    - Disease recurrence or relapse – measured in terms of disease-free survival (DFS) or event-free survival (EFS). Depending on the study, this usually includes death as an event.
    - Treatment discontinuation – measured in terms of time to treatment discontinuation (TTD). This usually includes death as an event.

Time-to-event data are special and cannot be analysed using standard statistical analysis as these are: (i) frequently censored (the concept of censoring is described Section 3.4); (ii) generally not symmetrically distributed (e.g. positively skewed).[29] This is illustrated in Figure 9 for the outcome of PFS in the BC dataset.

Figure 9 : Histogram for time to death from the BC dataset (CALGB 40502)



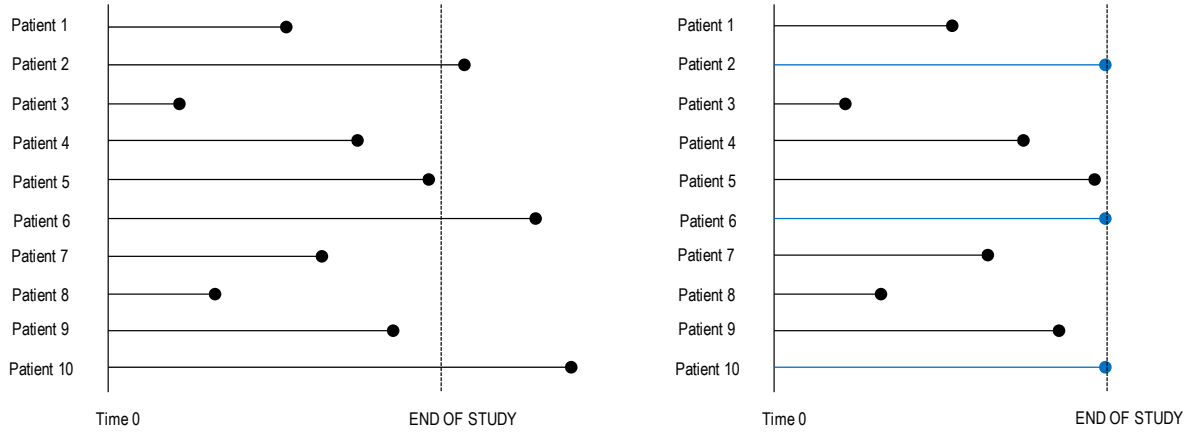
3.4 What is censoring?

The concept of censoring is central to the analysis of survival data. There are different types of censoring. In this thesis, the focus will be on right-censoring as this is the most common type of censoring in oncology trials.

Right censoring occurs when the event of interest (e.g. PFS) has not been observed for a particular individual and therefore the censored survival time is less than the actual, but unknown, survival time. This is called right censoring because the event occurs to the right of the last known observed time-point. There are a number of reasons for an event not to be observed within a trial. This could be because the study ended before the event occurred (also referred as administrative censoring). Alternatively, individuals in the trial could be lost of follow-up, thus, whilst no event was observed up to the last contact with the patient, it is unknown if an event happened subsequently as no information is available from this patient after their last contact. Patients may also experience a separate event (such as death) which prevents the event of interest (progression) from occurring.

The concept of right censoring is illustrated in Figure 10. On the left-hand side of Figure 10, the true time to event is presented. However, it can be seen that for Patients 2, 6 and 10, the event occurred after the end of study follow-up. Therefore, these individuals are said to be censored at the end of the study, as shown on the right-hand side of Figure 10.

Figure 10 : Illustration of right censoring



In the BC dataset, there were 58 censored PFS observations and 138 censored OS observations.

Less common is left censoring. Left censoring occurs when the subject is at risk for the event before the study start. Another form of censoring is interval censoring. Interval censoring occurs when the event of interest occurs within a time interval and the precise event time is not known.

Finally, informative censoring occurs when censoring cannot be directly assumed to be independent of the survival event and censoring provides additional information than the fact that survival time exceeded a certain time.[34] For instance, time to progression could be informatively censored by death.

3.5 The survivor, hazard and cumulative hazard function

Two central concepts of survival analysis are the survivor function and the hazard function. Let us denote T, the survival time, as a continuous random variable that can take any non-negative value. This random variable has a probability distribution with an underlying probability density function (PDF) $f(t)$. The cumulative distribution function (CDF) of T (or cumulative incidence function) is then given by:

$$F(t) = P(T \leq t) = \int_0^t f(u)du$$

with T, the random variable for survival time and t, the actual survival time. This function summarises the cumulative probability of an event occurring before time t.[29]

The survival function, $S(t)$, is the probability that the survival time exceeds time t and can be written as:

$$S(t) = P(T \geq t) = \int_t^{\infty} f(u)du = 1 - F(t)$$

In addition to the PDF, $f(t)$, and the survivor function, $S(t)$, the hazard function $h(t)$ is an important concept and can be defined as the instantaneous event or hazard rate at time *t*. It expresses the event rate at time t, conditional on the event not having occurred before t. This is described by Collett et al (2015)[29] as the probability that the random variable associated with an individual's survival time, T, lies between t and $t + \delta t$, conditional on T being greater than or equal to t.

$$h(t) = \lim_{\delta t \to \infty} \left(\frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t} \right)$$

The cumulative hazard function (i.e. the cumulative risk of an event occurring by time *t*) $H(t)$, can be derived from survivor function $S(t)$ as:

$$H(t) = \int_0^t h(u)du = -log\, S(t)$$

The hazard function $h(t)$ can also be estimated from the survivor function $S(t)$ (and vice versa), and the cumulative hazard function $H(t)$, so that:

$$h(t) = \frac{-dlogS(t)}{dt} = \frac{dH(t)}{dt} = \frac{f(t)}{S(t)}$$

3.6 Non-parametric estimates of the survival function: the Kaplan-Meier (KM) and Nelson-Aalen estimates

The survivor function, $S(t)$, is often estimated non-parametrically in the form of the Kaplan-Meier estimate. The Kaplan-Meir estimate of the survivor function is given by[35]:

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

With t_i denoting the time when at least one event happened, d_i the number of events that happened at time t_i, and n_i the total individuals at risk (have not yet had an event or been censored) up to time t_i.

In brief, a plot of the Kaplan-Meier estimate of the survivor function is a step, in which the estimated survival probabilities are constant between adjacent event time points and which decrease at each event time. Survival times are arranged in ascending order and the beginning of each time interval corresponds to the time at which an event occurred.

Whilst the Kaplan-Meier estimate of the survivor function is commonly used to represent survival data in clinical studies, an alternative estimate of the survivor function is the Nelson-Aalen estimate (or Altshuler's estimate)[36] with $\widetilde{H}(t)$ the Nelson-Aalen cumulative hazard rate estimator given by:

$$\tilde{H}(t) = \sum_{t_i \le t} \frac{d_i}{n_i}$$

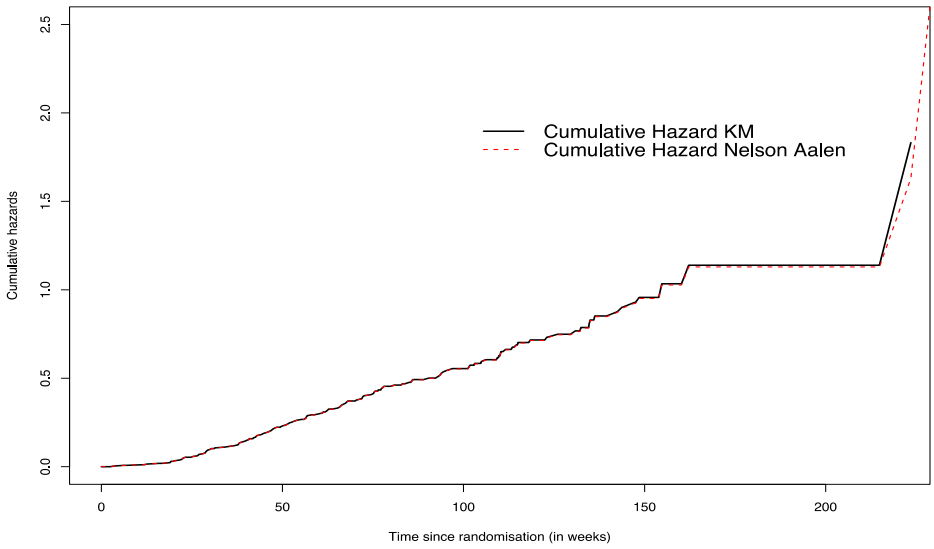
And $\tilde{S}(t)$ the survival function

$$\tilde{S}(t) = e^{-\tilde{H}(t)}$$

The Kaplan-Meier estimate is an approximation and will always be lower than the Nelson-Aalen estimate. However, both estimates are very similar, with the Nelson-Aalen estimate performing better in small samples.

Figure 11 shows the cumulative hazard for OS using the KM survivor function (black line) and Nelson-Aalen estimates (red dashed line). The cumulative hazard function can be obtained from the KM or Nelson Aalen estimate from the relationship between the survivor and cumulative hazard function. It can be seen that the two are very close to each other, with the Nelson-Aalen estimate being slightly higher than the KM estimate.

Figure 11 : Comparison of the cumulative hazard function from the KM and Nelson Aalen-estimate in the BC dataset (CALGB 40502)



3.7 The Cox regression model

Another important concept in survival analysis is the Cox proportional hazards model.[37] This is commonly used to study the effect of covariates on survival time. The Cox model is a semi-parametric model, which assumes a constant effect of the covariates on the hazard function, and thus, an implication is that this model is only appropriate when the survivor function for the groups compared (predictor variable) do not cross, unless time-varying covariates are included.

The Cox model defines the hazard function of an individual i by

$$h_i(t) = e^{\beta x_i} h_0(t)$$

where $h_0(t)$ denotes the baseline hazard function (the hazard function obtained when all covariates are set to zero); β denotes the logarithm of the hazard ratio and x_i denotes the value of the covariate

With the Cox model, we are only interested in estimating the hazard ratio (the β-coefficients). The hazard ratio (HR) is the ratio of the hazard rates for the groups compared according to the explanatory variable. The HR is commonly used to summarise the effect of a treatment in oncology trials, or the effect associated with other variables. The β-coefficients (the HR) could be estimated using the method of maximum likelihood or alternatively a Bayesian framework could be used.

It is important to highlight that the semi-parametric nature of the Cox model means that no assumptions are made about the form of the baseline hazard. Therefore, the Cox model is interested in estimating the effect of a covariate, not the baseline hazard. Once the β-coefficients (the effect of the covariates) have been identified, the baseline hazard function $h_0(t)$ and corresponding survivor function can then be estimated.

Whilst the Cox model provides an important interpretation of survival data, it cannot be used to extrapolate beyond the observed period of a study without further assumptions about the baseline hazard. As previously mentioned, an implication of the Cox model is also that the effects of the covariates are constants; which may not be valid. Therefore, it is important to test this assumption. A number of methods are available, which are not the focus of this thesis, including inspection of Schoenfeld residuals and/or log-cumulative hazard plots.

3.8 Parametric survival models

Whilst Cox proportional hazard models may be useful to estimate treatment effects, health economic models often require the extrapolation of outcomes beyond the study follow-up and therefore, we need to model the baseline hazard. In the Cox model, no probability distribution is assumed. In contrast, parametric survival models assume a specific probability distribution for the survival times and thus, can be used to analyse survival data and also to extrapolate beyond the observed period of a study.

A number of standard parametric models are often considered; these include: the exponential; Weibull; Gompertz; generalised gamma; gamma, log logistic and log-normal distributions. In addition to the standard parametric models, flexible parametric models can provide additional flexibility in the modelling of the shape of the hazard function.[38, 39]

Parametric models are often separated into proportional hazard (PH) models and accelerated failure time models (AFT). In the PH regression model, the effect of covariates is obtained on the hazard function.[40] PH models include the exponential, Weibull, Gompertz and some forms of spline model (hazard forms) and assumed that the effect of a covariate is constant. In the accelerated failure-time regression model, the effect of covariates on the logarithm of the survival time is assessed.[40] AFT models include the log-normal, log-logistic, generalised gamma and assume that the effect of a covariate is increasing (accelerating) or decreasing (decelerating) over time, and therefore is not constant (unless the AFT parameter is 1). It should be noted that some models can be parameterised as either PH or AFT (for example, the Weibull distribution and some spline models).

3.9 Maximising the likelihood function

Whilst most statistical packages allow for the simple fitting of standard parametric models to the data, it is important to understand how coefficients from these models are estimated. Under a frequentist approach, the parameters of the survival functions are typically estimated by finding the set of values which maximise the likelihood (the probability of the data given the model and specific parameter values). Under a Bayesian framework, we derive a posterior distribution for the parameters by combining the prior with the likelihood. Observations can be censored or not. Therefore, the likelihood function has two components: one for the observations that are censored (where no event occurred) and one for the observations that are not censored (where the event occurred). Events and censored times contribute differently to the calculation of the likelihood. For completely observed events, the contribution is the whole PDF. However, for right censored events, all we know is that they contribute at least up to a certain point, and therefore we use the survival function, S(t) for censored time. The likelihood of the data is then the product of the likelihood across all patients (observations and censored times). The overall likelihood is

$$L(\theta) = \prod_{\delta_i=1} f(t_i;\theta) \prod_{\delta_i=0} S(t_i;\theta)$$

$$L(\theta) = \prod_{\delta_i=1} h(t_i;\theta) S(t_i;\theta) \prod_{\delta_i=0} S(t_i;\theta)$$

$L(\theta) = \prod_{\delta_i=1}^{n} h(t_i;\theta)^{\delta_i} S(t_i;\theta)$ Where δ_i is the censoring indicator.

3.10 Competing risks

Another key concept in survival, notably in the analyses of multiple survival data (or endpoints) is the notion of competing risks. Competing risks occur when the occurrence of one event influences or hinders the occurrence of another event.

Standard survival analysis censors follow-up at the occurrence of the competing event. Therefore, it is implicitly assumed that individuals could subsequently experience the later event. This therefore assumes independence between the event of interest and the competing event (censored individuals are assumed to have the same survival as those who remain in follow-up), which may not be true. It is difficult to test for the independence of events.

In contrast, within the competing risk framework, competing events are not deemed independent and multiple events cannot occur simultaneously. In the presence of competing events, the marginal probability of each competing events can be estimated from the cumulative incidence function (CIF), which is derived from the cause-specific hazard. The CIF for an event c at time t_f can be written as[41]:

$$\mathrm{CIF}_c(t_f) = \sum_{f'=1}^{f} \hat{I}_c(t_f) = \sum_{f'=1}^{f} \hat{S}(t_{f'-1}) \,.\, \hat{h}_c(t_{f'})$$

Where $\hat{I}_c(t_f)$ is the incidence probability of failing from event-type c at time t_f, and $\hat{S}(t_{f'-1})$ the overall probability of survival at a previous time.

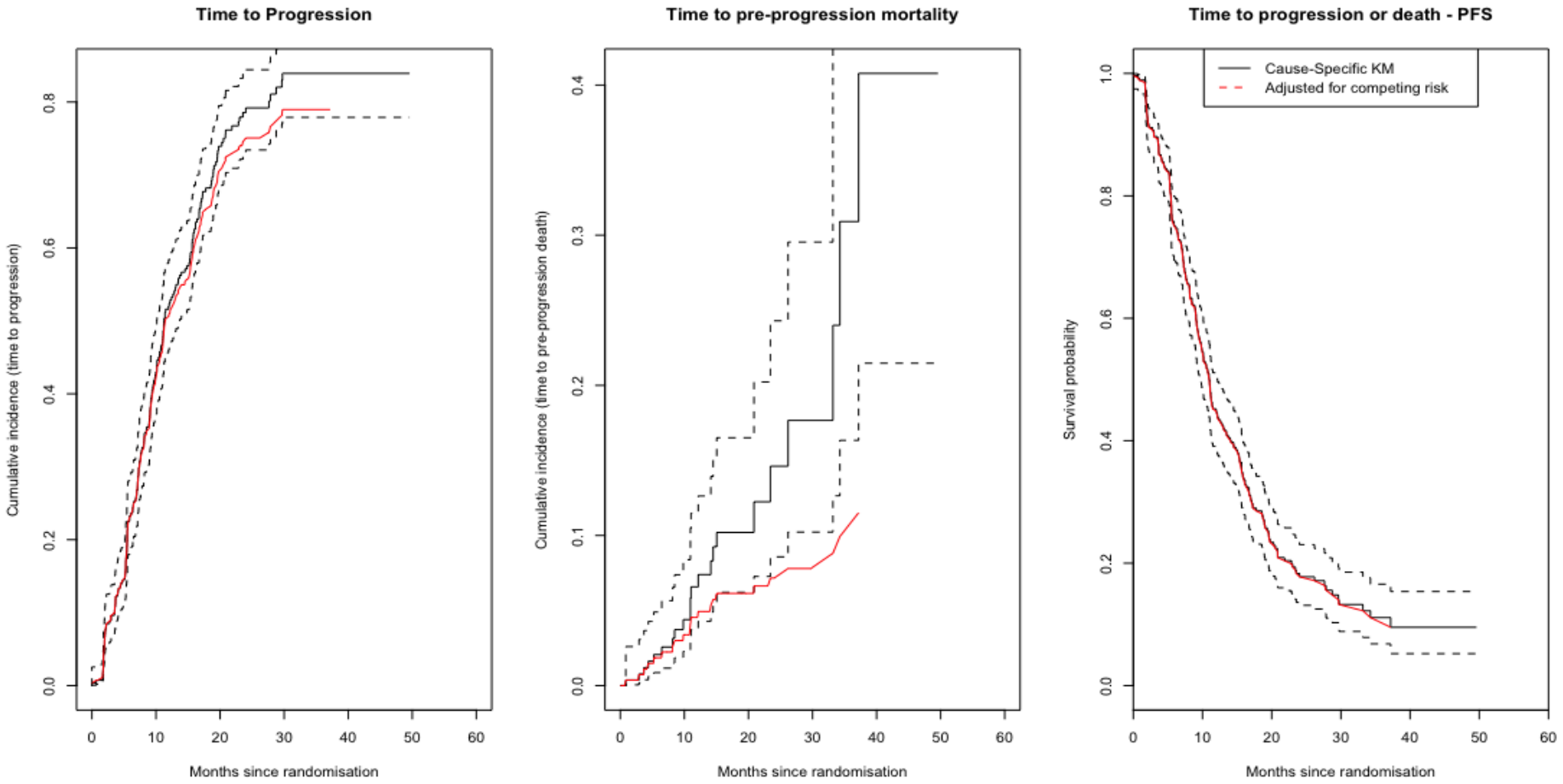
The CIF is equivalent to 1-KM estimator when there is no competing event.

It should be noted that the incidence function of the two competing events is equal to the cumulative incidence function. Therefore, considering PFS as an example, the incidence function for time to progression (TTP) and pre-progression mortality (PrePS) will be equal to the cumulative incidence function (PFS).

This is illustrated in Figure 12 using data on PFS from the BC dataset.

It can be seen that standard approaches and approaches that properly deal with competing risks are similar when analysing a combined composite endpoint such as PFS, but that a proper competing risk approach is required when analysing the component parts of PFS to avoid bias. In Figure 12, the red line represents the incidence function adjusted for competing risks for TTP and PrePS separately in the left and middle graphs, and is lower than the incidence functions estimated using standard survival analysis. In the graph on the right, the red line, represent the cumulative incidence function for both time to progression and time to pre-progression mortality combined, which is the same as PFS.

Figure 12 : Naive KM and KM allowing for competing risks for the cumulative hazard and the cause-specific hazard in BC dataset (CALGB 40502)*



* dashed line represents the 95% CI

It should be noted that within the competing risk framework, there is no longer a direct relationship between the cumulative incidence, survivor and hazard functions. Therefore, alternative approaches are required when examining the impact of an explanatory variable (the most common of which is the Fine and Gray model[42]).

Competing risks analysis is a large topic in its own right, and therefore is beyond the scope in this thesis. However, an understanding of the key principles is required to understand the MSM approach described in Chapter 4 and how this compares with the direct PFS fit in Chapter 5.

This chapter has provided a brief description of the key concepts used in survival analysis, including competing risks. In Chapter 4, I describe the MSM approach with a reference to the concepts described above and through reference to the BC dataset.

PART II: MODELLING THE PROCESS OF PROGRESSION

4 CHAPTER IV: MODELLING TRANSITIONS USING THE MULTI-STATE MODEL UNDER THE COMPETING RISK FRAMEWORK

4.1 Chapter overview

In this chapter, I focus on the implementation of the multi-state model (MSM) using two R packages (as R was the software used in this thesis for the simulation study in Chapter 8): the `msm` package and the `mstate` package. I describe the key differences and how transitions are combined under a competing risks framework.

Section4.2 introduces this chapter. The implementation of the MSM using the `msm` package when transitions are constant is described in Section 4.3. In this section, I describe how to implement the MSM using this package, and explain how transitions are estimated. I also briefly discuss how the Markov assumption (when transitions are constant) can be relaxed by using piecewise exponential models.

Section 4.4 details the implementation of the MSM using the `mstate` package. In Section 4.4.1, I describe how transitions are combined under a competing risk framework in the `mssample` function (which is part of the `mstate` package). In Section 4.4.2, I provide a summary of findings following a review of R code with an example of implementation in Section 4.4.3.

The term MSM will be used thorough this thesis to refer to the multi-state model, with `msm` used to refer to the package in R.

4.2 Introduction

As illustrated in Chapter 2 (Figure 3), three transitions are required in the three-state model typically used in the economic analysis of treatments for advanced/metastatic cancer:

- the transition from progression-free (PF) to progressed disease (PD) (λ_{12});
- the transition from PF to death (λ_{13}), and;
- the transition from PD to death (λ_{23}).

The transitions from the PF health state (to PD or death – λ_{12} and λ_{13}) are competing, in that the chance of experiencing one event (transition) is hindered by experiencing another event (transition). Therefore, transitions must be combined under a competing risk framework. This can be done through the MSM,

which is an extension of competing risk analysis[43, 44]. Multi-state modelling is widely used in the field of statistics but is rarely used in health economics. The reason for this is unclear and could be due to lack of awareness or its complexity compared with standard methods which are currently used.

A number of packages in R software are available to implement MSMs, making this a more attractive option and subject of research in health economics.[43, 45] For instance, the implementation of the MSM in health economics has been recently described by Williams *et al*[24] in a step-by-step tutorial using the `mstate` package developed by Putter *et al*.[43] An alternative package to implement MSMs in R - the `msm` package - has been developed by Christopher Jackson.[45] The `msm` package has been used in at least three NICE technology appraisals; TA586[46], TA587[47] and ID945[48] as highlighted later in Chapter 5.3.5.

More recently, a user-friendly MSM package has been developed in STATA software.[49] It uses the same principle as the function in the `mstate` package, and therefore, is not the focus in this chapter. However, it should be noted that the STATA function is perhaps more flexible and quicker to run and the decision to focus on R was taken as this was the software used for the simulation study.

In this chapter, I focus on the R packages and describe the implementation of the MSM using both the `msm` package[45] and the `mstate` package as these have different flexibilities and characteristics.[24, 50]

4.3 Performing multi-state modelling using the `msm` package when transitions are constant

4.3.1 How are transitions estimated and combined in the `msm` function?

The `msm` package is comprised of a number of pre-defined functions; the `msm` function is used to estimate the parameters for the MSM. A key characteristic of the `msm` function, is that it fits the MSM to the available time-to-event data directly using maximum likelihood estimation, therefore, few steps are required when implementing this form of model. The transitions (or parameters) are estimated and combined endogenously within the `msm` function. This contrasts with the `mstate` package (described in Section 4.4) whereby transitions (or parameters) are estimated exogenously from the function and combined within the function.

The `msm` package includes a number of options to fit the model to data from processes with arbitrary observation times (usually panel data on state-occupancy; observations obtained over multiple time periods[45]), exactly-observed transition times (usually time-to-event data where the time to transition is the event), censored states (i.e. information is not known about the state occupancy), or a mixture of these types of data. Therefore, the `msm` function is sufficiently flexible to deal with a variety of data types and interval censoring.

In the `msm` function, an initial transition matrix is specified with the parameters (that are unknown) being estimated by maximum likelihood via optimisation. The `msm` function is intuitively attractive as it estimates the parameters of the model directly from the data.

The default of the `msm` function is that the model is Markov, whereby event rates are assumed to be constant from start to the end (survival distributions are exponentially distributed) and are independent of previous transition events (although this can be relaxed, as described in Section 4.3.3). This is very restrictive; Christopher Jackson (author of the function) acknowledges that such an assumption was made to accommodate the computation of the likelihood for intermittently-observed processes within the function.[45] Nevertheless, the `msm` function offers the option to relax this assumption, for example, by assuming that transitions are different between time intervals, e.g. using piecewise exponential distributions conditional on time. And therefore, the function is very flexible.

Transition probabilities (which are assumed to follow an exponential distribution [or piecewise]) are estimated from the likelihood function which varies according to the type of data. The focus in this section will be on exactly observed transitions times as this reflects the type of data typically available from oncology trials (i.e. PFS and OS). It should be noted that the option of processes with 'arbitrary observation times' could be relevant, when data allows, as this will be similar to assuming interval censoring between two time intervals. I also focus here on the case whereby transitions are constant over time (rather than varying between time intervals).

The contribution of the likelihood for exact transition times[45] between states is written as:

$$L_{i,j} = \exp\left(q_{S(t_j)S(t_j)}(t_{j+1} - t_j)\right) q_{s(t_j)s(t_{j+1})} \qquad (4.1)$$

Where $S(t_j)$ represent the state an individual at time *t*, and $q_{s()s()}$ the instantaneous risk of moving between state

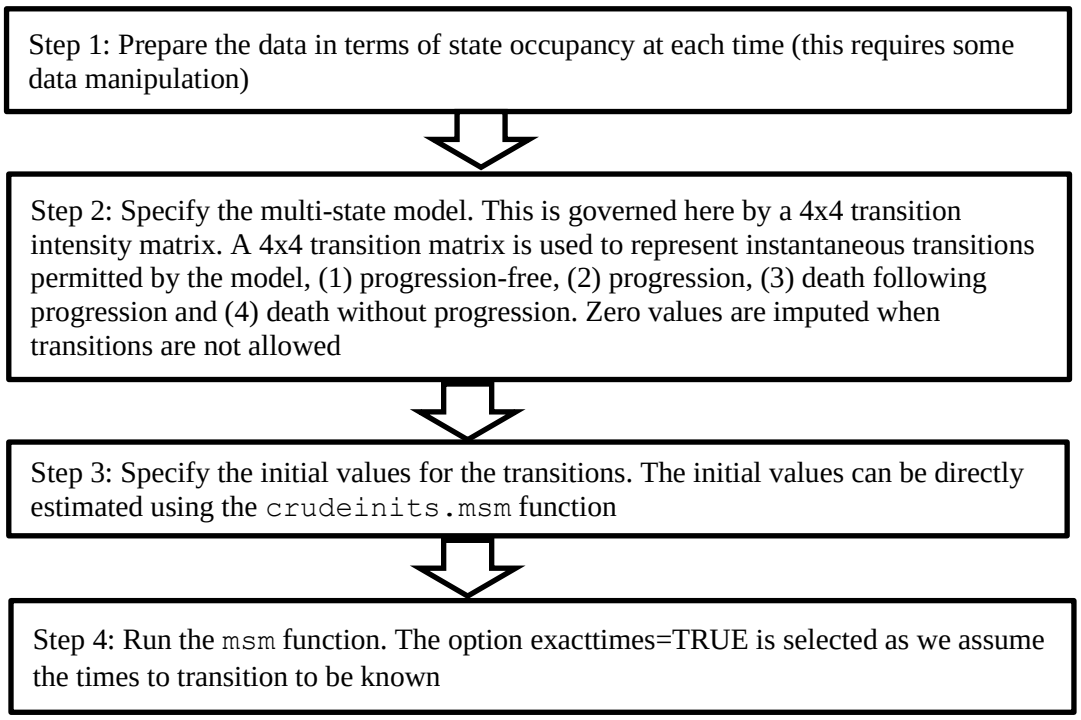
The expression for the likelihood in Equation (4.1) resembles the expression for the likelihood for general survival data and is composed of two elements: when transitions are observed or when transitions are censored. Thus, estimates for the non-competing transition (progression to death) will be the same within the multi-state model using the `msm` function or estimated externally from the data using PPS. Similarly, as shown in Section 3.10, PFS represents the cumulative incidence of the two competing transitions. Therefore, under the assumption of exponentially distributed event times, the probability of remaining progression-free in the MSM will be the same as the survival function for PFS. This is demonstrated in the Section 4.3.2.

4.3.2 Implementing the multi-state model using the `msm` function in the BC dataset (CALGB 40502) assuming transition to be constant (Markov model)

The implementation of the MSM using the `msm` function in the BC dataset is described hereafter using the option for "exactly-observed transition times" to reflect the type of data available. The implementation of the model using the `msm` function in the BC dataset can be summarised in four key steps, as shown in Figure 13.

Transitions are assumed to be constant throughout the patient's lifetime and therefore, the model is assumed to be Markov.

Figure 13 : Summary of key steps for implementing the multi-state using the `msm` function in the BC dataset (CALGB 40502) assuming the model to be Markov



Predictions for the state occupancy in PF, PD and time to death using the `msm` function are shown later in Figure 24 alongside the predictions from the `mssample` function and simplified STMs.

As described in Section 4.3.1, given the calculation of the likelihood function for exactly observed times, the transitions (leaving pre-progression and progression health state) estimated endogenously from the MSM using the `msm` function are expected to be the same as the transitions estimated externally assuming the survival distributions are exponentially distributed. To confirm this, I compared the estimate for the transition probabilities estimated endogenously using the `msm` function and the transition probabilities estimated externally from the data:

- The transition probability from the progression health state to death (state [ii] → state [iii]) was estimated to be 0.05697 using the `msm` function. The same estimate was obtained when PPS was estimated externally from the data assuming the survival distribution to be exponentially distributed,
- Similarly, the probability of remaining progression-free in the MSM was the same as the inverse of the PFS probability estimated externally from the data assuming the survival to be exponentially distributed.

4.3.3 Relaxing the Markov assumption

As described in Section4.3.1, the `msm` function assumes transitions to be constant and therefore the survival distributions are exponentially distributed (with the simple form assuming the model to be Markov, with transitions constant over the time horizon). However, it is possible to relax this assumption by fitting the MSM (using the `msm` function) to data for different time intervals to allow for time-varying transition probabilities (using piecewise exponential distributions). In the simple case, one may assume that the event rates are different between two time periods (e.g. the hazards change after some defined time-point). However, the `msm` function accommodates more complex forms, i.e. assuming transitions vary within more than two time intervals or assuming that only specified transitions are time-varied. For example, in TA586,[46] the `msm` was fitted to three time intervals in order to account for time-varying probabilities (Chapter 5.3.5). Functions are available within the `msm` package to assume directly piecewise exponentials, or constant transitions within a defined number of intervals. The `msm` function can also be extended to include the effect of covariates. This can be done by using the `pci` option within the `msm` function (to allow piecewise exponential) or the `piecewise.msm` function for more complex forms (when the effect of covariates is of interest). A key limitation with using piecewise exponential is that the extrapolation beyond the observed period will be based only on the fit and extrapolation of the last time period, rather than the entire data.

4.4 The `mstate` package; a flexible alternative to perform multi-state modelling to include any parametric distributions?

The `mstate` package was developed by Putter *et al*[43] and was subsequently described in a step-by-step tutorial for health economic modelling by Williams *et al*.[50] Compared with the `msm` package,[45] the `mstate` package[43] uses parameters for the transitions (cumulative hazard) as an input, rather than fitting the MSM to the data. Parameters for each transition are therefore estimated exogenously from the MSM, with transitions combined under a competing risk framework within the function. The `mstate` package, compared with the `msm` function, allows any form of survival distribution to be used.

Two key functions within the `mstate` package can be used to generate the MSM; (a) the `probtrans` function which uses a cohort approach, and (b) the `mssample` function which uses a simulation approach. These functions (`probtrans` *and* `mssample`) are used by Williams *et al*.[24] in her tutorial. Williams *et al* recommends the use of the `probtrans` function when the Markov property holds and the `mssample` function when the Markov property is violated. Nevertheless, whilst Williams *et al*.[24, 50] suggest the use of the `probtrans` function when the Markov property holds, following my own review of the function, I believe that the `mssample` function is equally appropriate in such situations, with transitions being constant (the survival distribution is exponentially distributed). This was confirmed following contact with the author, who suggested that when the Markov assumption holds (rare assumption), the key benefit of using the `probtrans` over the `mssample` function is that it uses a cohort approach (hence, no simulation is required). In this section, I focus on the implementation of the MSM using the `mssample` function as it encompasses both situations whereby the Markov assumption holds and where it is violated.

A key characteristic of the `mssample` function is that it uses data on the cumulative hazard for each cause-specific transition (estimated exogenously from the MSM), with data censored when the competing event occurs. Thus, considering the typical three-state advanced/metastatic cancer model structure, the `mssample` function uses three inputs;

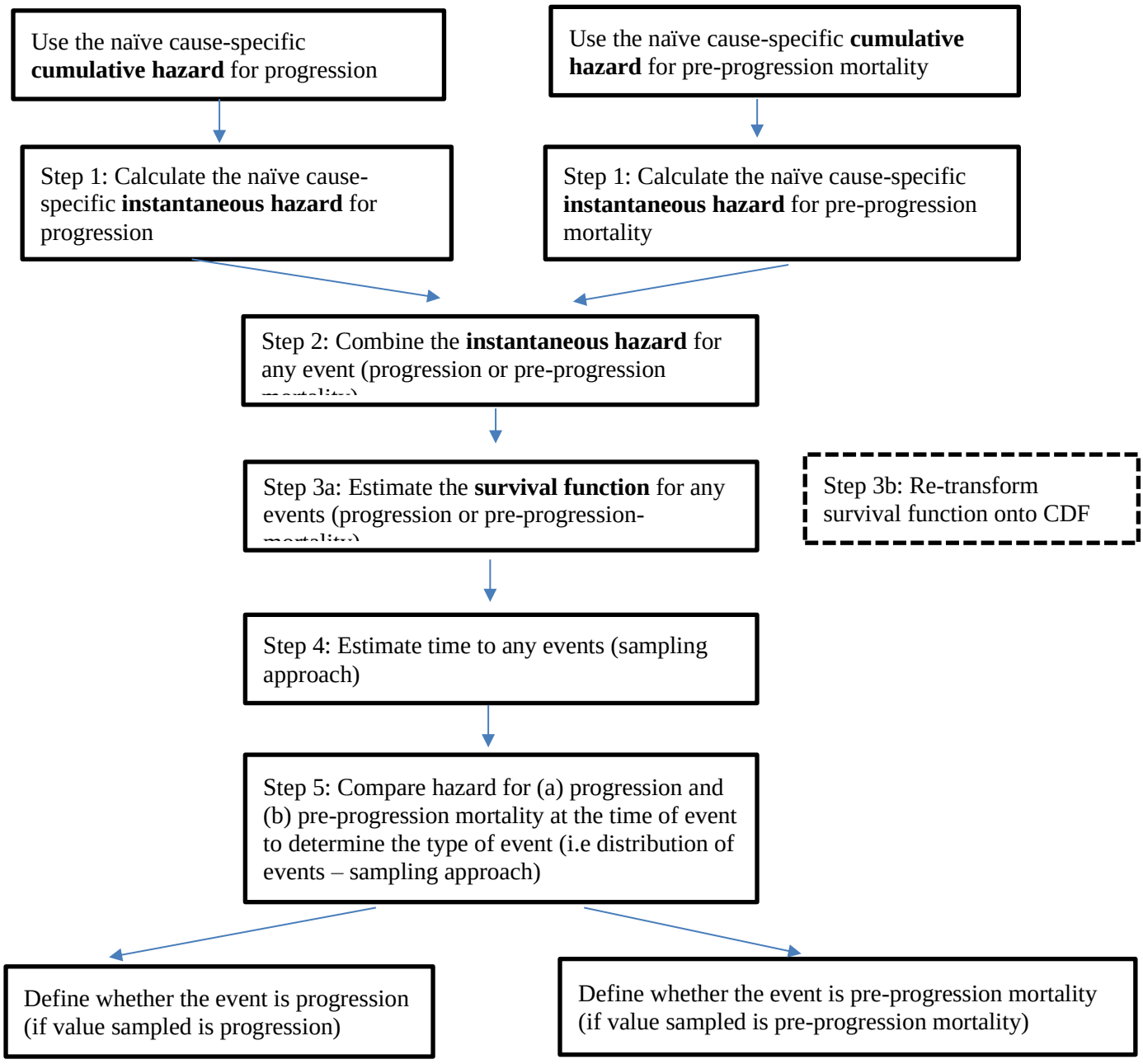
(i) The cumulative hazard for TTP. For this transition, PFS is censored for PrePS,
(ii) The cumulative hazard for PrePS. For this transition, PFS is censored for progression,
(iii) The cumulative hazard for the PPS.

These transitions (estimated exogenously from the function) are then combined under a competing risk framework within the `mssample` function. For this thesis, the R code of the `mssample` function (the key function used when generating the multi-state model) was reviewed in detail in order to understand how the inputs (the cumulative hazard function for each transition) to the function are transformed and then subsequently combined under a competing risk framework. It should be noted that the `mssample` function uses a number of associated pre-defined functions pertaining to the `mstate` package which were also reviewed in order to understand how the function works; in particular: the `mssample1, crsample, Hazsample, NAfix and to.trans2` functions. A description of the how inputs (cumulative hazard functions) for the competing transitions are combined is described in Section 4.4.1 with findings from the review of the R code highlighted in Section 4.4.2. The implementation of the MSM using the `mssample` function in the BC dataset is described in Section 4.4.3.

4.4.1 How are inputs (cumulative hazard) for the competing transitions combined under a competing risk framework in the `mssample` function?

The approach used within the `mssample` function to combine the competing transitions under the competing risk framework can be summarised in five key steps, as illustrated in Figure 14 for the outcome of PFS (cumulative incidence of progression and pre-progression mortality events). A narrative description is provided below.

Figure 14 : Illustration on how transitions are combined under the competing risk framework in the `mssample` function



In brief, the `mssample` function uses data on the 'naïve' cumulative hazards (not adjusted for competing risks) as inputs for the two separate cause-specific transitions (TTP and PrePS), with the competing risk censored. Transformations are then made inside the `mssample` function, with transitions combined under the competing risk framework following five key steps:

- **Step 1: Estimation of the 'naïve' instantaneous hazards (unadjusted for competing risk) for progression ($\mathbf{h_1(t)}$) and pre-progression mortality ($\mathbf{h_2(t)}$).** The instantaneous hazard is approximated by taking the difference between the cumulative hazard at time t and t-1. This is a discrete time approximation for the differential of the cumulative hazard over the relevant time interval:
- **Step 2: Calculation of the instantaneous hazards for the cumulative incidence of events (any events [progression or pre-progression mortality]).** The instantaneous hazard for the cumulative incidence of events ($h_3(t)$) is calculated by summing the hazard rates for progression ($h_1(t)$) and pre-progression mortality ($h_2(t)$)
- **Step 3a: Transformation of the instantaneous hazard for the cumulative incidence of events (PFS) onto a survival function $\mathbf{S(t)}$.** This is done by multiplying the survival at time t-1 by the 1 minus the instantaneous hazard at time t, $h(t)$. It should be noted that as described later in Chapter 4.4.2.1, the `mssample` function incorrectly uses the instantaneous hazard, instead of the probability ($p(t)$) when calculating the survival function. Doing so, will lead to the estimate of PFS (which is the cumulative incidence for both events).
- **Step 3b: Estimation of the cumulative hazard for PFS.** The survival function is then transformed back onto a cumulative hazard function $H(t)$ based on the relationship between $H(t)$ and $S(t)$ so that : $H(t) = -\log[S(t)]$
- **Step 4: Sample the time to any event (progression and pre-progression mortality).** The time to any event (progression or pre-progression mortality) is then sampled from the cumulative hazard curve to determine the time at which time an event (any) would occur.
- **Step 5: Define whether the event is progression or pre-progression death.** In Step 4, the time to any event is sampled. However, this could be either a progression event or a pre-progression mortality event. At the point of failure (the event), the hazard for the two competing events is compared to determine whether the event is progression or death (to identify which event is more likely). From the comparison of two hazards (progression and pre-progression death), the probability for the event to be one or the other is calculated. For instance, if the event is assumed to occur at time t, and the instantaneous hazard for TTP and prePS are equal at that time, there will be an equal chance for the event to be either progression or death. If the instantatenous hazard at time t is lower for TTP (compared with prePS), the probability for the event to be pre-progression mortality will be higher. However, the event could still be pre-progression mortality (despite lower hazard).

4.4.2 Findings upon review of the R code of the `mssample` function

As previously highlighted, for this thesis, in order to understand how the `mssample` function works and how competing transitions are combined under a competing risk framework, I reviewed the R code of the function (and associated functions).

Upon review of the R code, an inconsistency was identified during this process. A further clarification on how the function worked is also highlighted.

4.4.2.1 Rates in the `crsample` function appear to be incorrectly treated as probabilities

Upon the review of the R code, an inconsistency was identified in the `crsample` function (which is used in the `mssample` function) in that the instantaneous rates $h_{(t)}$ are treated as probabilities when deriving the survival function.

The survival function (`ci$S0`) is calculated as the cumulative multiplication result (`cumprod`) of 1 minus the instantaneous hazard (`ci$hazsum`)

This is not correct as shown in Figure 16, as the instantaneous rates need to be transformed to a probability first so that $p(t) = [1 - (1 - \exp(h(t))]$. The impact of this error is likely to be minimal when rates are very small as they become closer to probabilities. However, when the function is applied to models which use larger time increments, and when event rates are higher, the impact of the error could become more significant.

It is however possible to correct for this inconsistency simply by slightly amending the R code, so that instantaneous rates are first transformed onto probabilities prior to deriving the survival function so that:

$p(t) = [1 - (1 - \exp(h(t))]$ *instead of* $p(t) = [1 - h(t)]$

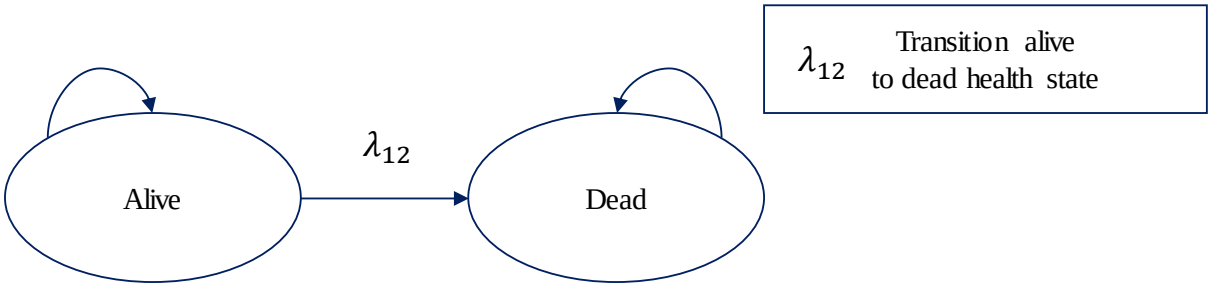
In the R code,

`ci$S0 <- cumprod(1 - -ci$hazsum)` (Original line of code)

needs to be amended to the following: `ci$S0 <- cumprod(1 - [1-exp(-ci$hazsum)])`

The presence of this inconsistency can be demonstrated easily by considering the simplest case of the MSM; the illness-death model, whereby individuals can either be alive or dead (Figure 15). Therefore, in this case, competing risks are not considered here, as the model includes only a single transition (from alive to dead).

Figure 15 : The illness-death model



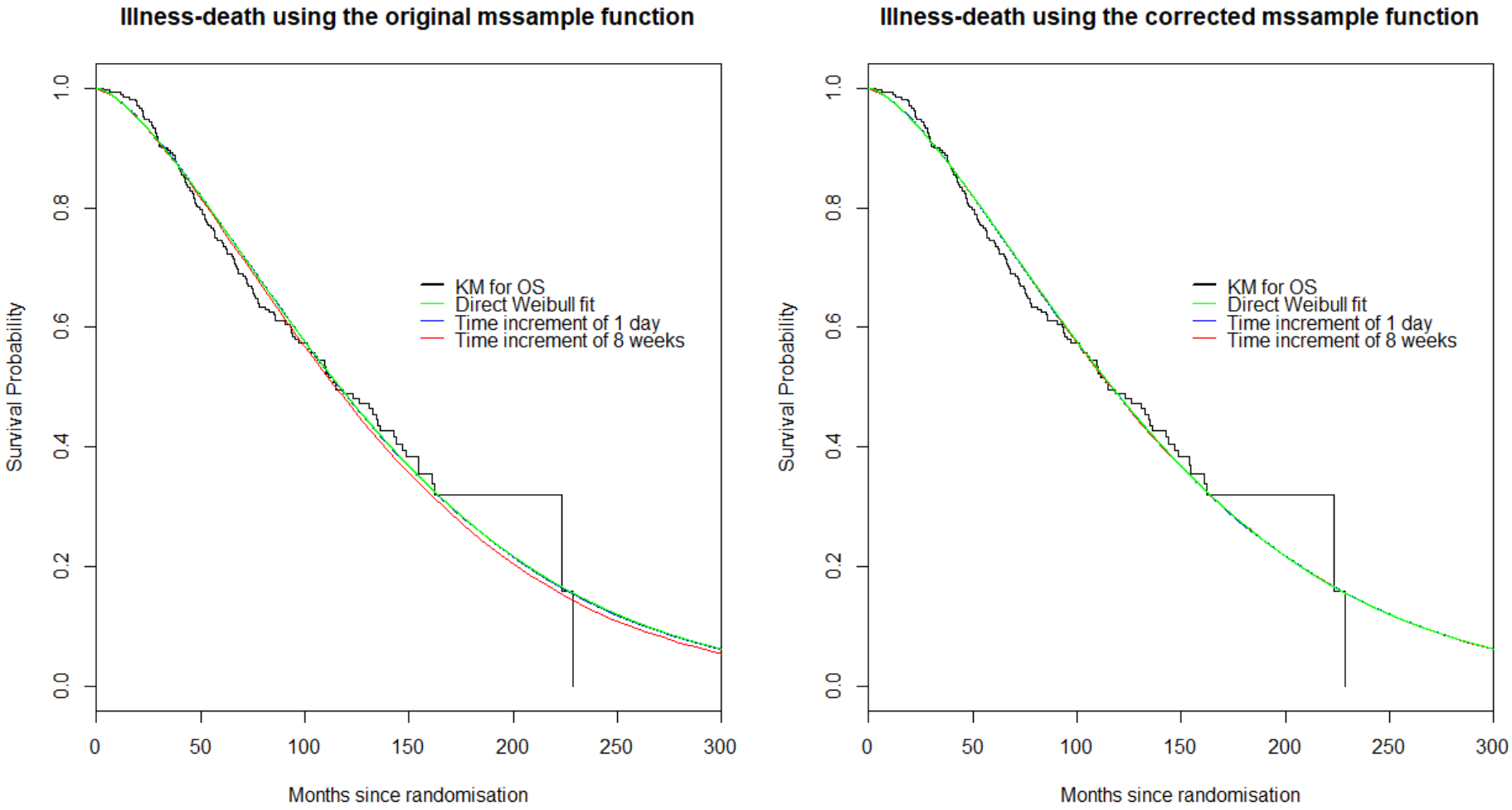
The MSM for the simplified illness-death model was run using: (i) the original `mssample` including the original `crsample` functions, and (ii) a corrected version of the `crsample` and `mssample` functions. For the purpose of illustration, the models were run assuming two extreme scenarios: (a) a very short time increment of one day, and (b) a longer time increment of eight weeks (56 days). The only transition in this model (death) was informed by a Weibull survival distribution fitted to data on OS from the BC dataset (described in Chapter 3.2) for illustration. A total of 150,000 patients were simulated within the `mssample` function, with the same random seed used to reduce any influence caused from sampling variation. Figure 16 shows the predictions from the two models (the original and corrected models) alongside the KM for OS and the direct fit to the data using the Weibull distribution.

In the absence of inconsistency in the `mssample` function, one would expect the predictions from the MSM to be identical to the direct Weibull fit irrespective of the time increment used. However, it can be seen from the left-hand panel in Figure 16 that using the original `mssample` function, different time increments provide different predictions, and that smaller time increments lead to predictions which are closer to the direct fit using the Weibull distribution, as expected. As previously highlighted, this inconsistency is likely to be attributable to the error made within the `crsample` function which incorrectly treats rates as probabilities when deriving the survival function.

The location of the error (transformation of rate onto probabilities) is also hinted given that predictions from the MSM get closer to the direct Weibull fit when the time increment is small (1 day). In contrast, following the amendment of the `crsample` function, it can be seen that predictions using the corrected version of the `mssample` function, shown on the right-hand panel of Figure 16 are more stable and, as expected, the predictions are identical to those derived from the directly fitted Weibull distribution , irrespective of the time increment used.

The author of the `mssample` function was contacted by email on the 19th October 2017 to highlight this error.

Figure 16: Predictions for the illness-death model using different time increments prior and after correction



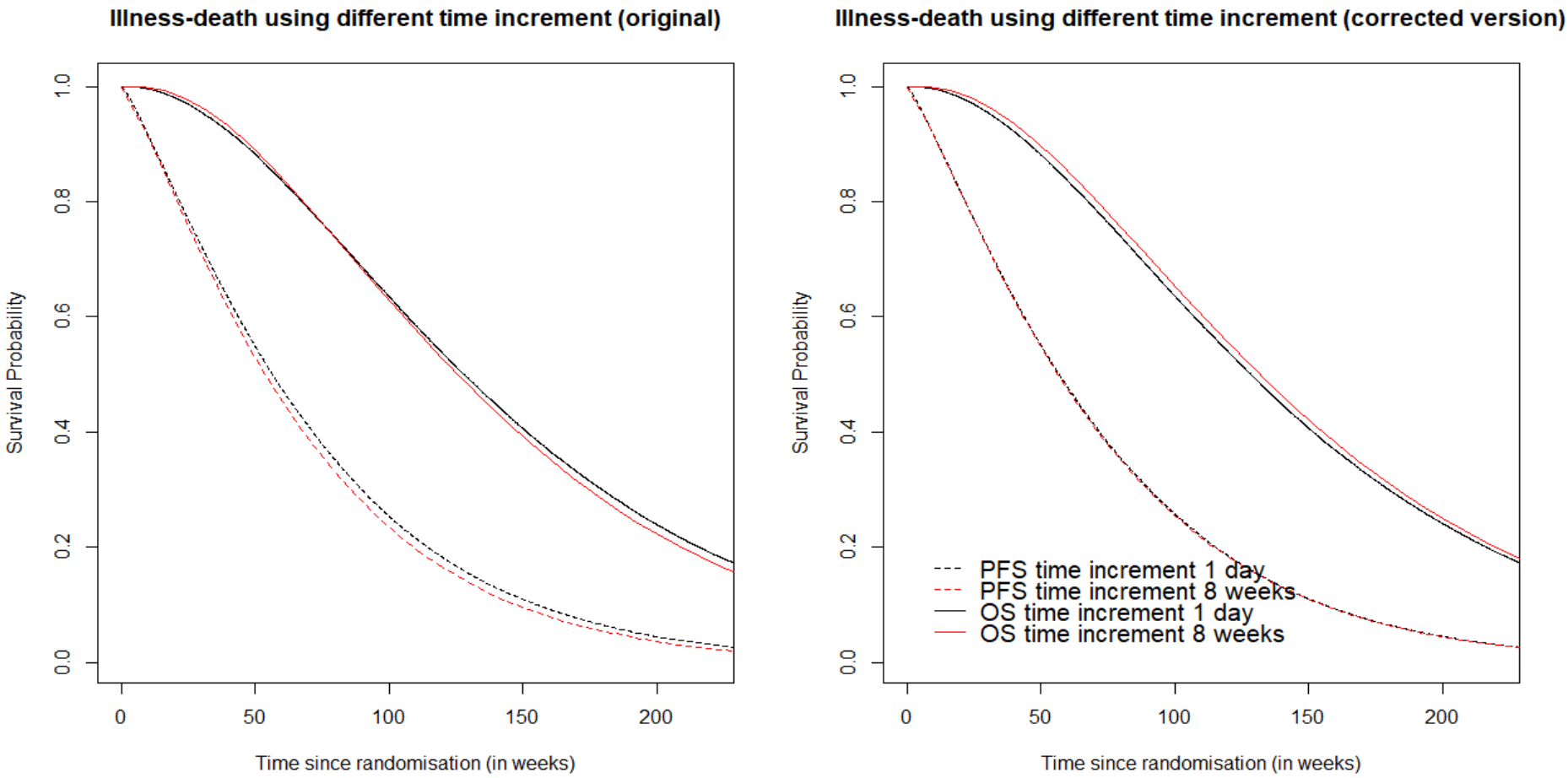
4.4.2.2 *Is time modelled continuously in the* `mssample` *function?*

Williams *et al*[24, 50] describes the MSM as using a continuous-time approach and states that in this type of model "*survival times are treated as continuous variables, rather than being measured in discrete cycles as is usually the case in decision-analytic modelling*". This statement is misleading as the `mssample` function uses direct data on the cumulative hazard over time (discretised), rather than the parameters of the survival function. In other words, the extrapolation for the transitions is made outside of the function with data discretised in time intervals defined by the user. The discretised cumulative hazards are then used in the `mssample` function to sample the time to events using the `Hazsample` function.

Upon reviewing of the R code, as expected, it appears that time is indeed sampled to a discrete value from the defined interval. The function returns a discretised value, rather than any values between cycles. For instance, assuming a time horizon of 1,000 days, with data structured in 10-days intervals (10, 20, 30…1,000), the possible time sampled will be a multiple of 10. In contrast, assuming the same time horizon of 1,000 days, but this time the data are structured in 1-day intervals, the possible time sampled could take the value of any integer between 1 to 1,000. Therefore, the use of the term "continuous time" is potentially misleading as the sampled time cannot take any possible value. Time is therefore treated in a similar way as in standard health economic cohort models, whereby, time is measured in discrete time cycles. However, the `mssample` function uses a simulation approach.

This approach to sampling time (using discretised time intervals) has a number of implications. Whilst predictions will be the same irrespective of the time increment for a single transition (e.g. time to progression), this is no longer true when considering consecutive times (e.g. time to death based on time to progression and time from progression to death). This can be illustrated by considering a simple three-state model, whereby all patients move to progression prior to death (therefore no patients die in the pre-progression health state [PFS=TTP]). Data from the BC dataset on PFS (used as a proxy for TTP) and PPS were used in this example. For the sake of simplicity, all transitions were assumed to be time-varying, based on Weibull distributions. The MSM was run using the `mssample` function (corrected for the inconsistency identified in Section 4.4.2.1), with predictions shown in Figure 17 assuming a time-step of one day and 8 weeks, respectively. For transparency, results are also presented using the original `mssample` function (prior correction).

Figure 17 : Illustration of the impact of using different cycle length on the predictions for the time to events for single and consecutive times



As anticipated, predictions for PFS are identical irrespective of the time interval used (as this is estimated from a single time-to-event outcome) using the corrected version of the `mssample` function. In contrast, predictions for OS (which are comprised of two consecutive times PFS/TTP+PPS) were different depending on the time interval used even using the corrected version (and even more so using the original function). This is because time is treated discretely rather than continuously. Whilst this is not a limitation of the function, this highlights that consideration needs to be taken when selecting the cycle length

4.4.3 Implementing the multi-state model using the `mssample` function in the BC dataset (CALGB 40502) and comparison with the `msm` function when the model is Markov

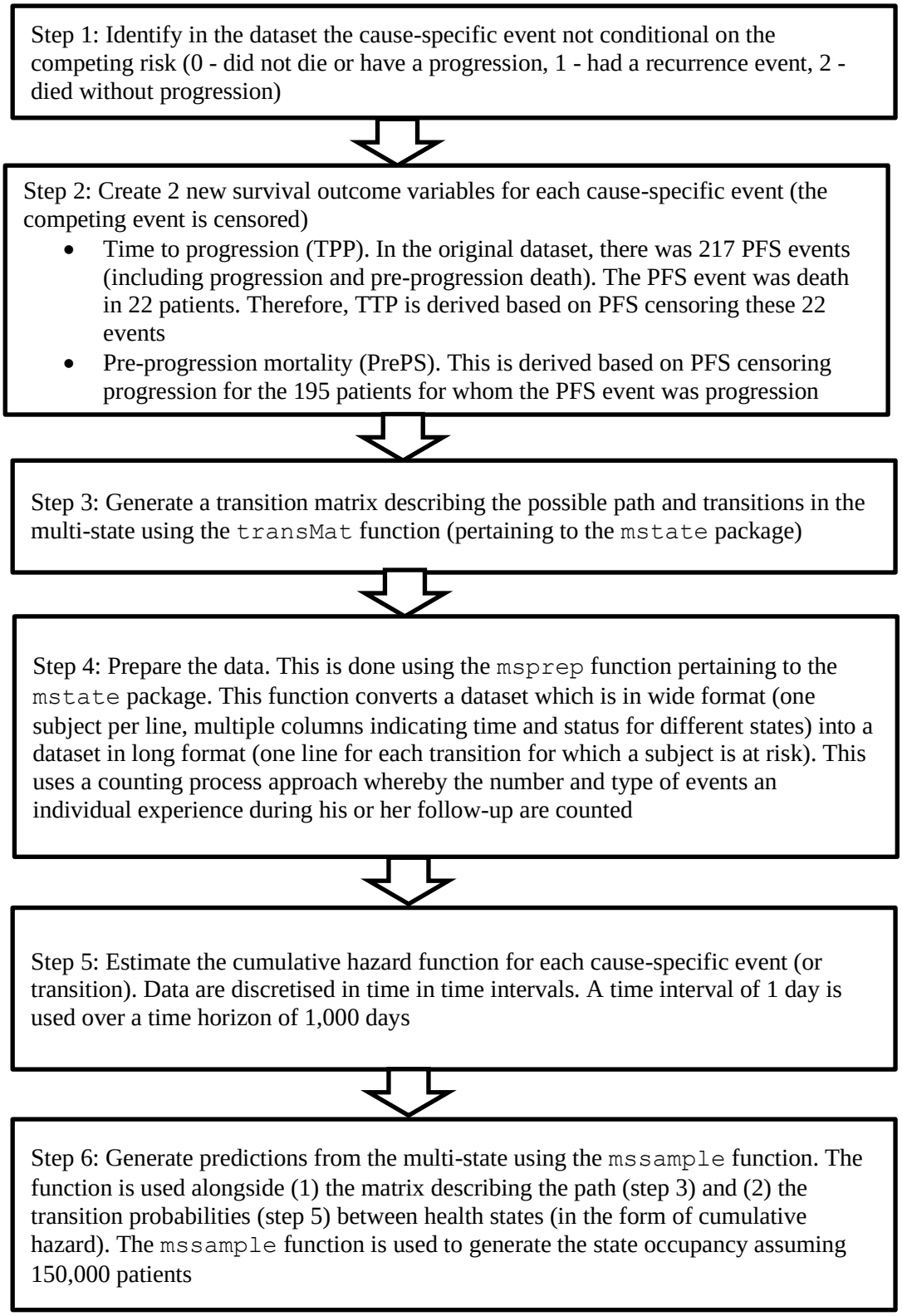
In this section, I describe the implementation of the MSM using the `mssample` function (part of the `mstate` package) in the BC dataset. Given (a) the inconsistency described in Section 4.4.2.1 and (b) the method used for sampling time to events (Section 4.4.2.2), the corrected version of the `mssample` function was used assuming a 1-day cycle length. It should be noted that whilst smaller time increments may reduce any further potential biases, the computational time is increased significantly. In addition, 150,000 patients were sampled to reduce any influence from sampling variation.

The MSM was run assuming that all transitions (TTP, PrePS and PPS) were constant (survival is exponentially distributed) - otherwise known as a Markov model. The survival distributions for each transition were transformed into their cumulative hazard functions and inputted into the `mssample` function. Predictions from the `mssample` function are plotted alongside the predictions from the `msm` function (described previously) as this also assumes constant transitions (exponential distribution).

It should be noted that transitions (TTP, PrePS and PPS) could follow alternative parametric time-varying distributions. The exponential distribution was selected to provide the same distributional assumption as that used in the `msm` function. In brief, the implementation of the MSM using the `mssample` function (part of the `mstate` package) is relatively straightforward and follows six key steps (see Figure 18). It should be noted that the earlier steps described in Figure 18 feed into the `mssample` function which is used to combine the transitions (Step 6 in the figure).

The survival functions for the three transitions together with their associated cumulative hazard functions using the exponential, are presented in Appendix 1. Predictions from the MSM using the `mssample` function (`mstate` package), the `msm` function and STMs are plotted against the KM plots of PFS and OS in Figure 24. As expected, when transitions are constant, predictions from the MSM using the `mssample` function and `msm` function are the same. It should be noted that in this example, the assumption of constant transitions did not fit the data well.

Figure 18 : Step by step implementation of the multi-state using the `mssample` function (part of the `mstate` package) in the BC dataset (CALGB 40502)



4.5 Summary

In this chapter I described the implementation and key assumptions when constructing an MSM. This is a powerful tool which allows for the modelling of transitions under a competing risks framework. Two R packages to perform multi-state modelling were described in this chapter; the `msm` and `mstate` package. As previously described, a function is also available in STATA and uses the same approach as the `mssample` function in R. It should be noted that the STATA function (which I have used separately – not shown here) is more efficient and return predictions more rapidly.

A key strength of the `msm` package is that parameters are estimated endogenously and the model is fitted directly to the data with transitions assumed to be constant. The `msm` package assumes constant event risks and that the model follows a Markov process. This assumption is rarely considered plausible in health economics. However, it is possible to relax this assumption by fitting the model using the `msm` function assuming that transitions are constant within defined time intervals. This can easily be done using the `msm` function using the `pci` option. However, when fitting the MSM and predicting health state occupancy over the patient's lifetime, the long-term extrapolation beyond the observed period of the trial will be based on the constant hazard observed in the last time interval considered – this assumption may not be considered appropriate and could lead to inappropriate extrapolations. In contrast, the `mstate` package allows for the use of any parametric distributions (standard or flexible). However, using this package, parameters for the transitions are estimated exogenously from the MSM and inputs (the cumulative hazard function for each transitions) are combined within the `mssample` function.

In addition to accounting for competing risks under a competing risk framework, a key strength with using the `msm` and `mstate` functions is that these are pre-defined functions which are easy to use, making them transparent and reproducible, which reduces the scope for implementation errors (as described in Chapter 5). Despite the strengths associated with multi-state modelling and the availability of packages in R, the approach is rarely used in health economic state-transition models. Instead, a number of simplifications are typically made in health economics to avoid the explicit modelling of the competing transitions.

In the next Chapter, I highlight the simplifications made in health economic models of anticancer therapies to estimate health state sojourn time using the STM approach and demonstrate how this compares with the MSM approach. I then discuss the key assumptions and implications for the simplifications typically made in health economics when modelling competing transitions between health states.

5 CHAPTER V: WHAT ARE THE IMPLICATIONS FOR THE SIMPLIFICATIONS MADE IN HEALTH ECONOMICS WHEN MODELLING COMPETING TRANSITIONS

5.1 Chapter overview

In this chapter, I describe how the STM approach is currently implemented in health economic models of treatments for advanced/metastatic cancer.

In Section 5.3, I present findings from a review of the implementation of the state transition model in NICE TAs of cancer interventions in the advanced/metastatic setting published in the last 10 years.

In Section 5.4, I describe the key assumptions made in health economics when modelling the competing transitions and highlight how this compares with the MSM.

In Section 5.5, I provide a direct comparison of the STM as currently implemented in health economics and the MSM when transitions are assumed to be constant to illustrate that these approaches provide similar estimates. A comparison when transitions are time-varying is presented in Section 5.6.

In Section 5.8, I draw conclusions regarding the extent to which the simplifications regarding competing transitions made in health economic state transition models are appropriate.

5.2 Introduction

Following a rapid review of STMs used in NICE TAs, I highlight the simplifications that are typically made in current STMs and discuss the key assumptions employed when modelling competing transitions. I also highlight how this approach compares with the MSM approach described in Chapter 4.

5.3 How are competing transitions combined in NICE TAs of anticancer interventions in the advanced/metastatic setting?

This section presents results from a rapid review of STMs used in NICE TAs of cancer interventions in the advanced/metastatic setting published in the last 10 years (May 2009 – May 2019). The review was subsequently updated for transparency and completeness to include an additional year (until May 2020).

5.3.1 Review objectives

A review of the implementation of the PSM is available Woods et al (2016). However, no review is available on how the STM is currently implemented in health economics. Consequently, a rapid review of STMs used in NICE TAs is presented in this thesis. The primary aim of this rapid review is to identify how competing transitions in the STM are modelled in health economic analyses, notably how the transition from TTP (λ_{12}) and PrePS (λ_{13}) are combined.

The secondary objective of the review is to identify the different structural assumptions used within current three-state STMs.

5.3.2 Initial search and selection strategy

The initial search was conducted on the 5th June 2019. For this thesis, I reviewed all NICE TAs of interventions for advanced/metastatic cancer completed or ongoing between May 2009- May 2019 (the last 10 years). Ongoing ("in development") appraisals were also included to ensure that potential alternative approaches were not missed.

The review was limited to NICE appraisals: (a) to be reflective of approaches used within the HTA context; (b) as enough details on the analytical approach and assumptions are typically available from NICE documentation, and (c) to keep the review manageable. Findings may therefore not be fully generalisable as alternative implementations may be available in the broader literature.

Only models reported in company submissions were considered within the review as these form the basis for the economic evaluation in NICE STAs. For NICE Multiple Technology Appraisals (MTAs), only the method reported by the AG was included. This is because only a summary of the company's model is typically published. Inclusion and exclusion criteria for the review are described below.

Inclusion criteria

- Appraisals (completed or ongoing) between May 2009- May 2019
- Modelling of an intermediate endpoint (PFS, TTD, EFS, TTP, RFS) and OS
- Appraisals for an intervention in the advanced/metastatic setting or mixed population including a large proportion of people with advanced disease.
- Progressive models (i.e. patients are not assumed to regress to a better health state)
- Presence of competing transitions, taken from the same source of data.

Exclusion criteria

- Appraisals in conditions other than cancer
- Early/adjuvant setting (when non-metastatic)
- Terminated appraisals or appraisals at scoping stage or prior ACD
- Models in which OS is estimated directly (e.g. using a PSM approach)
- Appraisals where there was insufficient detail (or lack of clarity) regarding how the competing transitions were considered
- Non-progressive models
- Absence of competing risks (for instance, no death in pre-progression)
- Competing risks modelled using data from multiple sources (for instance, PFS taken from one source and the proportion of events which are deaths taken from another source).

5.3.3 Subsequent update

Given the timing of the review, the original search was updated on the 16th June 2020 to include the most recent year (June 2019 – May 2020).

5.3.4 Data extraction and synthesis

A data extraction form was created in Excel with the key characteristics of relevant appraisals (condition, setting) and details of the modelling approach (for instance health states, structural assumptions) extracted.

A simplified version of the extraction from is provided in Appendix 2, summarising the key characteristics of the included appraisals, the approach taken and the assumption used.

Findings are summarised in a narrative form in the main body of this chapter. The methods to model/combine competing transitions are then categorised in terms of key structural assumptions.

5.3.5 Results

Figure 19 shows the results of the search. As of the June 2020, there 439 TAs ongoing/in the development list and 396 NICE TAs published between May 2009 and May 2020. Of these, 29 appraisals met the defined inclusion criteria; 28 in the published list of NICE TAs and 1 appraisal in the NICE development list. The included appraisals are listed in Table 4. The key reasons for exclusion were: non-cancer appraisals or non-advanced (n=182), appraisals at the scoping stage, in progress or terminated (n=473), direct OS modelling (n=135), other (n=16). It should be noted that only one reason for exclusion is recorded here; therefore, a study could also have been excluded for a different reason to the main reason recorded here.

Figure 19: Search results for NICE TAs included in the review

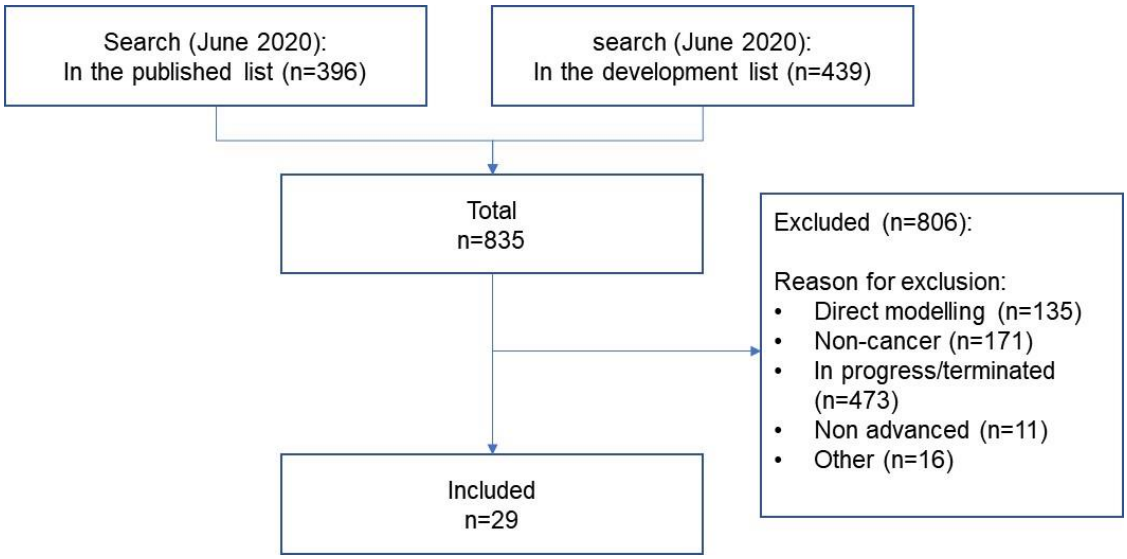


Table 4 : NICE TAs included in the review

Ref	Title	Date published
ID945	Abiraterone for treating newly diagnosed high risk metastatic hormone-naive prostate cancer [ID945]	TBC
TA604	Idelalisib for treating follicular lymphoma refractory to 2 treatments	Oct-19
TA593	Ribociclib with fulvestrant for treating hormone receptor-positive, HER2-negative, advanced breast cancer	Aug-19
TA586	Multiple myeloma - lenalidomide (post bortezomib) (part rev TA171) [ID667]	Jun-19
TA587	Multiple myeloma (newly diagnosed) - lenalidomide [ID474]	Jun-19
TA578	Durvalumab for treating locally advanced unresectable non-small-cell lung cancer after platinum-based chemoradiation	May-19
TA563	Abemaciclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer	Feb-19
TA513	Obinutuzumab for untreated advanced follicular lymphoma	Mar-18
TA502	Ibrutinib for treating relapsed or refractory mantle cell lymphoma	Jan-18
TA496	Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer	Dec-17
TA491	Ibrutinib for treating Waldenström's macroglobulinaemia	Nov-17
TA439	Cetuximab and panitumumab for previously untreated metastatic colorectal cancer	Mar-17
TA472	Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab	Aug-17
TA387	Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated	Apr-16
TA400	Nivolumab in combination with ipilimumab for treating advanced melanoma	Jul-16
TA386	Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis	Mar-16
TA384	Nivolumab for treating advanced (unresectable or metastatic) melanoma	Feb-16
TA380	Panobinostat for treating multiple myeloma after at least 2 previous treatments	Jan-16
TA381	Olaparib for maintenance treatment of relapsed, platinum-sensitive, BRCA mutation-positive ovarian, fallopian tube and peritoneal cancer after response to second-line or	Jan-16
TA370	Bortezomib for previously untreated mantle cell lymphoma	Dec-15
TA343	Obinutuzumab in combination with chlorambucil for untreated chronic lymphocytic leukaemia	Jun-15
TA263	Bevacizumab in combination with capecitabine for the first-line treatment of metastatic breast cancer	Aug-12
TA257	Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses	Jun-12
TA258	Erlotinib for the first-line treatment of locally advanced or metastatic EGFR-TK mutation-positive non-small-cell lung cancer	Jun-12
TA243	Rituximab for the first-line treatment of stage III-IV follicular lymphoma	Jan-12
TA226	Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma	Jun-11
TA214	Bevacizumab in combination with a taxane for the first-line treatment of metastatic breast cancer	Feb-11
TA193	Rituximab for the treatment of relapsed or refractory chronic lymphocytic leukaemia	Jul-10
TA174	Rituximab for the first-line treatment of chronic lymphocytic leukaemia	Jul-09

Consequently, a total of 29 appraisals were included in the review, including 28 TAs from the initial search in the last 10 years (May 2009-May 2019) and one TA in the most recent year (June 2019- May 2020).

Of the 29 appraisals included in this review, 6 were in breast cancer (BC; TAs 214, 257, 263, 496, 563 and 593),[19, 51-55] 5 were in follicular lymphoma (FL; TAs 226, 243, 472, 513 and 604),[20, 56-59] 3 were in multiple myeloma (MM; TAs 380, 586, and 587),[46, 47, 60] 3 were in CLL (TAs 174, 193 and 343)[61-63] 2 were in melanoma (TAs 384 and 400)[64, 65] 2 were in mantle cell lymphoma (MCL; TAs 370 and 502),[66, 67] 2 were in non-small-cell lung cancer (NSCLC; TAs 258 and 578),[68, 69] 2 were in prostate cancer (PC; TAs 387 and ID945)[48, 70] one was in myelofibrosis (MF; TA 386),[71] one was in ovarian cancer (OC; TA 381),[21] one was in colorectal cancer (CC; TA 439),[72] and one was in Waldenström's macroglobulinaemia (WM), TA 491.[73]

Fifteen of the included appraisals were conducted in the advanced/metastatic setting, eight related to a relapsed/refractory population, one was in patients with locally advanced disease only (stage III) with the remaining five appraisals including a mixed population in terms of stage or where it was unclear whether this was the advanced form of the condition.

The majority of models reviewed were centred around 3 key health states (PF, PD and death), with a proportion of those separating the time on and off treatment.

Evidence on the time to progression or death (PFS) and PPS was taken (mostly) from the same source (key trial) in 20 appraisals. Separate sources were used for PFS and PPS in 9 appraisals. This was because a pathway was typically modelled. In one appraisal (TA513[59]), the model was based primarily on the key trial, supplemented by external evidence.

5.3.5.1 Terminology used

Terminologies used to describe the modelling approach were sometimes absent, inconsistent or inappropriate, with the Markov terminology often used inappropriately, despite the model not being 'true' Markov. A range of terminologies were used including; Markov model, semi-Markov model, and (Markov) state-transition model.

5.3.5.2 Justification for the choice of approach

The key justification provided for the modelling approach compared with the PSM (or an alternative approach) focused on the modelling of the treatment pathway/natural history or use of external evidence in 17 appraisals. Fourteen appraisals (ID945,[48] TA593,[55] TA578,[68] TA563,[53] TA502,[67] TA496,[54]

TA491,[73] TA472,[20] TA400,[64] TA384,[65] TA381,[21] TA370,[66] TA343,[61] TA257[19]) made explicit reference to the immaturity of the data for justifying using an STM compared with a PSM approach. Four appraisals justified the use of the STM given the structural relationship between PFS and OS (TA586,[46] TA587,[47] TA578[68] and TA563[53]). One appraisal, an MTA, justified the use of the STM approach by the difficulty in representing the hazard of death using a parametric distribution (TA257[19]). Seven appraisals (TA604,[56] TA214,[52] TA193,[62] TA174,[63] TA513,[59] TA263[51] and TA258[69]) did not provide any clear justification for the choice of modelling approach, with the exception of mentioning that the selected approach is commonly used in oncology.

The majority of the STMs included in this review used a cohort approach, with 5 models using an individual based-simulation approach (TA593,[55] TA496,[54] TA387,[70] TA386,[71] TA243[57]) and an additional 3 using the `msm` package (ID945,[48] TA586,[46] TA587[47]). In the majority of cases, the intermediate endpoint was PFS. Two models used time to treatment discontinuation (TTD)[70, 71] with an additional appraisal using first subsequent treatment (FST).[21]

5.3.5.3 Modelling competing transitions

The majority of models reviewed (n=21) modelled a combined endpoint (typically PFS) directly as a single composite endpoint, accounting for the two competing events. In other words, taking PFS as an example, instead of explicitly modelling the transitions from progression-free to progressed disease and from progression-free to death (without progression), the majority of reviewed models used PFS directly and extrapolated this endpoint beyond the trial duration as a single outcome. PFS was then divided between the two competing events using a set of structural assumptions.

Three of the models reviewed (TA604,[56] TA400,[64] TA384,[65]) modelled the two competing risks separately: TTP - where deaths prior to progression were censored and; PrePS - whereby progression events other than death were censored. However, it is unclear from the model descriptions provided in the company submissions (CS) or the ERG reports how the two competing risks were combined and whether this was done appropriately. An additional model used TTP and estimated a constant rate of death from a negative binomial model (TA563[53]). Three appraisals used the `msm` package in R and therefore dealt with competing risks (ID945,[48] TA586,[46] TA587[47]). One appraisal used a discrete event simulation approach and compared times to the next event (TA387[70]).

Amongst the 21 models that modelled a combined endpoint (mostly PFS) as a single endpoint, three broad approaches were employed to separate the two competing events (using PFS as an example);

- **Approach 1 (n=8):** Use of PFS to model the combined transitions for people leaving the PF state, assuming that a proportion of PFS events are death or progression. The proportion was assumed to be constant (calculated as number of pre-progression deaths divided by total number of PFS events) in 5 appraisals (TA593,[55] TA496,[54] TA386,[71] TA381,[21] TA243[57]). In a further two appraisals (TA439,[72] TA257,[19]) this approach appeared to have been employed, but this was not fully clear. The proportion of pre-progression deaths was assumed to be time-varying in one appraisal (TA380[60]) whereby a logistic regression model using the log of time in PFS was used as a covariate.
- **Approach 2 (n=11):** Use of PFS to model the combined transitions for people leaving the PF state, with a constant probability of dying in PFS based on the rate of death in PFS (TA513,[59] TA502,[67] TA491,[73] TA472,[20] TA263,[51] TA258,[69] TA226,[58] TA214,[52] TA193,[62] TA174[63]) calculated as the number of deaths in PFS divided by the person-years in PFS (total PFS time; observation and censored time). In one appraisal (TA343[61]) it was unclear how the constant probability of dying whilst progression-free was calculated.
- **Approach 3 (n=2):** Use of PFS to model the combined transitions for people leaving the PF state; with
    - Approach 3a: the probability of dying in PFS at each cycle based on the pre-progression survival (PrePS) curve in one appraisal (TA370[66]). A parametric function was fitted to prePS data (PFS censored for progression) to obtain the probability of dying in PFS in each cycle.
    - Approach 3b: the probability of remaining in PFS in each cycle based on TTP in another appraisal (TA578[68]). A parametric function was fitted to the KM for TTP to obtain the probability of remaining in PFS at each cycle. In this appraisal, the company stated that "the key assumption made was that the TTP distribution was set to the same as for PFS". This assumption is likely to be necessary to limit inconsistencies in the shape between PFS and TTP.

5.3.5.4 Use of multi-state modelling

Three appraisals (ID945,[48] TA586,[46] TA587[47]) used a multi-state approach using the `msm` function described in Chapter 4.3. The observed time-to-event data were used initially, followed by an MSM in ID945[48] and TA587[47]. The `msm` was fitted to data for three time intervals in TA586[46] in order to account for time-varying probabilities. No appraisals used the `mstate/mssample` functions described in Chapter 4.4.

5.3.5.5 Use of post-progression survival and assumptions about the intervention

The majority of appraisals reviewed used a time-varying distribution to model PFS (or related intermediate endpoints) or to model the transition between progression and death (PPS). Seven distributions were generally considered: the exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma functions.

Appraisals which allowed time-varying transition in PPS used tunnel states (a series of states that can be occupied for only one cycle) in all cohort models. Tunnel states allow the probability to be different for each tunnel state, allowing transition probabilities to change depending on history. Tunnel states were not required for the individual patient-level simulation models. The use of an exponential distribution for PPS was justified in 5 appraisals (TA502,[67] TA380,[60] TA343,[61] TA258,[69] TA226,[58]) to avoid overcomplicating the model by the addition of tunnel states.

For the 15 appraisals which primarily used information from the key pivotal trial (i.e. the model is based on the trial for the intervention against a relevant comparator with little reference to external evidence), PPS was pooled across arms in 10 appraisals (TA593,[55] TA578,[68] TA513,[59] TA472,[20] TA370,[66] TA257,[19] TA214,[52] TA193,[62] TA174,[63] TA380[60]). The key justification for pooling PPS was the absence of significant or visual differences in the data between each arms. A different PPS function was used between the control and intervention arms in 6 appraisals (TA586,[46] TA587,[47] TA386,[71] TA381,[21] TA263,[51] TA257[19]) based primarily on evidence from the key trial. One appraisal (TA257) presented results from two models for different comparisons; one comparison where PPS was pooled and one comparison where PPS was not pooled. The same PPS was further assumed between treatments in a further 3 appraisals (TA502,[67] TA258,[69] TA604[56]) because the key trial was either a single trial arms study or had a different comparator to the one assessed in the pivotal trial.

In all of the included models in this review, PPS was based on survival models fitted directly to the trial data, without any adjustment (with the exception of general population mortality constraints). In other words, none of the included models estimated the time to death following progression conditional on the time to progression. The only models which recognised that PPS estimated from the trial only in the subset of patients who progressed may not be generalisable to the overall randomised population was in TA513[59] and TA380[60]. In TA513,[59] the company modelled early and late progressors separately, using information from the trial to represent early progressors (the first two years) and external evidence to represent late progressors (beyond two years). Similarly, in TA380, PPS (following lenalidomide) was separated between early and late progression.

5.3.5.6 Other assumptions

With the exception of the MSMs that used the `msm` function, transitions were typically estimated separately from each other (no common parameters), and therefore, there is an assumption of conditional independence (parameters for the transitions are estimated independently from each other).

In one appraisal (TA563[53]), transitions after progression were not explicitly modelled. Instead, a fixed pay-off (calculated from external data) that represented the outcomes in PPS (costs and QALYs) was applied at the point of progression. This assumes perfect surrogacy between delta PFS and delta OS.

Finally, the majority of models included background mortality constraints to ensure that the modelled mortality hazard for people with the disease does not fall below that for people without the disease. Such assumptions did not impact on the implementation in the model. Background mortality was typically implemented by taking the maximum between the probability of death for the general population and the probability of death from hazard from the trial (for logical consistency).

5.3.5.7 Key findings

This rapid review highlights some variation with respect to how competing transitions are considered in NICE TAs for anticancer therapies in the advanced/metastatic setting. The majority of economic models included in this review used a cohort approach and modelled the intermediate endpoint (PFS or an alternative related endpoint) as a single composite endpoint, accounting for the two competing events. Consequently, instead of explicitly modelling the two competing transitions (progression-free to progression and progression-free to death) as is done in the MSM approach described in Chapter 4, the cumulative incidence of events (PFS) was modelled instead for people leaving the progression-free health state, with a set of structural assumptions being applied to subsequently separate progression events from pre-progression mortality events.

Three appraisals used multi-state modelling using the `msm` function and fitted the model for part of the trial period, or for different time intervals, to account for time-varying hazards. Most of these appraisals were recent submissions/re-submissions. The review did not identify any appraisals which used the `mstate` (`mssample`) function. Three broad structural assumptions were used to separate the two competing transitions from the cumulative incidence of events (PFS or related endpoints) in people leaving the PF health state: (i) assuming a proportion of PFS events are deaths (which could be constant or time-varying); (ii) assuming a constant probability of dying in PFS calculated from the number of deaths divided by total person-PFS time, or (iii) assuming a probability of dying in PFS or remaining in PFS based on the PrePS or TTP curve.

This rapid review highlights that compared with the MSM approach described in Chapter 4, simplifications are typically made in health economic models in that not all transitions are explicitly modelled under a competing risk framework and that there are some inconsistencies in NICE TAs in terms of how competing transitions are considered.

In Section 5.4, I discuss the key assumptions and simplifications made in health economics to model the competing transitions and whether these have the propensity to alter decision-making compared with the formal multi-state framework.

5.4 What are the key assumptions commonly made in health economics when modelling the competing transitions?

The multi-state approach described in Chapter 4 is rarely used in health economic evaluation. As noted above, the `msm` function was used in three recent appraisals (ID945,[48] TA586,[46] TA587[47]). The key assumption is that transitions are constant within time intervals. The `mstate/mssample` function, which uses parametric extrapolation for each transition and combines them under a competing risk framework, has not been used in any of the appraisals included in this review. However, such functions could have been used in more recent appraisals or in expanded literature not covered by this review. As described in Section 5.3, the STM is typically implemented using PFS directly (as a composite endpoint for the two competing transitions) combined with structural assumptions on how to separate death events from progression events.

This section focusses on the two key structural assumptions made within the Simplified STM approach used in in health economic models and compares them against the structural assumptions made within the multi-state framework.

5.4.1 Assumptions regarding the estimation of the transition for leaving the progression-free health state (combined transition)

As highlighted in Section 5.3, a key difference between the MSM and the Simplified STM applied in health economics relates to how the cumulative incidence of event (PFS) is estimated:

- In the MSM, the cumulative incidence of events for leaving the PF health state is estimated as a function of the two competing events/transitions (TTP and PrePS) under a competing risk framework. Therefore, the probability of leaving the PF health state is given by two transitions and is therefore estimated indirectly from the PFS data.

- In contrast, in the Simplified STM applied in health economic models, the cumulative incidence of events for leaving the PF health state is estimated directly from the PFS curve fitted to the trial data. Therefore, the probability of leaving the PF health state is given by the PFS data directly (a single transition).

As described in Section 3.10, standard approaches and approaches that properly deal with competing risks are similar when analysing a combined composite endpoint such as PFS using a non-parametric approach. However, whilst this may be true from a non-parametric point of view during the observed period of a study, this relationship may no longer hold when:

- the hazard is determined parametrically during the observed period, and when,
- extrapolating beyond the observed period.

This is because during and beyond the observed period, the hazard is determined from a parametric function and fit to the data and no longer from the data itself directly. The only exception is when the rate of events is constant (i.e. survival is exponentially distributed). In such cases, using PFS directly to represent the cumulative incidence of events is the same as modelling the two competing transitions separately. This is because rates are additive and constant throughout the model duration.

In contrast, when the rate of events (for either PFS or competing transitions) is assumed to be time-varying, using PFS directly may lead to differences in predictions compared with modelling the two competing transitions under a competing risk framework. This is because the fit and extrapolation for PFS estimated from a single dataset (PFS) is not the same as the fit and extrapolation from two separate datasets (TTP and prePS) which are then combined. These potential differences will be reduced when the degree of extrapolation is minimal and when the fit of the model to the observed data is good. In contrast, larger differences will be expected when the parametric function provides a poor fit to the observed period or when the need for extrapolation beyond the observed period is greater due to the increase in uncertainty around of the shape of the extrapolation (as PFS will be a combination of the extrapolated hazards for two separate events).

Despite difficulty to provide a comparison, to illustrate this further, the PFS direct fit is compared against the estimated PFS under a competing risk framework in this Section. In addition to the BC dataset described in Chapter 3.2, data for the comparator arm from a trial in prostate cancer and lung cancer were obtained from the Project Data Sphere.[32, 33, 74] The prostate cancer trial included 470 patients, of whom 401 had a recorded PFS event (340 progression events and 61 deaths prior to progression).[32] The lung cancer trial included 479 patients, of whom 440 had a recorded PFS event (313 progression events and 127 deaths prior to progression).[33]

Eight parametric distributions (exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma and a restricted cubic spline (RCS) hazard model with one knot) were considered for each survival outcomes (PFS, TTP and PrePS). The fit for each parametric distribution for PFS, TTP and PrePS in the Lung and Prostate cancer dataset is shown in Appendix 3.

The estimated mean PFS using the direct PFS fit and estimated by combining TTP and PrePS under a competing risk framework is shown in Table 5 for the combination of transitions for each of the three separate datasets (breast , prostate and lung).

Table 5 : Estimation of mean PFS using direct PFS fit and combination of TTP and PrePS in the Breast, Lung and Prostate datasets

The values in bold (first row for each dataset) represent the mean PFS predicted using direct PFS (for instance, the mean PFS using a directly fitted Gompertz model for the BC dataset is 67.21 weeks). The next rows show PFS estimated as combination of the distribution selected for TTP (column) and presPS (row) under a competing risk framework, giving 64 possible combinations for each dataset. For example, in the BC dataset, mean PFS generated assuming TTP follows a log-normal distribution and PrePS follows a gamma distribution is 84.00 weeks.

	Distribution for PFS/TTP								
Breast cancer dataset (mean PFS in weeks)									
		exp	weibull	gompertz	lnorm	llogis	gamma	gengamma	spline
	PFS	70.18	66.55	67.21	101.42	84.78	67.02	67.76	75.17
Distribution for PrePS	exp	70.18	67.02	68.69	93.2	78.58	67.34	68.49	72.46
	weibull	68.57	66.46	67.6	81.93	72.36	66.67	67.47	68.65
	gompertz	67.4	66.05	66.84	75.88	68.99	66.16	66.69	66.4
	lnorm	69.77	66.92	68.41	91.65	77.82	67.22	68.28	71.98
	llogis	69.02	66.63	67.89	86.28	74.8	66.87	67.78	70.13
	gamma	68.85	66.56	67.78	84	73.5	66.8	67.66	69.37
	gengamma	67.97	66.26	67.23	78.11	70.26	66.42	67.08	67.28
	spline	68.87	66.56	67.78	85.33	74.26	66.8	67.67	69.79
Prostate cancer dataset (mean PFS in weeks)									
	PFS	44.95	44.09	43.69	69.31	67.03	44.47	43.63	50.64
Distribution for PrePS	exp	44.95	44.45	45.04	58.05	53.85	44.64	44.3	47.01
	weibull	44.34	44.01	44.41	52.61	49.92	44.15	43.9	45.35
	gompertz	43.38	43.27	43.4	47.1	45.95	43.33	43.22	43.39
	lnorm	45.28	44.74	45.39	62.2	57.13	44.95	44.58	48.27
	llogis	44.86	44.43	44.94	57.45	53.57	44.6	44.3	46.88
	gamma	44.54	44.16	44.62	54.18	51.06	44.32	44.05	45.86
	gengamma	43.61	43.42	43.64	48.37	46.84	43.51	43.35	43.81
	spline	44.39	44.06	44.45	53.77	50.84	44.2	43.96	45.7
Lung cancer dataset (mean PFS in weeks)									
	PFS	29.69	29.19	29.29	34.07	35.85	29.21	29.21	30.07
Distribution for PrePS	exp	29.69	29.26	29.31	29.36	28.82	29.02	29.38	29.38
	weibull	29.75	29.24	29.33	29.36	28.83	29	29.39	29.41
	gompertz	29.53	29.33	29.29	29.32	28.76	29.1	29.34	29.24
	lnorm	30.67	29.37	29.8	29.82	29.42	29.06	29.86	30.39
	llogis	30.39	29.46	29.72	29.77	29.33	29.18	29.8	30.13
	gamma	29.77	29.23	29.34	29.36	28.82	28.98	29.38	29.41
	gengamma	29.51	29.26	29.25	29.24	28.66	29.01	29.26	29.15
	spline	29.83	29.36	29.4	29.49	29.01	29.12	29.52	29.64

As expected, the estimated mean PFS is exactly the same when the exponential distribution is used for the direct PFS fit and when the exponential distribution is used for both TTP and PrePS and combined under a competing risk framework. This was the case in all three datasets (breast, prostate and lung cancer).

When looking at each individual dataset, the following could be noted:

- **Breast and prostate cancer dataset**: The log-normal, log-logistic and spline models had a longer tails compared with other distributions for both PFS and TTP, and thus led to higher estimates of TTP and therefore higher PFS. Because of the longer tail, for these distributions, the reliance on extrapolation beyond the observed period was much greater and therefore the extrapolation for PrePS has a more significant impact on the estimation of PFS, as shown in Table 5. While the same TTP is used, different prePS lead to larger variation for PFS.
- **Lung cancer dataset**: In this dataset, whilst the log-normal and log-logistic distributions had longer tails compared with other distributions, the tail remained minimal. PrePS was relatively similar between models up to week 75, by that time, most patients had progressed, limiting any impact from the use of different PrePS functions.

Although it is difficult to provide a like-for-like comparison, it can be seen from this simple illustration that when the fit to the transition is poor or the degree of extrapolation required is greater, there is more scope for the differences between the direct fit and competing risk approach. If transitions were estimated using a complete dataset with no censoring, no differences are expected in the estimation of PFS using the direct fit or combining TTP and PrePS. However, this is not possible in practice as data are incomplete (due to censoring). Consequently, differences in the estimation of PFS between these two approaches are the results of the selection of the appropriate survival function, rather than the approach itself.

5.4.2 Assumptions on the estimation on the separation of patients who experienced progression or death

Because PFS is used directly in the simplified STM (commonly used in HE) to represent the transitions for people leaving the PF health state, assumptions are required to separate progression to pre-progression death events. As highlighted in Section 5.3, three broad structural assumptions are used in the Simplified STM to separate PFS in terms of progression and pre-progression death events. These key structural assumptions are described in detail below. Whilst the number of PFS event will be the same irrespective of the approach taken, the estimated number of progression and pre-progression mortality events differs according to the approach used.

5.4.2.1 Approach 1: The Simplified STM assuming a proportion of PFS events to be death or progression

Perhaps, the simplest approach to separate PFS into the two competing events (progression and pre-progression mortality) is to consider that a proportion of PFS events are (a) progression or (b) death.

The probability that the event is either progression or death could be:

- Constant: calculated from the number of deaths in PFS divided by the total number PFS events
- Time-varying: calculated from a logistic regression model using PFS time as a covariate (on the log scale) or alternative forms.

The implementation is relatively straightforward. The number of patients leaving the PF health sate is first calculated at each cycle (based on the difference in PFS between time intervals), with a proportion of people leaving this health state assumed to be because of death.

The key strength of this approach is its simplicity. However, it is associated with two limitations:

(i) Assuming the probability that PFS events are deaths is constant is an over-simplification and is only appropriate when the model is Markov (i.e. when all transitions - PFS and the two competing transitions - are constant),

(ii) Whilst the probability of an event being death or progression could be assumed to be time-varying, through the use of a logistic regression model (which uses log of PFS time as covariate), such an approach is also likely to be biased for two reasons. First, the shape that the logistic model can take is restricted. Perhaps more importantly, the logistic model does not account for censoring, and therefore any estimates from the logistic regression model are likely to be different from data after accounting for competing risks.

5.4.2.2 Approach 2: The Simplified STM assuming a constant probability of dying in PFS based on the number of events divided by the total PFS time

The second approach uses a different set of structural assumptions in that the number of deaths in pre-progression at a given cycle is calculated from the number of patients who were progression-free in the previous cycle and a constant probability of dying. This constant probability of dying in PFS is calculated based on the number of deaths in PFS divided by the total PFS time (sum of PFS time, including censored observations and events). The number of progression events is then calculated from the number of people who progress during a given cycle minus the number of pre-progression deaths calculated above.

Compared with Approach 1 (outlined above), this approach is perhaps more consistent with the competing risk framework whereby PFS represents the cumulative incidence of the two competing events. Within this framework, the probability of death is calculated as the cumulative incidence of any event at time *t* multiplied by the cause-specific hazard for each event. However, such an approach is restrictive and only appropriate if the rate of pre-progression mortality is constant over time. Furthermore, this is different to properly dealing with competing risks, as PrePS does not affect PFS with this approach; instead it only influences the contribution of progression and pre-progression death events.

5.4.2.3 Approach 3: The Simplified STM assuming a probability of dying based on the pre-progression survival curve (PrePS) or remaining based on TTP

The third approach follows the same principle as the second approach in that the number of deaths in the PF state is calculated from the number of patients who were progression-free in the previous cycle and the probability of dying in PFS. However, compared with Approach 2, which assumes a constant probability of dying in PFS based on the number of deaths in PFS and total PFS time, Approach 3 uses the hazard from the pre-progression survival function (PrePS). PrePS is the "naïve" cause-specific incidence of pre-progression mortality prior to accounting for the existence of the competing event defined as the time to pre-progression mortality, with progression events occurring prior to death being censored.

Typically, PrePS is extrapolated beyond the trial using parametric functions. The probability of dying is then given by the PrePS hazard during each cycle. If PrePS follows an exponential distribution, the hazard of death prior to progression will be constant. If PrePS follows any other distribution, the hazard of dying whilst progression-free will be time-varying. It should be noted that if the exponential distribution is used and the hazard is constant, this is the same as calculating the hazard based on the number of death events in PFS and total PFS time under a Poisson distribution. Consequently, under the assumption of an exponential distribution for PrePS, the hazard of dying in PFS will be the same for Approaches 2 and 3.

As with Approach 2, the implementation is more consistent with the competing risk framework whereby PFS represents the cumulative incidence of the two competing events. However, compared with Approach 2, using PrePS to define the probability of dying whilst progression-free is more flexible and allows a better reflection of the data allowing the hazard to be time-varying. Nevertheless, as previously highlighted, this is different to properly dealing with competing risks, as PrePS does not affect PFS with this approach, but only the contribution of progression and pre-progression death events.

Furthermore, there is an important limitation with this approach in that PrePS and PFS are extrapolated independently from each other despite being correlated (PrePS events are a component of PFS events). Therefore, it is possible for PrePS to be mis-specified in that PrePS becomes greater than PFS (which is not plausible as PFS includes these deaths). Whilst a constraint could be added to a model to prevent the number of deaths in pre-progression from being greater than the number of PFS events, such constraints introduce biases compared with modelling the two competing transitions separately under a competing risk framework.

Similar to this approach, instead of using PrePS to determine the probability of dying whilst progression-free, an alternative approach was identified in the review whereby TTP is used to determine the probability of remaining alive in PFS (TA578[68]). This follows the same principle as above (approach 3), but uses TTP instead of PrePS to estimate the probability of remaining in the progression-free health state. However, using TTP instead of PrePS required additional strong assumptions, such that the same distribution had to be used for PFS and TTP to ensure consistency. This is approach is less likely to be appropriate and therefore is not discussed further within this thesis.

5.4.2.4 Estimation of the percentage of death in patients without progression using the different methods in the prostate cancer dataset

As highlighted in the previous section, different approaches are currently used to separate progression and pre-progression deaths from PFS. These different methods could have an impact on the estimation of OS depending on both the extrapolation for PFS and the number of pre-progression mortality events.

To illustrate this, data from the breast, prostate and lung cancer datasets were used to estimate the number of pre-progression mortality events using the following methods described above:

- Scenario 1: Assuming a proportion of PFS events are deaths (either constant or time-varying)
- Scenario 2: Using PrePS to estimate the probability of dying in PFS. Eight parametric distributions are considered (exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma, spline hazard model with one knot). As previously highlighted, the scenario using the exponential distribution is the same as using the number of deaths in PFS divided by the total PFS time.

For illustration, PFS is assumed to follow a log-normal distribution. This was selected because the log-normal distribution was generally associated with a tail in the datasets examined, and therefore any effect associated with the method to estimate the number of pre-progression death is likely to be clearer. Results in terms of estimated percentage number of pre-progression deaths over time for the breast, prostate and lung cancer datasets are presented in Figure 20, Figure 21 and Figure 22, respectively.

Figure 20 : Comparison in the estimation of the number of death not associated with progression using different simplified approaches used in health economics in the BC dataset (CALGB 40502)

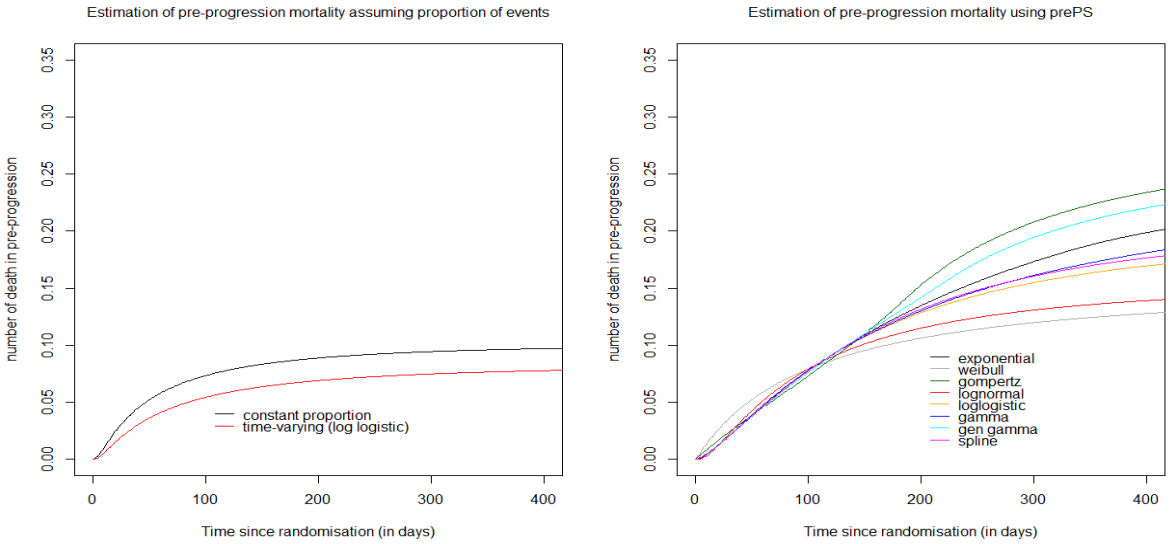


Figure 21 : Comparison in the estimation of the number of death not associated with progression using different simplified approaches used in health economics in the prostate cancer dataset

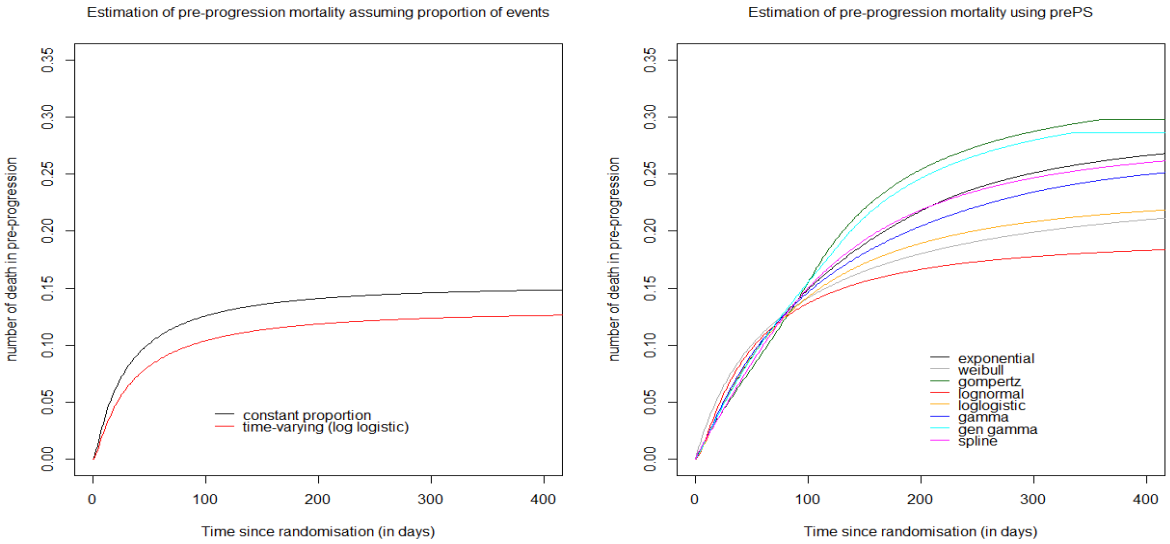
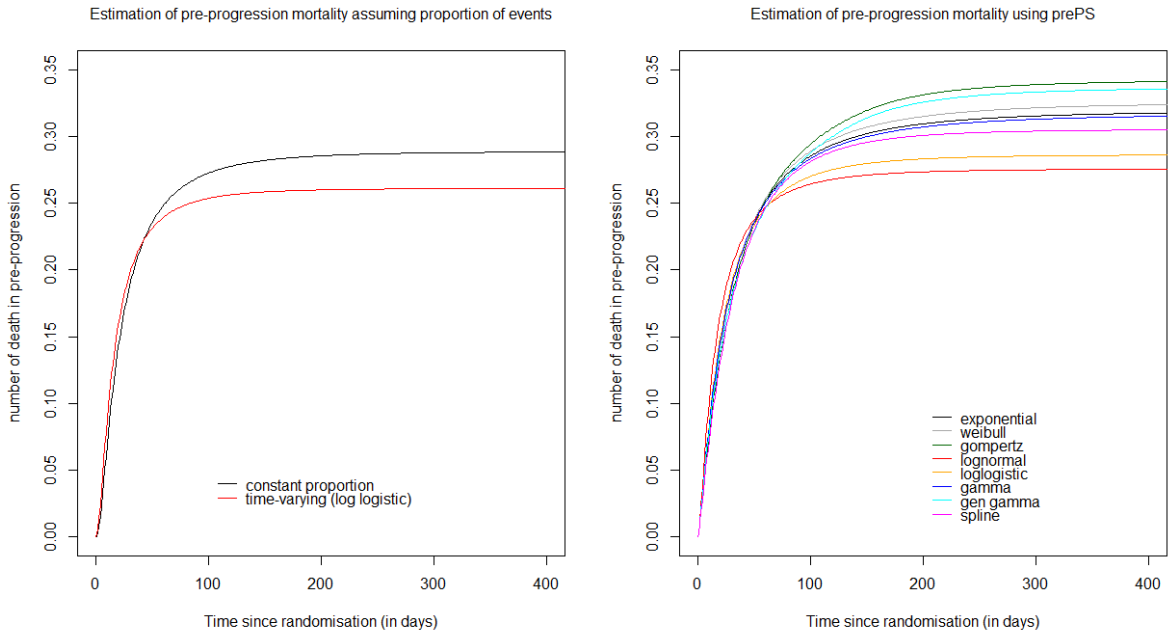


Figure 22 : Comparison in the estimation of the percentage of death not associated with progression using different simplified approaches used in health economics in the lung cancer dataset



In summary, it can be seen that different approaches could predict very different percentage of pre-progression deaths. Assuming a constant or time-varying proportion of PFS events are deaths led to different estimates (Scenario 1 – right-hand side of each figure). Similarly, for Scenario 2 (left-hand side of each figure), the percentage of estimated death events in people who are progression-free was very different depending the choice of parametric function used for PrePS. These differences would have a knock-on impact on the OS estimation.

5.5 Implementation of the Simplified STM (used in health economics) in the BC dataset and comparison against the MSM using the `mssample` function when transitions are constant (Markov)

Let us consider again the BC dataset, whereby the outcomes are PFS and OS. A Simplified STM was constructed under each of the three different structural assumptions described above and these were compared against the MSM (using the `mssample` function).

For simplicity, a cohort approach is used. It is expected that the choice between the cohort or simulation approach would not affect conclusions as both approaches would provide very similar predictions (with differences attributable to sampling variation) under the same assumptions. In addition, in order to reflect the implementation of the Simplified STM, transitions are assumed to be conditionally independent i.e. parameters for the transitions are estimated independently from each other. In order to provide a like-for-like comparison against the MSM, all transitions rates are assumed to be constant (Markov). A comparison against the MSM when transitions are time-varying is presented in Section 5.6. The Simplified STM was implemented using R. The key steps in implementing the Simplified STM are summarised in Figure 23. A time horizon of 520 weeks was used assuming a cycle length (time interval) of 1 day.

As expected, predictions were the same (Figure 24) between all the different implementations of the Simplified STM and the MSM using the `msm` or `mstate` package when the model is assumed to be Markov (transitions assumed to be constant). Curves are superimposed, and therefore other lines are obscured.

Figure 23: Summary of the step-by-step implementation of the Simplified STM (based on the approach commonly used in health economic models)

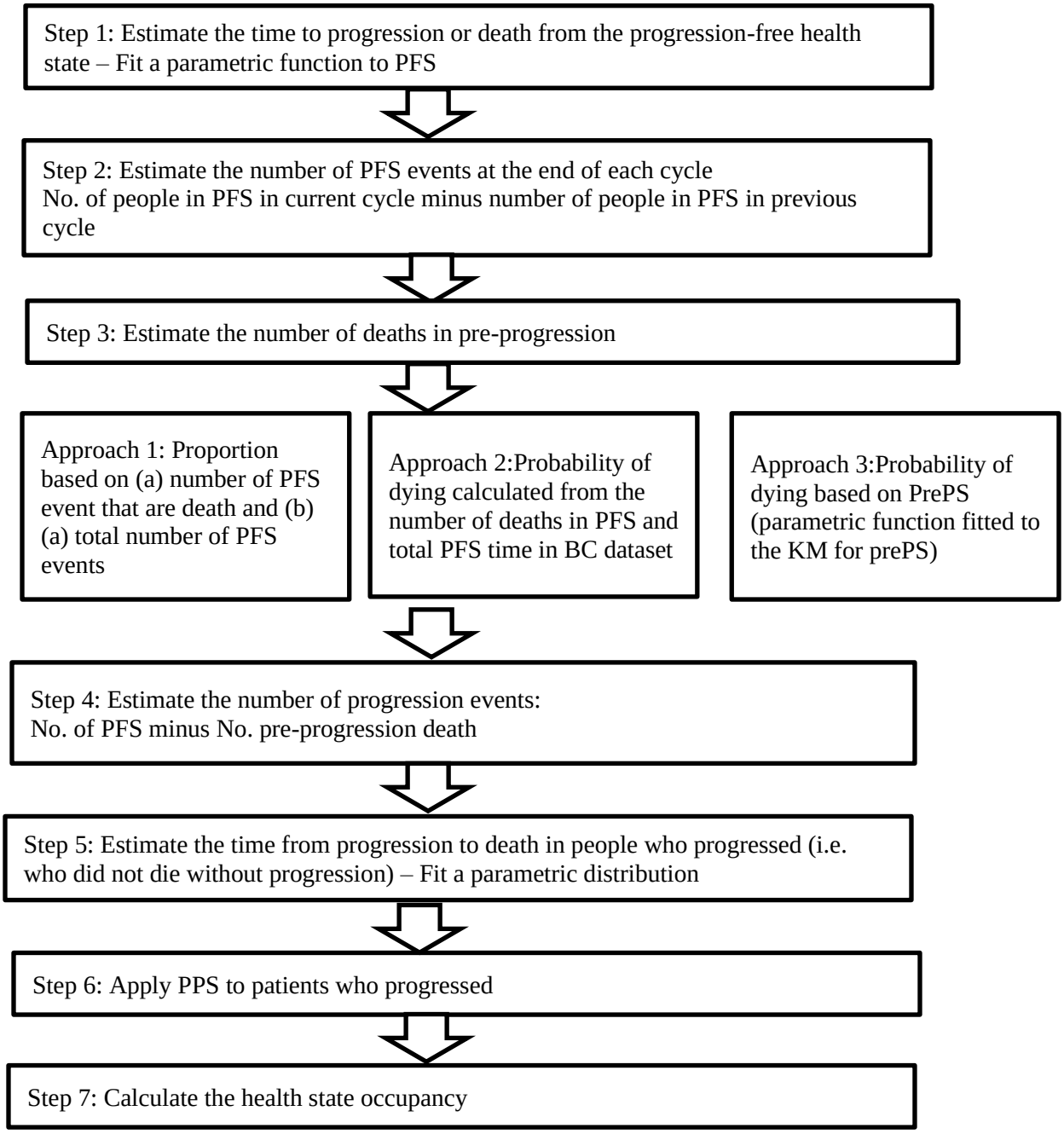
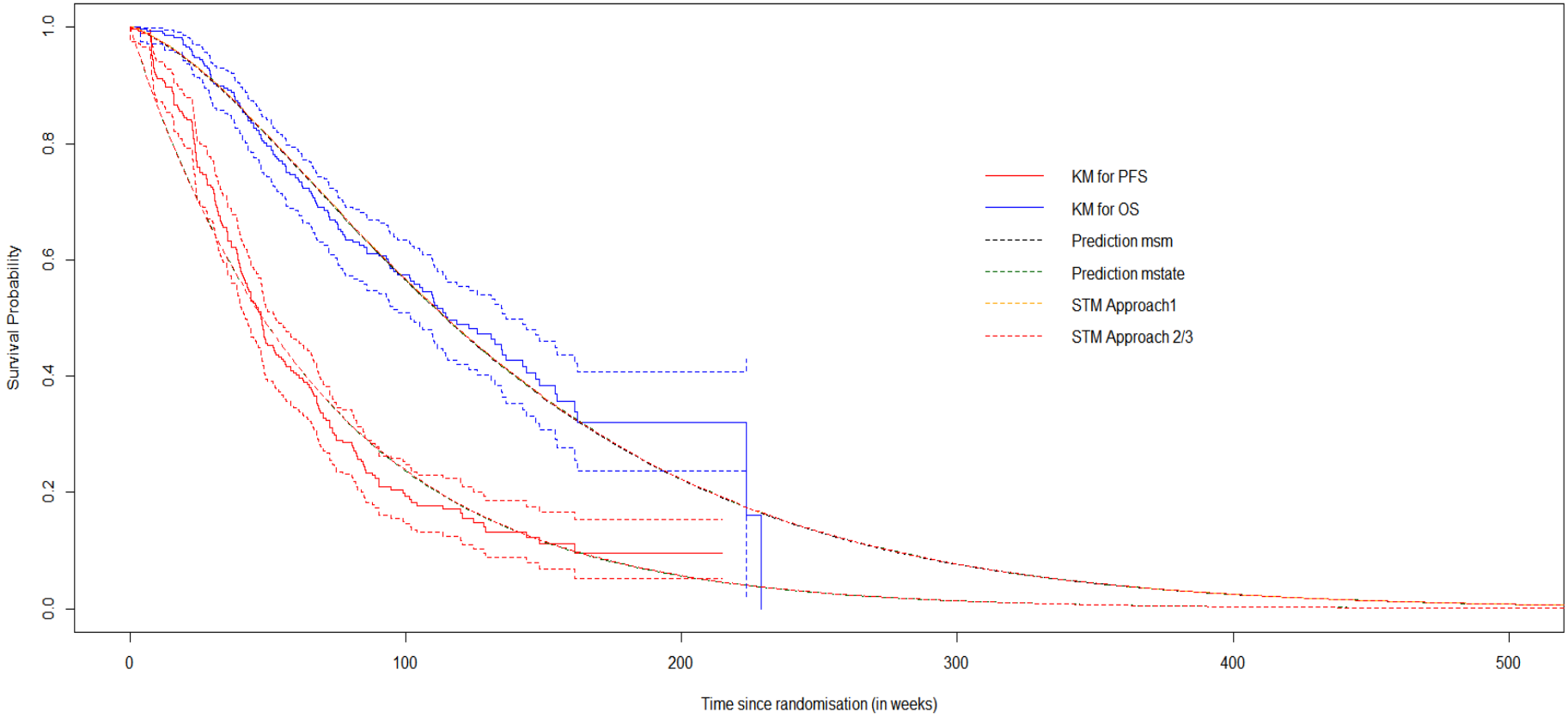


Figure 24 : Comparison of predictions for PFS and OS assuming transition rates are constant (curves are superimposed) in the BC dataset (CALGB 40502)



5.6 Health state sojourn time estimated with each approach

Health state sojourn times (time in PF, PD) estimated from the multi-state using msm, the mssample function, and the three implementations of the simplified STM are summarised in Table 6. As approach uses different inputs, the uncertainty is captured by bootstrapping the BC dataset rather than the value of each input parameters. This is to ensure that any differences observed would be attributable to the approach itself, rather than the differences in input parameters.

Table 6 : Summary of health state sojourn times (BC dataset boostrapped)

		msm	msmsample	simplified STM1	simplified STM2/3
Time in PF (in weeks)	point estimate	70.16	70.11	70.16	70.16
	LCI	61.60	61.77	61.62	61.62
	UCI	79.91	79.93	80.05	80.05
Time in PD (in weeks)	point estimate	68.14	68.63	68.57	68.63
	LCI	56.01	56.57	56.39	56.44
	UCI	81.74	82.21	82.14	82.20
Total LY (in weeks)	point estimate	138.30	138.75	138.74	138.79
	LCI	122.77	123.33	123.20	123.25
	UCI	156.22	156.39	156.66	156.71

In summary, approaches generate the same health state sojourn times.

5.7 Implications when transitions are time-varying

It is challenging to provide a like-for-like comparison between the current implementation of the Simplified STM and the MSM given that these approaches uses different input parameters. As described in Section 5.4, a single extrapolation is used for PFS in simplified STM. In contrast, PFS in the MSM is a function of the extrapolation of both TTP and PresPS. Different structural assumptions to separate progression from pre-progression mortality could also be made (Section 5.4.2).

As previously described in Section 5.4.1, differences in predictions between the MSM and the simplified STM are likely to be larger when: (a) data are less complete; (b) the fit of the model to the observed data is poor, and/or (c) where the reliance on extrapolation is greater.

Despite the difficulty in comparing the two approaches, I ran the MSM (using the `mssample` function from the `mstate` package) using eight parametric extrapolations (exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma and spline hazard model with one knot) for TTP and PrePS, giving 64 combinations of predicted PFS. PPS was assumed to be constant in all analyses.

In parallel, I also ran the Simplified STM as currently implemented in HE assuming:

- both PrePS and PFS follow eight possible parametric extrapolations, leading to 64 possible combinations of functions (referred to as Approach 3 in Section 5.4.2.3).
- PPS was assumed to follow an exponential distribution as in the implementation of the MSM.

As previously discussed, there is an alternative implementation of the STM, assuming a proportion of PFS events are deaths (Approach 1 in Section 5.4.2.1). Little difference is expected (See Section 5.4.2) against the implementation the STM selected here given that PFS will be the same, with the only difference being the number of deaths occurring pre-progression. Therefore, for ease of interpretation, this is not explored here.

Approaches were applied to the BC dataset as well as the prostate and lung cancer datasets, with results reported in Appendix 4 (due to the size of the tables). In summary, taking the BC dataset as an example, the mean OS predicted by the simplified STM ranged from 133.3 weeks to 166.5 weeks in the BC dataset, depending on the choice of parametric distributions assumed. In contrast, the mean OS predicted by the MSM using the `mssample` function ranged from 132.8 to 159.1 weeks. Similarly, the mean PFS predicted by the Simplified STM ranged from 66.6 weeks to 101.4 weeks, depending on the

choice of parametric distributions assumed. In contrast, the mean PFS predicted by the MSM using the `mssample` function ranged from 65.6 weeks to 93.6 weeks.

Whilst it is difficult to interpret, differences in predictions in LYGs between the two approaches were over 5% only when PFS/TTD followed a log-normal or log-logistic distribution. This is because these distributions were associated with a long tail (Appendix 1 for BC dataset) and the hazard was no longer based on the data itself but based on the extrapolated hazard. As expected, when the mean LYGs were calculated for the first 100 weeks only (to avoid the need for extrapolation), predictions between the MSM and the simplified STM were closer (not shown).

The process was repeated in the prostate and lung cancer datasets. Similar findings were obtained, in that larger differences (defined as >5%) were seen only for those scenarios where PFS/TTD was assumed to follow a log-normal or log-logistic distribution, due to the longer tail associated with these distributions.

It should be noted that it is difficult to make inferences based on a limited number of datasets and that the MSM and simplified STM use different inputs. Whilst it is challenging to compare predictions between approaches, this exploratory comparison confirms that some differences are expected between approaches. However, these differences are likely to be minimal, but could be larger depending on the parametric function used, the fit of the model to the observed data and the degree to which the model relies on extrapolation beyond the observed period.

5.8 Should we be concerned with the simplifications made in the Simplified STM (commonly used in HE)?

Despite multi-state modelling offering a convenient way to combine competing transitions under a competing risk framework, simplifications are usually made in health economic applications to avoid the need to model the competing transitions. The rapid review of NICE cancer appraisals also indicated that the STM has been implemented inconsistently between appraisals and that different assumptions could be made regarding how to separate PFS into its two constituent components (progression and pre-progression deaths). These different implementations can lead to differences in the predicted number of pre-progression death events, and therefore may also impact on predictions of OS.

As demonstrated in Section 5.4.1, the simplifications made in health economics have few implications when all transition rates are assumed to be constant. However, the implications are more significant when transitions rates are time-varying. Whilst it is difficult to directly compare the MSM and the Simplified STM due to the differences in inputs, larger differences in predictions could occur between implementation depending on the characteristics of the data and the need for extrapolation, as demonstrated in the exploratory analysis in the BC dataset in Section 5.6, but also in two separate datasets in patients with prostate and lung cancer.

This chapter has demonstrated that when transitions are selected appropriately and the fit to the data is good, it makes little difference whether transitions are modelled under a competing risk framework or using the assumptions typically made in HE. However, if transitions are selected poorly and do not fit the data well, differences in predictions are likely to be increased. This is because one approach relies on the fit and extrapolation of a transition in one dataset (PFS) compared with the fit and extrapolation in two datasets (TTP and PrePS). Therefore, differences are more likely to be attributable to the choice of parametric function and extrapolation, rather than the approach itself.

This will be demonstrated in Chapter 8, where the MSM is compared systematically against the STM in a simulation study, where I attempted to reduce potential biases in interpretation by selecting parametric functions using a similar process.

Prior to this, in the next chapter, I describe a review of methods to jointly model progression and survival outcomes in order to identify whether alternative approaches could be used to estimate health state sojourn time.

PART III: JOINT MODELLING OF PROGRESSION AND SURVIVAL OUCTOMES

6 CHAPTER VI: THE JOINT MODELLING OF PROGRESSION AND SURVIVAL OUTCOMES: A REVIEW OF METHODS

6.1 Chapter overview

This chapter aims to summarise methods that can be used to jointly model progression and survival outcomes.

Section 6.2 introduces this Chapter. Objective of the review are described in Section 6.3. Inclusion and exclusion criteria for the review are described in Section 6.4. In Section 6.5, I describe the search strategy, challenges associated with searching the methodological literature and how I addressed those challenges. In Section 6.6, 6.7, 6.8 I describe the screening process, the framework for the review and how data were synthesised respectively. Findings from the review are presented in Section 6.9.

6.2 Introduction

As described in Chapter 2, Woods *et al*.[9] and Bullement *et al*.[18] two key general approaches are currently used in health economics (HE) to estimate health state sojourn times and associated quality-adjusted life years (QALYs); (i) the partitioned survival approach (PSM) and (ii) the state-transition approach (including the MSM or the "simplified" STM which uses PFS directly). Under these currently used approaches, progression and survival outcomes are estimated independently from each other; thus, the current implementation in health economics does not typically consider the dependence structure or correlation between PFS and OS.

Indeed, in the current implementation of the PSM, OS and PFS are typically modelled as two independent processes, whereby parametric functions are fitted independently to the OS and PFS data from the trial, despite PFS events including death occurring before progression. Parameters are estimated for each outcome separately.

Similarly, within the current implementation of the STM approach, OS is estimated indirectly, as a result of three possible transitions, with the transitions typically estimated one at a time and often independently from each other, with the possible exception of PPS which could be estimated as a function of the time to progression TTP, although this is rarely done. Similarly, the dependence between

TTP and PrePS is often not considered within the standard STM used in HE; however, it is accounted for within the MSM framework whereby competing risks are considered.

The dependence is ignored primarily because estimates for each transition are estimated using separate datasets, rather than through the use of a jointly-fitted model. The impact of ignoring this dependence structure between progression and survival outcomes in health economic models is unknown. Consequently, prior to testing the performance of the different methods to estimate health state sojourn times (Chapter 8), I undertook a systematic review of the literature to identify methods that could be used to jointly model progression and survival outcomes that could be relevant to health economics. This review examined methods used within a range of other disciplines including operational research, statistics, engineering and environmental modelling.

6.3 Objectives for the review

The primary aim of this review is to identify methods that can be used to jointly model progression and survival outcomes (and associated parameters) when estimating health state sojourn time to improve on methods currently used in health economics.

A secondary objective was to identify methods that could be used to induce dependence between transitions in order to generate 'reflective' trial data that exhibit different degrees of dependence for use in the simulation study presented in the next chapters (Chapter 7 and 8). The review followed a systematic process in that it is reproducible and documented. However, it should be noted that there are a number of challenges associated with the conduct of a review of methods, which are described further below.

6.4 Inclusion/exclusion criteria for the review and search

Strict inclusion and exclusion criteria were applied. This is because a large number of citations and studies were expected. Inclusion and exclusion criteria are described below.

Inclusion criteria

- Full papers describing a methodology, not already identified (from sources already known or identified through the searches) to jointly model progression and survival outcomes or related survival endpoints,

Exclusion criteria

- Conference abstracts or presentations in which sufficient details are not available to replicate the approach
- Application of a method rather than its development; when the method has been previously identified
- Methods that do not relate to time-to-event outcomes
- Methods developed to account for the dependence between a longitudinal measurement e.g. blood pressure) and a time-to-event outcome (e.g. OS)
- Methods that cannot be used to extrapolate time-to-events outcomes (e.g. non-parametric)
- Methods developed to estimate the effect of covariates, rather than the joint prediction of health state sojourn time
- Methods to deal with competing risks or interval censoring only.

6.5 Search strategy

6.5.1 Method for searching

The methods to systematically search for evidence relating to the effectiveness of clinical interventions are well established, with searches typically being based on a defined PICO (population, intervention, comparator, outcome) specification of the research question. The types of studies to search for are known and therefore key terms can be defined *a priori*. For this type of systematic review, the aim is typically to identify all possible studies according to the PICO question given a set of search terms.[75, 76]

In contrast, methods for searching within the methodological literature are less established and are particularly challenging. This is due to the absence or limited knowledge of the studies that might be relevant for inclusion in the review. When searching for studies within a methodological review, whilst the broad concept of the method searched is known, the name of the method may be described using a variety of terms and therefore it is not possible to search for the method directly.

The challenges associated with conducting searches for methodological research have been discussed by Schlosser *et al* (2006),[77] Booth *et al* (2008),[78] Hutton and Ashcroft (1998),[79] Edwards *et al* (1998)[80] and Paisley *et al*.[81]

Unlike systematic reviews of interventions, there is no gold standard on how to conduct searches for methodological research. I rapidly searched the literature to identify previous examples of methodological reviews to provide an understanding of some of the approaches that have been used when searching for methodological papers.[82] I also looked at the methods used in similar PhD theses.[11, 83] Whilst the description for the method used to search for the methodological literature is often limited, these sources[11, 82, 83] used an iterative approach.

As part of the process of understanding methods for searching for a methodological review, I talked to an information specialist (Dr Suzy Paisley, ScHARR, HEDS, University of Sheffield) with expertise in searching methodological literature. Dr Paisley confirmed that an iterative approach would be the most appropriate method to conduct the searches for this review. With this approach, the searches are not fixed and evolve at the same time as the researcher develops a deeper understanding of the topic and the different methods available. This type of approach for searching, also known as "pearl growing" has been widely discussed in the literature by Schlosser *et al* (2006),[77] Booth *et al* (2008),[78] Hutton and Ashcroft (1998),[79] Edwards *et al* (1998)[80] and Paisley *et al*.[81] An advantage of this approach is that this is not static, and there can be some variation depending on what the searches are trying to achieve. Therefore, I considered that using such an approach would provide some flexibility and would allow the searches to be adapted to meet my needs.

In brief, I adopted the following process:

1. Relevant key papers were identified ('pearls') through known sources by myself and my supervisory team;
2. Key terms under which the key papers are indexed, and key terms used in the title and abstract were identified;
3. A keyword search was then conducted in a search database (Web of Science [WoS]) based on the key terms identified in the 'pearl' papers (through keywords used and reference searching).
4. In addition to the keyword searching, citation searching was undertaken with the aim of identifying papers that cite the identified 'pearl' papers. Citation searching is a useful alternative to subject searching which allows the identification of key papers that include the identified 'pearl' paper in their bibliographies.

5. The reference lists of the identified 'pearl' papers were also searched to identify potentially relevant published literature.
6. Papers identified through database, citation and reference searching were then screened to identify potentially relevant papers.
7. The process was repeated twice again until saturation was reached (further detail is available in Section 6.5.4). These potentially relevant papers were, in turn, assessed and new search terms were defined based on the keywords used in the papers and reviewing the reference list. Citation searches are also conducted to identify papers that reference the 'new' 'pearl papers.
8. Validation with experts to ensure that key methods have not been missed.

6.5.2 Identification of the initial 'pearl' papers

Five initial 'pearl' papers were used. The 'pearl' papers were primarily identified through known sources by myself and my supervisory team. The initial 'pearls' are summarised in Table 7.

Table 7 : Initial 'Pearl' papers

Author	Year	Title
Andersen[44]	2002	Multistate models for event history analysis
Glasziou[84]	1998	Quality adjusted survival analysis with repeated quality of life measure
Putter[43]	2007	Tutorial in biostatistics: competing risks and multi-state models
Williams[24]	2017	Estimation of survival probabilities for use in cost-effectiveness analyses: A comparison of a multi-state modelling survival analysis approach with partitioned survival and Markov decision-analytic modelling.
Williams[50]	2017	Cost-effectiveness analysis in R using a multi-state modelling survival analysis framework: A tutorial.

6.5.3 Initial keyword search

A search was initially conducted in the ISI Web of Knowledge on the 19th February 2018. The search was subsequently re-run on the 7th June 2020 in the ISI Web of Science core collection (formally Web of Knowledge) to ensure the review was up to date. The review was limited to peer-reviewed publications and therefore conference abstracts or grey literature including unpublished or ongoing research was excluded from the review.

To keep the review manageable, I searched for keywords included in the title only. This was a pragmatic decision which was taken due to the large number of citations retrieved when searching across all fields due to the use of broad terms. The number of citations retrieved by the searches was already high when searching titles only.

Initial search terms were identified through keywords used to describe the methods in the initial 'pearl' papers (described in Section 6.5.2) and papers included in their reference lists.

The first component of the keyword search involved searching for terms related to time-to-event outcomes. The second component involved searching for terms related to dependence structures. These search terms are then combined with an AND statement.

This was then combined with an additional filter related to methods in order focus the search on papers describing the method itself and its assumptions, rather than any applications of that method. Variations to search terms were used. For instance, to describe the terms 'dependent', related (synonym) search terms such as 'conditional' and 'joint' were considered.

In addition to keywords related to dependence and outcomes, additional keyword searching was conducted using terms describing previously identified approaches. This was done to identify potential papers which compare an alternative method to an approach previously identified. The search strategy is described in Table 8, with the additional search terms used for the second iterations of the search highlighted in bold. The search strategy was discussed with an information research specialist prior to the search being conducted.

6.5.4 Additional search terms (2nd iteration)

Search terms used for the second iteration are presented in Table 8 above highlighted in bold. Whilst the initial title keyword search was already sensitive, the search was slightly amended to include additional search terms related to outcomes and dependence. The initial search was then complemented with a search using more specific keyword terms.

It should be noted that the process can be repeated indefinitely and that this is can be a slow process. Therefore, a pragmatic approach was employed whereby a maximum of two iterations was considered to avoid repeating the process indefinitely with diminishing marginal returns. This would allow a large amount of literature to be captured whilst keeping the review manageable. This was also supported by the small amount of literature deemed relevant from the second iteration, as shown in the PRISMA diagram in Figure 25. Petticrew and Roberts (2006) and Schlosser *et al* (Schlosser 2006) consider saturation as an indicator for stopping further literature searching. Edwards *et al* (1998)[80] further notes that the marginal returns associated with reviewing additional papers diminish very quickly after a certain point. Citations were uploaded onto Endnote X8 reference management software.

Table 8 : Initial and second keyword search

Search	Search term	1st search	2ndsearch
#1	TI=(event* OR failure* OR survival* OR duration* OR hazard* **OR process OR processes OR occurrence**)	801,674	**1,691,665**
#2	TI=(dependenc* OR correlat* OR associat* OR join* OR relationship OR conditional* OR linked* OR clustered* OR connection*)	1,903,768	**2,540,902**
#3	TI=(model* OR method* OR approach* OR statistic*)	4,018,460	
#4	#3 AND #2 AND #1	1,850	**4,395**
#5	TI=(multistate OR "multi-state" OR markov OR "illness death" OR "partition* survival" **OR copula* OR frailt***)	29,609	**44,469**
#6	TI=(compar* OR versus OR alternative*)	1,569,753	
#7	#6 AND #5 AND #3	386	**475**
#8	#7 OR #4	2,234	
#9	TI=(progres* OR PFS)	245,271	
#10	TI=(overall survival OR OS OR death)	214,122	
#11	#10 AND #9 AND #2	285	
#12	#11 OR #8	**2,510**	
#13	TI=((multivariate* OR bivariate* OR cluster*) near/0 (survival* OR risk* OR event* OR failure* OR endpoint* OR time* OR distribution*))		**2,101**
#14	#13 AND #3		**605**
#15	#14 OR #7 OR #4		**5,392**
#16	TI=(semi competing or semicompeting)		**90**
#17	TI=((successive* OR sequential* OR consecutive* OR serial*) near/0 (survival* OR risk* OR event* OR failure* OR endpoint* OR time* OR distribution*))		**346**
#18	#17 OR #16 OR #15		**5,823**
#19	#18 NOT #12		**3,588**

6.5.5 Citation and reference search

In addition to keyword searching, within each iteration, the references of relevant papers were assessed to identify further papers. Further, citation searches were conducted to identify papers that cite the paper that was deemed relevant.

6.5.6 Expert advice

Due to challenges when conducting a review of methods, it is possible that some methods may not have been identified due to the lack of sensitivity of search terms, different terminology used in different disciplines or because the searches were restricted to titles.

Expert advice was sought when the review of methods was completed. Key authors of papers who have published in the area and experts in statistics were emailed a brief outline of the review objectives together with the list of key methods identified alongside the key publications to check that no relevant methods had been missed by the search. Experts were asked to validate whether any methods were missing and suggest further research for inclusion in the review. Experts were identified through the my supervisory team and authors of the papers identified.

6.6 Screening process

The titles and abstracts of the citations identified from the searches were screened and relevant full-text papers were obtained according to the inclusion criteria defined in Section 6.4. Papers were screened by one reviewer (myself).

Given that the focus of the review was on methods that could be used to jointly model progression and survival outcomes in order to estimate health state sojourn time, I did not aim to incorporate all applications of these methods to different case studies. However, excluding all applications outright may exclude methods that had not otherwise been identified. Therefore, I attempted to ensure that a method only described in an application was not excluded. In addition, whilst the focus is on methods that could be used to jointly model progression and survival outcomes, the identified methods were included when deemed relevant if they were used for different purposes. Important modifications of a method previously identified were also included.

6.7 Framework for the review

In addition to challenges associated with the searches, systematically reviewing methods papers poses similar challenges, and there is no ‘gold standard’ approach. Therefore, the review needs to be adapted

to its particular context. Hutton and Ashcroft (1998)[79] and Edwards *et al* (1998)[80] describe some of the challenges associated with conducting a review of methods. Edwards *et al*. (1998)[80] proposes that the review of methods must be considered according to an explicit framework and that the process by which literature is obtained and synthesised should be methodical, transparent and explicit. Hutton and Ashcroft (1998)[79] and Edwards *et al* (1998)[80] suggest that methods could be assessed in a framework in terms of validity, practical applicability, reliability, mathematical properties and theoretical arguments.

Consequently, a framework was developed to assess and describe the studies included in this review of methods (Table 9). The framework includes the questions I tried to answer for each of the method included in this review. The developed framework is based on criteria set out by Hutton and Ashcroft (1998)[79] and Edwards et al (1998)[80] and adapted from the frameworks used in previous PhD theses by Dr Nick Latimer[11] and Dr Jon Tosh[83] when searching for methods for adjusting for treatment switching and methods for optimisation, respectively.

It should be noted that answers to questions from this framework are typically limited to information available in the paper identified or additional information provided following discussion with experts. For instance, for a large number of the methods identified, limited details were available, reducing the scope for assessment of the method.

The framework focuses on four key elements of the identified methods:

(i) its development/origin;
(ii) its theoretical/mathematical properties;
(iii) its application and performance,
(iv) its applicability to health economics.

The ”development/origin” domain concerns whether the method was originally developed to jointly model PFS and OS and to estimate health state sojourn time, or whether it was originally developed to account for the dependence between other survival endpoints or other types of data (for example, longitudinal or binary outcomes). The domain also refers to whether the method is an extension of another method, how the method has been extended and how it compares with the original method.

Table 9 : Framework for the review of methods

Factor	Consideration
Development of the method	• What was the method originally developed for? • Is the method an extension of another method?
Theoretical properties	• What are the key assumptions? • What are its mathematical properties? • What are its key limitations?
Application of the method for the joint modelling of PFS and OS	• Has the method been used to jointly model PFS and OS? • Has the method been tested in a simulation study or a real-life example? • How did the method perform? • Are the authors aware of any limitations relating to its performance? • When the method has not been used for the joint modelling of PFS and OS, how did it perform when jointly modelling two processes (similar to PFS and OS)
Applicability to HE	• Is the method applicable and sufficiently flexible to be used in health economic evaluation? If not, why not? • Is an example of implementation provided by the authors? • Is the implementation of the method transparent?

The "theoretical/mathematical properties" domain concerns the key assumptions, mathematical properties and key limitations associated with the method. It is important to understand the theoretical properties of the methods in order to assess whether they are suitable for application in health economic evaluation.

The "application and performance" domain concerns whether the method has been used for the joint prediction of PFS and OS (or similar outcomes). This also concerns its performance when jointly modelling PFS and OS and whether the method has been tested in a simulation study or a real-life example. It should be noted that whilst the review is focussed on methodological papers, applications relevant to the joint modelling of progression and survival outcomes would have also been included if there had been any.

Finally, the "applicability to health economics" domain relates to whether the method is likely to be adopted for use in health economic evaluation in the near future. Applicability to health economics is an important domain given that the lack of details or examples on how to implement the method (example of implementation in a suitable software package such as Excel or R) will likely prevent the method from being widely adopted in health economics. The importance of the applicability to health economics is illustrated by the case of the MSM. Despite the method being originally described by Andersen *et al* (2002),[44] and described in a tutorial by Putter *et al* (2007)[43] using the `mstate` package and Jackson et al (2011)[85] using the `msm` package over a decade ago, only recently has interest increased in the use of this method as a vehicle for health economic evaluation, spearheaded by the publication of a tutorial by Williams *et al* (2017) on how to model PFS and OS using the multi-state framework in health economics.[50] Despite the publication of a thorough tutorials on the method and its application in the context of health economics, this approach is still rarely used. The applicability to health economics domain also discusses whether the approach is flexible enough to be routinely used, or if its use is likely to be restricted, for instance, if it is limited to specific parametric distributions or to individuals with specialist skills (e.g. statistics).

6.8 Data extraction, assessment and synthesis

A template, based on the framework defined in Section 6.7, summarising the relevant details of each study included in the methods review was completed by myself, which has not been provided as this contains exactly the same information as provided in the main body of text in this chapter. Indeed, in the absence of relevant assessment criteria when conducting a review of methods, each method was judged based on its theoretical properties; in particular, with respect to the main assumptions, how well the method is described and potential key limitations. The assessment of the theoretical properties of a method helps in understanding how a method can be used to jointly model PFS and OS.

Details are synthesised directly in a narrative form in the main body of this chapter. Key characteristics of the methods identified are described as well as their mathematical form. Methods were grouped where possible according to key characteristics.

6.9 Summary of methods identified within the review to jointly model PFS and OS

This section summarises the results of the systematic review of methods identified in the literature that could be used to jointly model progression and survival outcomes. The flow diagram depicting the number of records identified, screened and included is shown in Figure 25. The initial keyword search yielded 2,510 citations, of which 12 were included following sifting of the titles, abstracts and papers (Belckacemi et al, 2014;[86] Dejardin, 2010;[87] Fleischer et al, 2009;[88] Fu et al, 2013;[89] Krol et al, 2017;[90]

Li et al, 2015;[91] Mazroui et al, 2013;[92] Meller et al, 2019;[93] Oakes et al, 1982;[94] Rondeau et al, 2007;[95] Rondeau et al, 2012;[96] Weber et al, 2019.[97]). No further relevant citations were obtained following the initial screening of the 1,490 citation searches and 212 reference searches.

At the second round of searches, the keyword search yielded 3,588 additional citations, of which one additional citation was included following sifting of the titles, abstracts and papers (Sildnes *et al*, 2018[98]). The citation and reference searches yielded to 526 and 389 citations, respectively; however, none of these were considered relevant. The citation and reference searches for the paper identified in Step 4[98] yielded to one and 32 citations, respectively; of these, none were considered relevant. I considered that saturation was achieved given that only one citation was considered relevant in Step 4, and therefore, no further searches were conducted.

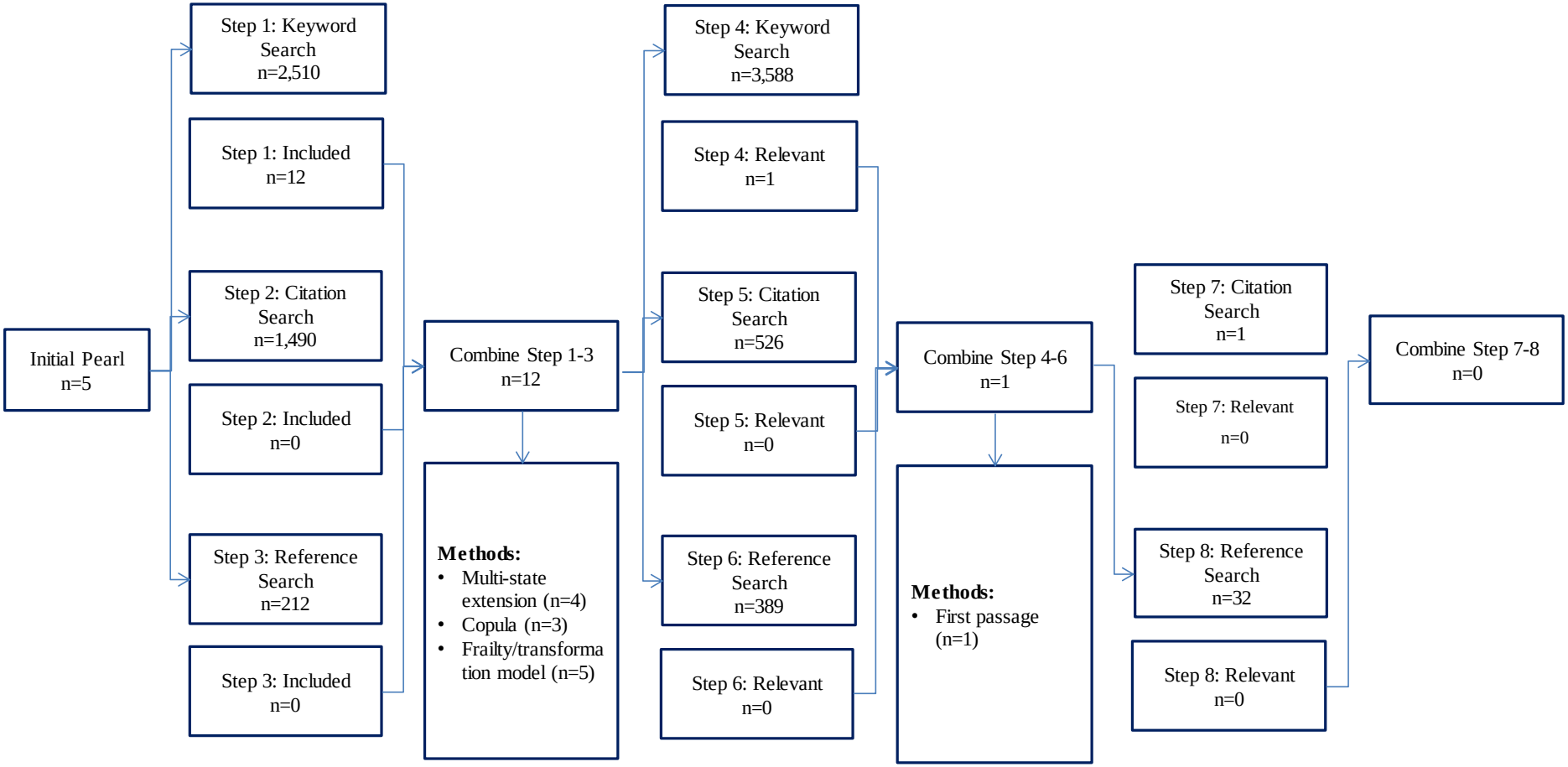
Given the large number of records, the specific reason for exclusion was not explicitly recorded for each citation. However, the key reason for exclusions were: lack of relevance of the paper (e.g. analysis of trials); methods related to the modelling of longitudinal and time-to-event data (for which the volume of literature is very large), methods for meta-analysis or applications of the same method previously identified.

Some papers included in the review described the same methods (in particular for the frailty and copula models) but provided additional details. This is one of the challenges of conducting a systematic review of methods. Consequently, rather than focusing on a description of each individual paper, papers were grouped according to the methods they describe.

It should be noted that a large number of papers are available describing each of these methods, in particular for the frailty and copula models. A large number of applications were identified, but were not formally included. Furthermore, additional papers describing these methods may not have been identified during this review. Consequently, the included papers may not necessarily reflect the most comprehensive papers that describe the method, but rather the papers identified during the systematic review process and considered to be relevant by myself for the description of the method, or that I considered for the identification of the method (the joint modelling of PFS and OS).

Given the challenges associated with conducting a review of methods, flexibility was required when including and excluding studies. When describing the methods in this thesis, in addition to the papers that were formally included during this process (e.g. papers included in Appendix 5), additional papers that were formally excluded (for instance because the paper described an application of a method previously identified) were also retrieved informally to provide supplementary information about the method when necessary.

Figure 25 : Flow diagram for the review of method



Approaches identified during this systematic review process were grouped into four categories; these are described in the next sections:

1. General extensions of the multi-state model to jointly model progression and survival outcomes (described in Section 6.9.1)
2. Methods to jointly model progression and survival outcomes, inducing the dependence from a random effect term (frailty) in PH models or its extension using transformation models for AFT models (described in Section 6.9.2)
3. Methods to jointly model progression and survival outcomes, using a copula model (described in Section 6.9.3)
4. Semi-competing risk by means of first passage times (described in Section 6.9.4).

Throughout the systematic review process, a number of specific bivariate models were also identified; however, these were excluded upon further inspection, as they were special cases of more general copula models (Group 3 above). Similarly, a general approach that has been used for the analysis of bivariate models, described by Henderson *et al* (1995), was not included in this review as the general concept was known (e.g. when including TTP as a covariate when estimating PPS) and consisted of modelling the conditional distribution of a failure time (e.g. T1) given another failure time (T2) as covariate.

Identified approaches fall further under two umbrellas:

1. Those where the dependence is included between transitions (under the illness-death model)
2. Those where the dependence between OS and PFS is included by modelling the survival outcomes under a semi-competing risks model.

As part of the systematic review process, seven experts were contacted to ensure that no additional relevant approaches were missing. Four of the seven experts contacted as part of the review process responded. They believed that the list of methods identified during the systematic process was generally exhaustive. Experts suggested a number of additional papers of interest; however, following inspection, the majority of the papers suggested related to evidence synthesis (using bivariate models or frailty models) for the joint modelling of PFS and OS, rather than the joint prediction of progression and survival outcomes within a single trial (as is the focus in this thesis). These methods (bivariate/frailty models) are already included in this review but are versatile and are used for a variety of purposes, therefore, including all possible variations is outside the scope of this thesis. For transparency, the reasons for the exclusion of papers suggested by the key experts is described below:

- All four experts suggested the addition of further papers for inclusion on the application of the multi-state model.[43, 45, 50, 99, 100] However, these did not represent new methods (and were already included) and therefore were not considered further.
- An expert suggested that there was an increasing interest in predicting OS based on tumour imaging data and suggested a paper for potential inclusion.[101] This was not considered further as this was outside the scope of the review and did not meet the review inclusion criteria.
- A paper describing a Bayesian approach for jointly modelling correlated outcomes was also suggested as a possible approach to use in health economics.[102] However, following examination of the paper, the proposed approach was similar to a bivariate model and therefore was not considered further. The authors of this paper were contacted and confirmed that the approach used was a bivariate joint distribution model.
- An expert suggested for inclusion some of work carried out on multi-level mixed effect parametric survival analysis as a potential approach. However, following examination of some of the papers found, the approach described used a frailty (random effect) model and was therefore not considered further as this general approach was already included.
- Additional papers on joint modelling in evidence synthesis (NMA) were also highlighted, but were outside the scope of this review[103-105]
- An approach used to jointly model PFS and OS employed in a previous NICE STA (TA561[106]) was also suggested by one of the experts for possible inclusion in this review for completeness. In this appraisal, the company modelled PFS and OS jointly across both arms, assuming proportionality and the same parametric form between OS and PFS. In other terms, PFS was included as a covariate when estimating parameters for the parametric function for OS. The key assumption in this approach is proportionality between PFS and OS. Little detail on the approach was available or included in TA561.[106] No literature was identified describing this approach, limiting any assessment in this review. The expert further considered this approach to be crude, and therefore it was not formally included in this review.
- Finally, one expert mentioned the use of a non-parametric approach to combine evidence on time-to-event outcomes and have used a constraint to ensure that PFS is less than OS which has been applied to an appraisal in lung cancer.[107] The area under the curve (AUC) was calculated for PFS and OS, with the correlation between the two endpoints included in the NMA using a bivariate model. Patients alive after 5 years were then assumed to be remission as the curve converged in the trials with external evidence used after that time point. Whilst this approach has been used in a NICE appraisal,[107] the expert highlighted that this approach is not yet been published. While the method was unpublished, this was still considered as new methods may address limitations of currently published methods. However, from the description, this

approach (using non-parametric AUC using a bivariate model) appears to be mostly relevant when synthesising OS and PFS from different sources, rather than the joint modelling of progression and survival outcomes within a single trial. Strong assumptions are also required to extrapolate beyond the trial period. Consequently, in the absence of further details, this approach was not considered further.

Following discussion with other analysts, an additional paper of potential interest was suggested relating to the use of moment-generating functions in health economics.[108] However, this paper did not meet the inclusion criteria as it was not related to the joint modelling of progression and survival outcomes.

As no additional methods included in this thesis were identified through expert opinion, the final taxonomy was not sent back to expert for validity testing. The key characteristics of methods identified during the systematic process are summarised in Table 10.

6.9.1 General extensions of the multi-state/illness-death model for the joint modelling of PFS and OS

Three general extensions of the MSM were identified to jointly model PFS and OS. Parameters for each transition are estimated jointly. These methods are described in turn below.

6.9.1.1 The model proposed by Li et al (2015)[91]

The authors aimed to extend the model developed by Fleischer *et al* (2009)[88] based on the assumption of exponential distributions to a Weibull distribution to include the dependence between progression and survival outcomes in a multi-state model.

6.9.1.1.1 Development of the method

This is an extension to the statistical model developed by Fleischer *et al* (2009)[88] which is, in turn, an extension of the multi-state framework. Fleischer *et al* (2009)[88] developed a statistical model based on exponential distributions that describes the dependence structure between OS and PFS. The model developed by Li *et al* (2015)[91] generalises the exponential to the Weibull distribution, in order to provide additional flexibility.

This model was developed to predict PFS and OS; but could be extended to other time-to-event outcomes. This section focuses on the model developed by Li *et al* (2015) as this includes both the case where the transitions follow an exponential distribution (Fleischer's model) or the Weibull distribution (Li's model). Consequently, despite being identified, the model by Fleischer *et al* (2009)[88] is not described here as this is the same model, but uses an exponential rather than the Weibull.

Table 10 : Summary of approaches identified to jointly model PFS and OS

	Extension of the multi-state			**Frailty model/extension**	**Copula models**	**First passage**
	Li *et al* (2015);[91] Fleischer *et al* (2009)[88]	Belkacemi *et al* (2014)[86]	Meller *et al* (2019)[93]	Krol *et al* (2017),[90]Mazroui *et al* (2012),[92] Rondeau *et al* (2007),[95] Dejardin *et al* (2010)[87]	Fu *et al* (2013),[89]Weber *et al* (2019),[97] Oakes *et al* (1982)[109]	Sildnes *et al* (2018)[98]
Key properties/assumptions	Transitions follow a Weibull distribution with same shape parameter	· Transition follow an exponential distribution · Conditional distribution	Likelihood estimated using counting processes	Frailty acts multiplicatively on the hazard	Form of bivariate model	The joint distribution must satisfy random sign censoring
Has the method been applied to jointly model PFS and OS	Yes – within an illness-death model	Yes – within an illness-death model	Yes – within an illness-death model	Yes – a frailty term shared between TTP and PPS	Yes – under semi-competing risk	Yes – semi-competing risk
Key limitations	· Limited to Weibull distribution · Same shape assumed between transitions	Limited to exponential	Unpublished at the time of writing of this chapter	· Limited to PHM (possible to extend to AFT) · Choice of frailty distribution and model	Large number of copulas available	Terminal and non-terminal event follow same underlying process
Example of implementation available? and/or key barriers for use in health economics	Example of R code provided by the authors	Exponential is too restrictive	Code not available	Challenging to implement in the absence of a step-by-step tutorial – given the different possible formulations	Example of R code provided by the authors (Fu, 2013) Different copula would require different formulation	No example of implementation in a suitable package Gamma process (restrictive)

Abbreviations: AFT: accelerated failure time; OS: overall survival; PFS: progression-free survival

6.9.1.1.2 *Theoretical properties*

The model proposed by Li *et al* (2015) can be represented as an MSM with transition intensities which follow a Weibull distribution. The model has four parameters $\lambda_1, \lambda_2, \lambda_3 \; and \; \alpha$; with the same shape parameter (α) being shared between transitions. Consequently, the dependence/correlation between transitions is induced by transitions sharing the same shape parameter. As the Weibull is an extension of the exponential, the same results will be obtained by Fleischer *et al* (2009)[88] when $\alpha = 1$.

The working assumption is that TTP and OS are independent (this is the "maximal independence assumption"), with PFS given by the minimum of TTP and OS. If death occurs before progression, then PFS=OS (for an individual patient). If progression occurs first (TTP<OS), OS would be equal to the sum of TTP and a new variable OS'(latent time).

Equations described here are reproduced from the paper directly.[91] Further information is available in the paper.[91]

This can be written down as:

$$TTP \sim f_1 = Weibull(\alpha, \lambda_1) \; OS_{orig} \sim f_2 = Weibull(\alpha, \lambda_2), PFS = min\,(TPP, OS_{orig})$$

$$OS' \sim f_3 = Weibull(\alpha, \lambda_3), OS' \perp\!\!\!\perp TTP \perp\!\!\!\perp OS_{orig}$$

$$OS = \begin{cases} PFS, if\ PFS \neq TTP \\ TTP + OS', if\ PFS = TTP \end{cases}$$

Where $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$ and $\alpha > 0$

For this model, the survival function for OS is given by [equations reproduced from the paper[91]]:

$$S_{OS}(x) = \exp[-(\lambda_1 + \lambda_2)x^{\alpha}] + \lambda_1 \alpha \int_0^x \gamma^{\alpha-1} \exp\,[-(\lambda_1 + \lambda_2)\gamma^{\alpha} - \lambda_3(x-\gamma)^{\alpha}]d\gamma$$

The parameters $\lambda_1, \lambda_2, \lambda_3$ and α are estimated using maximum likelihood estimation (MLE), with the likelihood calculated for four types of patients (represented by δ_n):

1. those, who progress and then are censored without death (δ_1),
2. those, who progress and then die (δ_2),
3. those who die before progression (δ_3),
4. those, who are censored without progression or death (δ_4).

The likelihood for these four patients' categories can then be calculated based on the density $f()$ and the survival function $S()$ so that:

if $\delta_i = 1$ then

$$L_i^{(1)}(\theta) = P(TTP = t_{i1})P(OS_{orig} > t_{i1})P(OS' > t_{i2}) = f_1(t_{i1})S_2(t_{i1})S_3(t_{i2})$$

if $\delta_i = 2$ then

$$L_i^{(2)}(\theta) = P(TTP = t_{i1})P(OS_{orig} > t_{i1})P(OS' = t_{i2}) = f_1(t_{i1})S_2(t_{i1})f_3(t_{i2})$$

if $\delta_i = 3$ then

$$L_i^{(3)}(\theta) = P(TTP > t_{i1})P(OS_{orig} = t_{i1}) = S_1(t_{i1})f_2(t_{i1})$$

if $\delta_i = 4$ then

$$L_i^{(4)}(\theta) = P(TTP > t_{i1})P(OS_{orig} > t_{i1}) = S_1(t_{i1})S_2(t_{i1})$$

with the overall log likelihood being the sum of the log likelihood across all subjects.

It should be noted that the model proposed by Li *et al* (2015),[91] and by extension Fleischer *et al* (2009),[88] is considered to use a latent failure time approach by Meller *et al* (2019),[93] given that the estimation of the transition for PFS can occur after OS (but this is adjusted as PFS cannot be greater than OS in the final model). Meller *et al* (2019) suggest that this could make the interpretation of the estimate difficult, as the predictions may not be as meaningful from a clinical point of view.

Although not presented here, it is also possible to derive analytical correlations amongst TTP, PFS and OS from Li's model.

6.9.1.1.3 Application of the method to the modelling of PFS and OS

Li *et al* (2015) applied their method in three cancer trials.[91] Overall, the authors reported that their model provided a good fit to the data and, as expected, a better fit compared with the exponential distribution. They also demonstrate that whilst the assumption of a constant shape parameter α could be questionable, the model can still, in general, provide a good fit to the data during the observed period of a study.

In addition, the authors conducted a simulation study to evaluate the fit of the Weibull and exponential models using hypothetical data sets generated from various distributions:

1. Assuming the simulated event times follow an exponential distribution, the authors report that both the Weibull and exponential functions provided an accurate fit to the data.
2. When the simulated event times are generated from a Weibull distribution with the same shape parameter (across transitions), the Weibull model provided a good fit to the data. In contrast, the exponential distribution fitted the data less well.
3. Assuming the data are generated from a Weibull distribution with different shape parameters; the Weibull model provided a reasonable fit, despite the assumption of common shape parameter being violated. In contrast, the exponential provided a poor fit.
4. When the simulated event times were generated from a log-logistic distribution (hazard had a non-monotonic shape), the Weibull model provided a reasonable visual fit to the KM and performed better than the exponential, as expected,
5. Similarly, when the simulated event times were generated from a log-normal distribution, the Weibull model provided a reasonable visual fit to the KM.

The model developed by Li *et al* (2015)[91] was also evaluated in Meller *et al* (2019)[93] using a simulated dataset, as well as in a real-life example, using data from the CLEOPATRA trial; a Phase III RCT in HER2-positive metastatic breast cancer involving 808 patients. Using the simulated data, Meller *et al* (2019) reported that the model developed by Li *et al* (2015) tended to over-estimate the correlation coefficient between PFS and OS. Using real data from the CLEOPATRA trial, Meller *et al* (2019) reported that whilst the model proposed by Li *et al* (2015) provided a slight over-estimate of the transition between progression and death, the model provided a reasonable fit to PFS and OS.

6.9.1.1.4 Applicability of the method to health economic evaluation

The method was specifically designed for estimating jointly PFS and OS and therefore can be applied easily to an RCT dataset for use in health economic analyses. The statistical model proposed by Li *et al* (2015)[91] is a general illness-death model. Compared with the model by Fleischer *et al* (2009),[88] a key strength is that it uses Weibull models which are more flexible compared with exponential models. The Weibull model is expected to provide a better fit with a more accurate estimation of the correlation (formula for analytical estimation of correlation not shown here). Whilst its implementation is more complex, a copy of the R code is available from the author on request and is relatively straightforward

to implement (as demonstrated in the simulation study in Appendix 12). Despite the primary aim of the author to estimate OS, the code is easily adaptable to extract PFS predictions as well.

A key limitation is that transitions are assumed to follow a Weibull distribution, which may not always be appropriate. Furthermore, the same shape parameter is assumed between transitions, which is a simplification. However, simulation studies conducted by the authors showed that even when the underlying distribution has a non-monotonic hazard, the proposed Weibull model (using the same shape parameter) could still provide an adequate fit to the observed data. The use of the same shape parameter is justified by the author on the basis of mathematical convenience. The authors also argue that usually the shape of the hazards for progression and death are similar; therefore, this is a plausible assumption. However, even when the common α assumption is violated, the authors show that the proposed Weibull model with the same α parameter fits the data adequately. The authors suggest this could be relaxed and it is possible to use different shape parameters. However, the correlations among the survival endpoint and OS can no longer be derived analytically.

In summary, the model proposed by Li *et al* (2015)[91] presents a potential alternative to the STM as currently implemented in health economics as it allows for the joint modelling of PFS and OS. The model is relatively straightforward to implement in R. Whilst a key limitation is the use of the Weibull distribution, the Weibull distribution is often used in health economic analyses and appears to provide a reasonable visual fit to the observed data, even when data are generated using other distributional forms. Nevertheless, the plausibility of the predictions beyond the trial period generated using this method remain unclear.

6.9.1.2 The model proposed by Belkacemi et al (2014)[86]

The authors aimed to develop a model, whereby PFS and PPS are linked using a conditional exponential distribution to test the existence of an association between PFS and PPS to better understand the process of improvement or decrement of OS.

6.9.1.2.1 Development of the method

The model proposed by Belkacemi *et al* (2014)[86] is also an extension of the multi-state framework, whereby OS is modelled based on PFS and PPS, but using a conditional distribution.

6.9.1.2.2 Theoretical properties

The model proposed by Belkacemi *et al* (2014)[86] considers the conditional association between PFS and PPS. In other words, the dependence structure between PFS and PPS is represented by a conditional

distribution. This is different to the model developed by Flesicher *et al* (2009)[88] and Li *et al* (2015) which did not consider conditional distributions.

The working assumption of the model proposed by Belkacemi is that:

- If both progression and death are observed, OS is composed of two survival times; (a) PFS and (b) PPS,
- If only death is observed, this is counted as a progression event and therefore PFS $\leq$ OS
- PFS and PPS are statistically dependent.

Equations described here are reproduced from the paper directly.[86] Further information is available in the paper.[86]

Belkacemi *et al* (2014) describe the survival function for OS as:

$$S_T(t) = S_X(t) + \int_0^t S_{Y|X}(t-x|x) f_X(x) dx$$

Where:

T represents OS, X is PFS and Y is PPS and $S_{Y|X}$ is the conditional survival function of Y given X

Estimates for the survival function are then generated using MLE. The likelihood is calculated by dividing the contribution of patients in 5 categories [equations reproduced from the paper[86]], where, for the ith patient, $\delta_{i(i=1,\ldots,N)}$ is a categorical indicator, n_j the number of patients in each category, C the administrative right censoring time,

For $\delta_i = 1$, both progression and death are observed

$$L_1 = \prod_{i=1}^{n_1} f_X(x_i)\, f_{Y|X}(y_i|x_i)$$

where $y_i = t_i - x_i$

For $\delta_i = 2$, only progression is observed

$$L_2 = \prod_{i=1}^{n_2} f_X(x_i)\, S_{Y|X}(y_i^c|x_i)$$

where $y_i^c = c_i - x_i$

For $\delta_i = 3$, only death is observed (related to cancer)

$L_3 = \prod_{i=1}^{n_3}\{1 - S_X(t_i)\}f_T(t_i)$

For $\delta_i = 4$, only death is observed (unrelated to cancer)

$L_4 = \prod_{i=1}^{n_4} S_X(t_i)\, f_T(t_i)$

For $\delta_i = 5$, neither progression nor death are observed

$L_5 = \prod_{i=1}^{n_5} S_X(c_i)\, S_T(c_i)$

The model proposed by Belkacemi *et al* (2014)[86] assumes that both PFS and PPS (conditional on PFS) are exponentially distributed with parameters λ for PFS and θ(x) for PPS.

As PPS is conditional on PFS, θ(x) is assumed to follow a conditional exponential distribution with 2 parameters β and v such that:

$\theta_{Y|X}(x) = \alpha.\exp(-\beta x)$

where $\alpha = \exp(v)$ and $\beta < \lambda/2$

The correlation coefficient β provides an indication of the correlation between PFS and OS and PFS and PPS. The authors observed that if PFS and OS are positively correlated, an improvement in PFS would therefore lead to an improvement in OS. However, the authors state that this is different to assuming that PFS is positively correlated with PPS, because the improvement in PFS may translate to lower PPS.

The hazard function for PPS can be written as:

$h_Y(y) = \frac{\int_0^\infty \theta_{Y|X}(x)g(x,y)dx}{\int_0^\infty g(x,y)dx}$

The hazard function for OS can be written as :

$h_T(t) = \frac{\int_0^t \theta_{Y|X}(x)g(x,t-x)dx}{\frac{exp(-\lambda t)}{\lambda}\int_0^t g(x,y)dx}$

where $g(x,y) = \exp\{-[\theta_{Y|X}(x)y + \lambda x]\}$ and the assumption that X and $Y|X$ to be exponentially distributed with parameters λ and $\theta(x)$

6.9.1.2.3 Application of the method to the modelling of PFS and OS

The authors applied their method in a Phase III clinical trial of patients with NSCLC.[110] Parameters were found by optimisation methods using two R software packages; `DEoptim` (a package to perform global optimisation, used to find starting values), and `alabama` (which includes functions that use the augmented Lagrangian and adaptive barrier minimisation algorithm in which constraints are allowed). The model is applied with (explanatory variables considered relevant by the authors including tumour stage, performance status and toxicity) and without covariates. In their dataset, the β parameter (the correlation coefficient) was statistically significant, indicating a link between PFS and PPS confirming the adequacy of the conditional model.

Overall, the authors show that the proposed model provided a reasonable fit to OS, PFS and PPS when no covariates are considered. When considering covariates in their dataset, the authors suggest that their model performed better for the estimation of OS (especially at the tail) compared with the Cox-semi Markov model. The authors state that results were consistent with Fleischer *et al* (2009)[88] in the same dataset; despite the statistical models being different. It should be noted that both Belkacemi *et al* (2014)[86] and Fleischer *et al* (2009)[88] use exponential distributions, but the model proposed by Belkacemi *et al* (2014)[86] takes into account the dependency between PFS and PPS, that is not considered by Fleischer *et al* (2009).

6.9.1.2.4 Applicability of the method to health economic evaluation

Whilst this method could be easily applied to an RCT dataset, the model developed by Belkacemi *et al* (2014) has limited application to the HE context as it is very restrictive and is limited to the exponential distribution only (applied as a conditional exponential distribution). The exponential distribution was used for mathematical convenience. The authors suggest that the generalisation to the Weibull distribution is numerically possible, but that the estimate of the correlation becomes analytically impossible to derive. The absence of code to reproduce results limits its applicability to health economics.

6.9.1.3 The model proposed by Meller et al (2019)

6.9.1.3.1 Development of the method

The model proposed by Meller *et al* (2019)[93] is also an extension of the multi-state framework. However, compared with the models developed by Fleischer *et al* (2009)[88] and Li *et al* (2015),[91] which according to the authors, use a latent failure time approach, the model developed by Meller *et al* explicitly considers all "transition intensities between all states." The model proposed by Meller *et al* (2019)[93] has been specifically developed for the joint modelling of PFS and OS, but is applicability is not restricted to the modelling of PFS and OS and could be used for any illness-death model.

6.9.1.3.2 Theoretical properties

The model proposed by Meller *et al* (2019)[93] follows a general illness-death MSM approach. PFS is defined as the waiting time in the initial state. OS is defined as the time until reaching the death state.

The authors state that no assumptions are made, with the exception that there are no progressions after death. PFS < OS if a progression event occurs, whilst PFS = OS if a patient transition directly from the initial state to death (i.e. the patient dies prior to progression). PFS is first estimated based on the hazard of patients moving from the initial state to either the progression or death health state. In the second step, a binomial distribution is used to define if PFS<OS. The residual time until death in people for whom PFS<OS is then simulated. Neither latent times nor copulas are used in the model developed by Meller *et al* (2019).[93]

The model proposed by Meller *et al* (2019)[93] is therefore very similar to the multi-state framework used by Jackson *et al* (2012)[45] and Putter *et al* (2013)[43] whereby all transitions are explicitly modelled.

However, whilst the general framework is similar, parametric inference for the model proposed by Meller *et al* (2019)[93] relies on MLE for counting processes, thus, it is based on the contributions of the three transitions. This contrasts with the multi-state framework in Putter *et al* (2013), whereby transition intensities are estimated independently of each other one at a time and then combined under a competing risk framework.

In addition to the estimation of the transition intensities, the authors demonstrate how the correlation between PFS and OS can be computed (see Meller *et al* (2019)[93] for further details). This is not described here as it is beyond the scope of the review.

6.9.1.3.3 Application of the method to the joint modelling of OS and PFS

The authors compared their model with the model by Fleischer *et al* (2009)[88] and Li *et al* (2015)[91] in both a simulation study and a real case example. In the simulation study, the authors show that their model provided a better estimation of the transition between the PD state and death, and therefore did not result in such an over-estimation of the true underlying correlation. This is explained by the fact that the shape of Weibull distribution was not assumed to be constant between transitions.

In a real-life example, using data from the CLEOPATRA trial; a Phase III RCT in HER2-positive metastatic BC involving 808 patients, the authors found that:

- for the transition between PF and PD, all models provided a very similar and satisfactory visual fit
- for the transition between PF to death, the model proposed by Meller provided a slightly better visual fit
- for the transition from PD to death, no direct comparison is provided as estimates are shown on different scale. The authors stated that the model by Meller slightly over-estimated the hazard for that transition. The model by Fleischer *et al* (2009)[88] and Li *et al* (2015)[91] provided a reasonable visual fit to the observed data (Nelson-Aalen); however, in the absence of a direct comparison with the method of interest, it is unclear how they performed compared with the model proposed by Meller *et al* (2019).[93]
- when considering the fit to the PFS and OS survival functions, the estimate for PFS was generally similar irrespective of the approach used. As the authors pointed out, this is not surprising, since: (a) all approaches provide similar estimates for the transition between PF and PD, and (b) the number of transitions from PF to death is relatively much lower than the number of transitions to PD.
- the authors also found that estimates for OS are also very similar between approaches, although they do not comment on the implications of this finding.
- finally, the authors note that the estimated correlation coefficients are similar between approaches, but are associated with large confidence intervals, indicating the difficulty in estimating the correlation coefficient with precision irrespective of the approach used.

6.9.1.3.4 Applicability of the method to health economic evaluation

The model proposed by Meller *et al* (2019)[93] was developed to jointly model PFS and OS and therefore, is highly relevant to the context of health economics.

A key advantage of the model proposed by Meller *et al* (2019),[93] compared with other extensions of the multi-state framework discussed previously (Li *et al* (2015)[91] and Fleischer *et al* (2009)[88]) is that there is no restriction on the parametric distribution and strong assumptions are not made e.g. the assumption used by Li[91] regarding the same Weibull shape parameter for all transition intensities.

Despite its flexibility, results from the simulation study and real-life case study example remain inconclusive, with the authors noting that the estimates for PFS and OS are relatively similar irrespective of the approach used in the case study example. It should be noted that this could be attributable to the data used, and that the methods might perform differently using different datasets. The evaluation of the performance of the method was also limited to the goodness-of-fit to the observed data (visual fit) and therefore does not account for the plausibility of the long-term extrapolation.

A key barrier for the adoption of the method described by Meller *et al* to health economics is the lack of detail provided regarding its implementation. Whilst the authors describe their general model and the approach used for statistical inference, few details are provided by the authors on how to reproduce the results, leaving the reader/analyst to program the implementation based on their understanding of the method described by the authors. No code is available in the final paper published in *Statistics in Medicine*; this is a key barrier for its immediate adoption in health economics. Analysts typically in charge of developing such health economic models may not have the technical skills to implement such an approach without a comprehensive tutorial. A copy of the code used to generate results in the paper was requested from the main author, but was not shared and therefore information is limited to what is contained in the paper.

6.9.2 Jointly model PFS and OS and inducing the dependence by the introduction of a random effect – the frailty/transformation model

6.9.2.1 Development of the method

This method is not specific to the joint modelling of progression and survival outcomes but is a general approach to account for the dependence between two survival endpoints. Frailty models[111] are an extension of the Cox PH model. They account for the heterogeneity caused by unmeasured covariates. The Cox model can be considered as a frailty model without a random effect term. Thus, the frailty acts as a random effect.

Frailty models can be used for a number of purposes, but are widely used for clustered survival data when modelling the dependence between two processes belonging to the same cluster. For instance, frailty models are used extensively in family disease studies for instance, where the time to disease onset for individuals within a family (cluster) are correlated, possibly due to sharing similar environmental and/or genetic conditions.[112]

6.9.2.2 Theoretical properties

Frailty models were originally introduced for PH models and were later extended to AFT models. The key assumption in the frailty model is that the survival times are conditionally independent given an (unobserved) frailty term (or random effect). Therefore, in the case of the PH model, given an unobserved frailty (random effect), the hazard for each survival time is assumed to follow a PH model, with the frailty effect acting multiplicatively on the baseline hazard. The joint survival function can then be obtained by integrating out the frailty (using an appropriate frailty density and its corresponding Laplace transformation – see details below).

There are a number of classes of frailty models (shared, additive, nested or joint frailty models). Brief descriptions of the shared frailty model and the joint frailty model are provided in the next sub-section, as these two models are the most relevant when considering the dependence between progression and survival outcomes.

In addition to the different classes of frailty, different frailty densities can be used; the most common being the gamma, positive stable or log-normal distribution. The gamma distribution is possibly the most commonly used due to its simplicity compared with other models. However, the log-normal model is more flexible than the gamma distribution, but requires the use of a more complex MLE procedure to estimate the parameters.

The baseline hazard can be non-parametric, semi-parametric or parametric. Parameters are usually estimated through the marginal likelihood via the marginal distribution (the likelihood integrated with respect to the frailty term).[113] As previously highlighted, frailty models were originally developed for PH models, with the frailty term (random effect) acting multiplicatively on the hazard, inducing positive correlation. For AFT models, the frailty is included as an error term. The frailty could be considered as an unobserved covariate that is additive on the log failure time scale and describes some reduced or increased event times for different clusters.[114] Similarly, multi-level mixed effects parametric survival analysis described by Crowther *et al* (2017)[100] is an extension of parametric frailty survival models, allowing any number of normally distributed random effects to be used, including the exponential, Weibull, and Gompertz PH models, and the Log-Logistic, Log-normal and Generalised Gamma AFT models.

6.9.2.2.1 The shared frailty model

The shared frailty model is appropriate when observations within a cluster share a common unobservable frailty. Typically, a single frailty is assumed, implying a positive correlation within each group. This can be relaxed by considering two independent frailty terms. Furthermore, the frailty assumes that the unobserved factors are the same within a group of clustered observations, and therefore the correlation is assumed to be constant between all individuals. Therefore, the shared frailty model is particularly useful in the context of the illness-death model where transitions share a frailty term.

The hazard function for the shared frailty for PH model can be written as follows:[96]

For the j-th (j = 1, ..., ni) individual of the i-th group (i = 1, ..., G),

$$\lambda_{ij}(t|v_i) = v_i\lambda_0(t)\exp\left(\beta^T X_{ij}\right) = v_i\lambda_{ij}(t)$$

where λ_0 is the baseline hazard; X_{ij} is the covariate vector associated with the vector of regression parameter β, and v_i the random effect for the i-th group.

For AFT,[114] the shared frailty can be expressed as a log-linear model for the logarithm of the event time such as:

$$\log \quad T_{ij} = x'_{ij}\beta + b_i + \sigma\varepsilon_{ij}$$

where β is a vector of fixed effects corresponding to covariate vector x_{ij}, σ is a scale parameter, the ε_{ij}'s are independent and identically distributed random errors, and the b_i's are the cluster-specific random effects.

6.9.2.2.2 *The joint frailty model*

In contrast to the shared frailty model described above (see Section 6.9.2.2.1), the joint frailty model[90, 96] can be used irrespective of whether the observations are clustered. This is particularly useful when an event may be terminated by loss to follow-up, end of study, or a major failure such as death. Therefore, the joint frailty model is particularly useful in the context of semi-competing risks.

The joint frailty model considers the joint evolution of two survival processes whereby one event (event 2) impedes the process of another event (event 1); treating the terminal event (event 2) as informative censoring.[96] It considers the dependency between the survival processes and respects that event 2 is a competing event for event 1.

Compared with the shared frailty model, in the joint frailty model, the frailty (υ_i), which links the two processes is assumed to act differently on the two survival functions. This is made possible by the introduction of an additional parameter, α.[90, 96]

The hazard functions for the joint frailty model, for the intermediate and terminal events can be written as:[96]

$r_{ij}(t|\upsilon_i) = \upsilon_i r_0(t) \exp(\beta_1^T X_{ij}) = \upsilon_i r_{ij}(t)$ (intermediate event)

$\lambda_i(t|\upsilon_i) = \upsilon_i^{\alpha} \lambda_0(t) \exp(\beta_2^T X_i) = \upsilon_i^{\alpha} \lambda_i(t)$ (terminal event)

where $r_0(t)$ is the intermediate event baseline hazard function and $\lambda_0(t)$ the terminal event baseline hazard function.

It should be noted that when $\alpha = 1$, the frailty has an identical effect on the risk of recurrent events and on the risk of terminal event. When $\alpha > 0$, the recurrent events rate and the terminal event rate are positively associated. Finally, when $\alpha = 0$ the terminal event is independent of the recurrent events.

6.9.2.3 *Application of the method to the joint modelling of PFS and OS*

A large number of applications of frailty models were identified. However, the large majority of applications of frailty models identified during this review process focused on: (i) the estimation of the frailty term; (ii) the estimation for the effect of covariates for PFS and OS rather than health state sojourn time, typically using non-parametric distributions, or (iii) the modelling of a recurring event and death.

Thus, this section focuses on a limited number of cases that were considered relevant to the prediction of PFS and OS and the estimation of health state sojourn time. Perhaps the most relevant example for

the use of frailty terms for the joint modelling of PFS and OS has been described by Dejardin *et al* (2010).[87] The authors proposed an illness-death model whereby PFS and OS are jointly modelled through the use of frailty term which is shared between TPP and PPS. The key assumption was PH between progression and death. Transitions shared a common parameter and therefore, the survival distribution for the PPS was partly determined by the survival distribution for TTP. Key characteristics of the model developed by Dejardin *et al* (2010) are: (a) it incorporates interval censoring, and (b) it assumes that death can only occur after progression (therefore, there is no direct transition from the PF state to death). The authors assumed piecewise marginal distributions (semi-parametric) and examined two frailty distributions: the gamma and positive stable frailty distributions. Overall, the authors reported that their model estimate for time to death was close to the non-parametric KM estimator, using either the gamma frailty or positive stable distribution, with the exception of the tail of the curve (which is important to consider in health economics). A similar model was also used by Rice *et al* (2017)[115] using a non-parametric distribution.

A similar approach (frailty shared between transitions in an illness-death model) was employed by Xu *et al* (2010)[116] using data from a RCT of nasopharyngeal cancer to estimate the effect of covariates. A frailty term (gamma distribution) was shared between all transitions (non-parametric maximum likelihood estimation [NPMLE]). The authors compared their model with: (a) the Markov model and (b) a restricted model (semi-competing model). They found that the Markov model and the general frailty model gave broadly similar results, both of which were more realistic compared with those of the restricted model. Similarly, Han *et al* (2014) proposed a Bayesian Markov Chain Monte Carlo methods (MCMC) for model fitting and a frailty term with normal distribution (in WINBUGS).

Examples of the joint frailty model have been identified mostly for the modelling of recurring events and death to estimate the effect of covariates; these have been described for examples by Liu *et al* (2015),[117] Rondeau *et al* (2007)[95] and Mazroui *et al* (2012).[92]

6.9.2.4 Applicability of the method to health economic evaluation

Frailty models are useful tools when modelling the dependence between two processes and have been extensively studied. Whilst frailties were originally developed for PH models, they can be extended to some AFT models. Furthermore, whilst not specific to the modelling of PFS and OS, frailty models can be used when considering the dependence between TTP/PFS and OS/PPS with;

- the shared frailty model, whereby TTP and PPS share a common unobservable frailty,
- the joint frailty model, whereby the dependence between TTP and OS is considered under a semi-competing risk framework, with TTP terminated by death.

A key limitation is that a single frailty induces positive correlation, which may not always be appropriate. This can however be relaxed by the addition of another frailty term. The use of a frailty term also assumes that the unobserved factors are the same within a group of clustered observations, and therefore the correlation is assumed to be constant between all individuals within that group. Furthermore, the marginal survival functions in the frailty model contain the association parameter of the frailty distribution,[118] and thus can be challenging to couple with other frailty models.

When considering the frailty model, in addition to the class of model, assumptions are also needed regarding the distribution of the frailty term (gamma, log-normal etc.), which specifies the type of dependence between the two processes. A key barrier for the adoption of frailty models (shared or joint) in health economics is the absence of clear step-by-step tutorial on how to jointly predict PFS and OS using frailty models. Whilst Dejardin *et al* (2010) reports an example of the use of a frailty model to jointly model PFS and OS, the code used to reproduce results is not available online. Strong assumptions were also made by the authors. The author was contacted, but no response was provided at the time of writing of this thesis. The only application identified in which the code was available in the appendices was from Han *et al* (2014). However, the approach is programmed in WINBUGS (statistical software for Bayesian analysis using MCMC methods), which is likely to be a barrier for many analysts responsible for development cost-effectiveness models in HTA.

It should be noted that there are a number of R packages that are available to fit general frailty models such as `frailtypack`, `frailtySurv` or `parfm` that could potentially be used to jointly model progression or survival outcomes. However, these packages do not provide clear examples for direct use in health economic analyses. The implementation is also likely to be different according to the parametric distribution assumed, the frailty distribution used and type of frailty model; this limits the immediate adoption of the approach in health economic evaluation.

6.9.3 Bivariate models: the Copula model

6.9.3.1 Development of the method

This method is not specific to the joint modelling of PFS and OS, but is a general approach to account for the dependence between two survival endpoints.

Sklar's theorem is the foundation principle for copulas (which establishes the connection between a joint d-dimensional distribution function and its univariate marginal distribution).[119] Copulas are a class of bivariate distribution whose marginal distribution function are uniform on the unit interval. Copulas can be described as functions which enable the combination of univariate distributions to obtain a joint distribution given a specified dependence structure. Therefore, any bivariate model can be considered to represent a form of copula.

6.9.3.2 Theoretical property

The copula model compared with the frailty model deals with the joint survival function whereas the frailty model is a conditional hazard model that has a multiplicative factor.[120] Indeed, Nelsen *et al* (2006) describe the copula, C, as a function that "*joins or couple multiple distribution functions* to their one-dimensional marginal distribution functions".[119] Sklar's theorem states that for a given joint distribution and univariate marginal distribution, there exists a copula function to couple them.[119, 121]

In mathematical terms[121], for $H_{X,Y}(x, y)$ - a joint distribution, $F_X(x)$ - the marginal distribution of X and $F_Y(y)$ - the marginal distribution of Y, there exists a copula $C(u, v)$ such that:

$$H_{X,Y}(x, y) = C(F_X(x), F_Y(y))$$

Each pair, (x, y) are associated with a point $(F_X(x), F_Y(y))$, in the unit square $[0,1] \times [0,1]$, corresponding to a number, $H_{X,Y}(x, y)$, in $[0,1]$.[121]

In the copula model, there are no constraints with the form of the marginal distribution (in contrast with the frailty model). This is because the correspondence is independent of the marginal distribution.

Copula are effective tool to obtain the joint CDF from individual marginal distribution. A key difference of the copula compared with frailty model is that the marginal survival function does not include the association parameter (the copula). In other terms, the marginal function remains unchanged, which is an attractive property as it is flexible and allows for the separation of the measure of dependence from the marginal distribution.

A copula having a domain of I^2 has the following properties.[121]

- It is 2-increasing[121]

For every u_1, u_2, v_1, v_2 in I such that $u_1 \leq u_2$ and $v_1 \leq v_2$, then

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Where I is the identity matrix.

- It is grounded[121]

For every u, v in I,

$$C(u, 0) = 0 = C(v, 0)$$

- Finally, for every u, v in I,[121]

$$C(u, 1) = u$$

$$C(1, v) = v$$

Several families of Copulas have been described, with the Archimedean copulas being the most popular to model the dependence in reliability engineering, medicine, climate and weather research, hydrology research. The Archimedean copulas (associative class of copula) is particularly popular as only one parameter is needed to govern the strength of dependence, it can be easily constructed and admit an explicit formula.[122] There are a varieties of Archimedean copulas such as Ali-Mikhail-Haq, Clayton, Franck, Gumbel, independence and Joe.[123, 124]

Another families of copulas often used is the Gaussian copula.[123, 125] As described in Wikipedia,[123] it is a distribution over the unit cube $[0,1]^d$ and constructed from a multivariate normal distribution over $\mathbb{R}^2$, by using the probability integral transform.

This family of copula has been used to jointly model PFS and OS by Fu et al.[89]

As described in Wikipedia,[123] for a given correlation matrix $R \in [-1,1]^{d \times d}$, the Gaussian copula with parameter matrix R can be written as:

$$C_R^{Gauss}(u) = \Phi_R(\Phi^{-1}(u1), \ldots., (\Phi^{-1}(ud))$$

The inferential procedure for the copula model follows two possible approaches;

- A two stage approach.[124] The marginal survival functions are first estimated. The estimates for the marginal distribution are then used to estimate the dependence parameters (copula function parameters) by maximisation of the likelihood with respect to the copula function. This is particularly useful when modelling the marginal survival functions in a semi-parametric or nonparametric way.
- Simultaneous MLE for all the parameters (the parameters of the marginal survival functions and the parameters of the copula). This is usually done when the marginal survival function is modelled parametrically. The simultaneous estimation would provide a different estimation for the baseline hazard accounting for the dependency.[120]

6.9.3.3 Application of the method to the modelling of PFS and OS

Copula models have been extensively studied and have typically been used to assess the surrogacy between PFS and OS in meta-analysis or used for evidence synthesis.[126, 127]

However, an example of the use of a copula to jointly predict PFS and OS was reported by Fu *et al* (2013).[89] The authors describe the use of a Copula function to study both the correlation structure between PFS and OS and to predict OS based on PFS. The authors proposed a normal induced copula estimation model and used a Gaussian copula to link the marginal distributions of TTP and OS, under a Bayesian framework.

In this model, PFS and OS are combined under a semi-competing risk framework, with OS (the terminal event) censoring the non-terminal event (PFS), with the copula acting as the dependence parameter between the terminal and non-terminal events. The authors proposed a normal induced copula estimation model and used a Gaussian copula (bivariate normal) to link the marginal distributions of TTP and OS, under a Bayesian framework. Both the marginal and copula terms are estimated simultaneously. The choice of a bivariate normal distribution is justified by the authors as it is simple to interpret and because the dependence can be determined by the correlation coefficient.

In terms of calculation of the likelihood and estimation of parameters, within Fu's model, parametric functions are fitted to both TTP and OS at the same time, with the likelihood function calculated based on the contribution of 4 type of individuals:

- those who both progressed and died subsequently
- those who progressed only but did not die

- those who died prior progression
- those who neither progressed nor died.

The total likelihood is therefore the sum of the likelihood estimated for these four type of individuals in the dataset, with parameters for both the marginal distributions (e.g. parameters for the Weibull for OS and Gompertz for TTP) and copula parameter estimated as the parameters that maximise the likelihood.

This contrasts with the independent model, whereby parametric functions for PFS and OS are fitted separately to each outcome, with the likelihood function for both events only incorporating information on patients who had the event of interest or were censored.

Fu *et al* (2013)[89] performed a simulation study to determine whether the performance of the method in terms of OS prediction was improved using joint modelling compared with the independent model (direct fit to PFS and OS). The authors concluded that:

- when there is no correlation in the data ($\rho = 0$), the copula model performed less well (in terms of mean squared error and biases) compared with the working independent model. This is justified by the author because of the introduction of an additional parameter,
- both the copula model and working independent model provide an unbiased estimate of the median OS and are not significantly different from each other (irrespective of ρ). However, the authors note that the variance with the NICE model is usually smaller, therefore providing a more accurate estimate of median OS,
- when $\rho \neq 0$, the copula model performs better (i.e. it is more accurate estimate and unbiased) for TTP. This is because TTP can be dependently censored by OS, and therefore the correlation parameter ρ becomes important.

Fu et al (2013)[89] also demonstrate how to estimate OS and PFS given that their model is based on TTP and OS. This could be summarised in three steps:

- Step 1: Obtain M posterior samples for ρ and the hazard functions of TPP and OS
- Step 2: for each k (sample), generate TTP and OS as $\{T_{1i}, T_{2i}\}_{i=1}^{10,000}$ on the basis of $\{\lambda_{TTP,k}, \lambda_{OS,k}, \rho_k\}$
- Step 3: derive PFS and OS such as $\{\min(T_{1i,}T_{2i}), T_{2i}\}_{i=1}^{10,000}$

The code used by the authors is available in the supplementary appendix and therefore, the model can be used by other analysts. Although it is unclear whether the Fu's model example was used, as this was

only available in abstract form, Felizzi *et al* (2018)[128] conducted a study using a Gaussian copula (bivariate normal) to link OS and PFS survivals. Another relevant example for the use of copulas is provided by Rotolo *et al* (2013). Rotolo *et al* (2013) used Copulas to generate simulated data within an MSM that would include: (a) the dependence of time for successive time, and (b) the dependence of time for competing events. Clayton copulas were used to induce the dependence for the competing event and another Clayton copula was used to induce the dependence for the successive transitions.

6.9.3.4 Applicability of the method to health economic evaluation

Copulas are versatile and are suitable methods for use in health economics to induce the dependence between survival outcomes. Copulas are equally useful when including the dependence of time for successive events or semi-competing events. A key strength of copula is that the marginal distribution remains unchanged and therefore it is possible to couple different copulas to induce different levels of dependence, as shown by Rotolo *et al* (2013).[129] In the copula model, there are also no constraints with the form of the marginal distribution (in contrast to the frailty model). This is because the correspondence is independent of the marginal distribution. This is a key strength.

Fu *et al* (2013) provide an example of implementation (R code) for the Gaussian copula for the joint modelling of PFS and OS. Whilst the exponential distribution is used, the code provided by the authors can be easily amended by the user to any parametric distribution (as shown in the simulation study in Appendix 12). Therefore, the method by Fu *et al* (2013) using a Gaussian copula to jointly model PFS and OS under a semi-competing risk can be easily and immediately adopted in health economic analyses. For instance, the same model was used recently by Felizzi *et al* (2018)[128] However, it should be noted that a large number of other copula functions exist, which would require a different implementation and formulation. The Gaussian copula, whilst simple, may not always be the most appropriate copula for the data (depending on the tail of the distribution of the data). The Gaussian copula is said to be tail independent,[130] thus a key limitation for this type copula is the inability to capture extreme values. Other copulas may be more appropriate. However, from a practical point of view, it is not feasible to look at all copula functions. Researchers suggest that the choice of copula is often based on familiarity, ease of use and analytical tractability.[122] The choice of copula function could also be informed by goodness-of-fit tests, such as the Akaike Information Criterion (AIC).[131] Exploring the type of dependency pattern exhibited within the data can also be useful to limit the number of copula functions examined.

6.9.4 Semi-competing risk by means of first passage times of a stochastic process

This method[98] was originally developed for engineering problems and is an extension of threshold analysis and the single-threshold model proposed by Paroissin and Salami (2014). The approach proposed by the authors models terminal and non-terminal event under a semi-competing framework.[98]

The description presented in this section is very brief, as I wasn't able to understand properly this method based solely on the description provided by the authors. No software for implementation was also provided, meaning I could not check or confirm my understanding using an example of implementation in software.

In brief, the two events are coupled through the introduction of a threshold. A threshold (c) is assumed for the terminal event at time X. A threshold (S) is assumed for the non-terminal event at time Z.

The authors propose a model whereby the time to the terminal event (X) is the first passage time to a fixed level *c* in a stochastic process, while the time to the non-terminal event (Z) is represented by the first passage time of the same process to a stochastic threshold *S*, which is assumed to be independent of the stochastic process.

The method was applied in a simulated dataset (data generated using the method examined itself [threshold model]) .[98] The authors showed that estimated parameters using MLE were close to the true parameters in their simulated data.

The authors also applied their method to a case study in bone marrow transplantation. The threshold model was first fitted to the data assuming S follows a log-normal distribution. Following examination, the authors found that a normal distribution was more appropriate. Parameters for S and the underlying process $D(t)$ (gamma process) were then estimated using MLE.

The authors state that the parametric curves seem to fit fairly well to the non-parametric curve (I believe this is debatable when examining the fit against the KM), but less well compared with that in the simulated data (which were simulated using the method itself). The authors also compared their estimate for the marginal distribution for Z (the non-terminal event) against estimation by Fine *et al* (2001) and state that their predictions using their model was within the confidence interval estimated by Fine *et al* (2001), but did not reflect the plateau at the end. Agai,n I believe this is debatable.

While this approach[98] to deal with semi-competing risks is interesting, its applicability to health economics is likely to be very limited given the strong assumptions upon which the model relies. In particular, both events are assumed the follow same underlying process ($D(t)$). As acknowledged by

the authors this may not always be appropriate. In particular, the process may change when the non-terminal event occurs. Although this is debatable, the approach did not appear to fit the data well (according to my interpretation) even when data were simulated using the approach itself (and approach subsequently used estimate the underlying parameters).

Perhaps more importantly, I wasn't able to fully comprehend this method based on the description in the paper. No implementation in a software is provided, and therefore I had to rely on the description in the paper; the same challenges are likely to be face by most analysts in health economics. The absence of the code in a suitable statistical package to replicate results is an important limitation to its use and quick adoption to health economics.

6.10 Discussion and conclusions

This chapter aimed to identify and summarise approaches that could be used to jointly model progression and survival outcomes in order to estimate health state sojourn times in health economic models for anticancer therapies. A systematic review process was employed, using an iterative approach, in order to account for challenges associated with searching and reviewing the methodological literature. Only one reviewer screened and extracted results, which is a limitation. Despite some overlap, identified methods could be categorised according to two groups; (1) methods that include the dependence between transitions in an illness-death model (joint/conditional modelling of transitions) and (2) methods that include the dependence between PFS and OS under a semi-competing risk framework. These could be further separated into four broad categories; (i) general extensions of the MSM; (ii) methods where the dependence is induced by a frailty/error term; (iii) methods where dependence is induced using a copula or bivariate model, and (iv) the first-passage method.

All of the approaches identified within this review are subject to certain limitations. The extensions of the multi-state method identified during the review process are limited to specific parametric distributions, typically the exponential and Weibull.[86, 91] This may therefore be a strong assumption which may not be consistent with the underlying distributions from which transitions are drawn. An additional recent paper[93] was identified describing a method to jointly model PFS and OS (using the multi-state framework); however, details are currently insufficient for its immediate adoption in health economic analyses.

The frailty model, which is typically used for PH models, allows for the induction of the dependence between two survival outcomes. This can be extended to some AFT models using an error term. However, this requires transformation. There are several types of frailty models, with the shared frailty and joint frailty models being possibly the most relevant to the health economics context. Within the shared frailty approach, the dependence can be induced between consecutive transitions (for instance between TTP and PPS).[87, 95, 96, 132] The joint frailty model is useful in the context of semi-competing risks whereby a terminal event (OS) censors a non-terminal event (TTP). The introduction of a frailty in the shared frailty model typically induces a positive correlation, as this acts multiplicatively on the hazard. However, this can be relaxed by the introduction of a second frailty term. In addition to the type of frailty, there are different distributions for the frailty term that can be used, including the gamma or log-normal distribution.

The copula model is an alternative to frailty model.[89, 118, 121, 129, 133] Copulas are essentially a type of bivariate model. Copulas are perhaps more flexible compared with the frailty model, as there is no constraint on the choice for the marginal distribution. Compared with the frailty model, the marginal distribution in the copula model also remains unchanged. The copula model also induces both positive and negative correlation. Similar to the frailty model, copula models can be used to induce the dependence between consecutive transitions (such as TTP and PPS) or to deal with a semi-competing risk situation whereby a terminal event (OS) censors a non-terminal event (TTP). OS and TPP are therefore estimated jointly, accounting for their dependence. Despite copulas being flexible, a large number of copulas exists, and the choice between copulas can be difficult. This can be informed by statistical goodness-of-fit.

In Chapter 8, the performance of approaches commonly used in health economics will be examined in addition to alternative approaches that could be employed to jointly model progression and survival outcomes and to estimate health state sojourn time. A key consideration when examining a method is whether the method is likely to be adopted in the first place. Given the need for transparency and technical skills of analysts typically in charge of building or reviewing cost-effectiveness models, a method is unlikely to be adopted if there are no examples/tutorials on how to implement the method in a suitable software package. Even with a tutorial (and examples available in a suitable package),[45, 50] adoption is not guaranteed as illustrated by the slow adoption of the MSM. This is also reflected by the restriction on the software use when companies submit to NICE. Consequently, for pragmatic reasons, only methods that provided code or a clear tutorial for its implementation in a suitable package were included in Chapter 8.

Whilst a number of these identified methods have potential, these are unlikely to be adopted immediately to health economics, either because they need to be further developed or a thorough tutorial is required to guide analysts on how to implement these approaches (in particular given the different possible formulation). The model proposed by Li *et al* (2015); a MSM using Weibull distributions with transitions sharing a common parameter, and the model proposed by Fu *et al* (2013); Bayesian Gaussian copulas are methods that could be immediately adopted in health economics and therefore will be included in the next stage of this thesis.

Despite its high degree of relevance, the approach proposed by Meller *et al* (2019)[93] is unlikely to be adopted immediately and easily in health economics given the insufficient information to implement the approach (the code is not available in the final version of the published article and was not provided upon request to the author). The approach also did not show to be particularly superior to the Li' model in a simulation study conducted by the author. A number of general methods are available to jointly

model two survival outcomes (PFS and OS) under a semi-competing risk framework. Of particular interest are the frailty model and the use of a copula. Both are widely used in the field of statistics, in particular for evidence synthesis, bivariate analysis or the modelling of a recurrent event and a terminal event. There are however a large number of copula and frailty models. Although the volume of literature on this topic is large, none of the models seem to hold any clear advantage compared to the other. There are also a large number of different specifications, making any direct comparisons impossible. It should be noted that a comparison of the copula and frailty model is outside the scope of this thesis, and therefore, the focus in this thesis is on exploring whether the joint modelling of PFS and OS could potentially improve predictions compared with the separate fit (as currently done in the PSM) and therefore focusing on the Fu *et al* (2013)[89] was deemed reasonable as this approach was also used by Felizzi *et al* (2018).[128] Despite the approach being limited to one copula distribution (Gaussian), this approach could be easily and immediately adopted in health economics. The use of other forms of copula would require a different formulation. The Gaussian copula is tail independent in that it may not capture extreme values, which is likely to be appropriate when considering PFS and OS. No papers were identified providing a clear example of the implementation of the frailty model under a semi-competing risks framework, despite some packages being available. Whilst it could have been interesting to examine both copula and frailty models in Chapter 8 when jointly modelling PFS and OS under a competing risk framework, I had to be pragmatic in this thesis and focus on approaches that are most likely to be adopted and which have been implemented in a suitable statistical package (for conclusions to be helpful for decision-makers and analysts, rather than conclusion to be limited to a research exercise). Whilst this could be considered as a possible limitation, using the Fu model in Chapter 8 (given the availability of a clear tutorial) could inform future research in health economics on whether predictions are improved sufficiently to justify using complex models (given the different formulations) which may make decision-making more complicated. Indeed, at present, decision-making is already challenging by the choice of parametric functions. If different approaches/formulations were also to examined, this would introduce additional challenges.

In the next chapter, I describe the possible limitations/biases associated with the use of PPS estimated in the subset of patients who experienced progression to reflect the overall trial population when developing a model based on information collected in an RCT only.

7 CHAPTER VII: LIMITATIONS ASSOCIATED WITH THE USE OF POST-PROGRESSION SURVIVAL (PPS) WHEN DEVELOPING A MODEL BASED ON INFORMATION COLLECTED IN AN RCT ONLY

7.1 Chapter overview

In this chapter, I describe the potential biases associated with the use of PPS estimated only in a subset of patients who progress when generalised to the overall randomised population in the same randomised control trial (RCT). This may be an issue for any STM.

Section 7.2 introduces this Chapter. In Section 7.3 I discuss the possible limitations and illustrate those using hypothetical simulated data. In Section 7.4, I use real datasets to illustrate the possible limitations. Finally, in Section 7.5, I discuss a possible simple approach to reduce potential biases associated with PPS and test this within a real dataset to assess whether it improve predictions of OS.

7.2 Introduction

As described in Chapter 4.5, within the STM approach (including MSMs), OS is estimated indirectly through the explicit modelling of every transition between health states. This is an attractive approach, as this allows for a more natural and explicit modelling of the natural history/underlying disease progression process of cancer. This contrasts with the PSM which does not involve modelling the disease process, but instead involves directly fitting parametric functions to data on PFS and OS. Modelling the underlying disease progression process appears more naturally appropriate rather than fitting a curve directly to OS. However, when developing a model based only on information collected in an RCT, transitions are not observed for all randomised individuals, which could introduce a series of biases.

Indeed, when developing a model based on information collected in an RCT only, the transitions from the PF state to the PD state (using TTP) or death states (using PrePS) are estimated using data relating to all randomised patients. Despite data not being fully complete due to censoring (administrative or random), (right) censoring is assumed to occur at random. Consequently, the estimate for the transitions from the PF state (to progression or death) can be considered ‘unbiased’ as this is estimated amongst all randomised individuals in the trial. The likelihood function for the survival function uses information both from patients who have had the event and from those who are censored. It should be noted however that TTP/PFS would be biased if censoring was not random.

Conversely, when developing a model based on information collected in a single RCT only, the transition from the PD state to the death state (PPS) is estimated only in the subset of patients who progressed during the observed period of the trial; thus, only a subset of randomised patients contributes information to the likelihood function for PPS, and those who did not progress within the observed trial period are excluded from the estimation of this transition.

Consequently, whilst the estimate for this transition could be considered unbiased within those who progressed, the estimate could be biased when representing the overall population (those who progressed and those who did not yet progress) for two reasons. First, patients who experienced progression by the time the trial ended (also referred as early progressors) may have a different survival prognosis following progression compared with patients who had not yet progressed by end of the trial (from which data are not observed). Secondly, those who progress later are more likely to be censored in the PPS dataset. This would suggest that there is time-dependent bias in the form of informative censoring.

This chapter aims to: (a) demonstrate and describe potential biases associated with the use of PPS estimated in the subset of patients who experienced progression when developing a model based on information collected in an RCT only; (b) describe some of the implications of these biases and; (c) discuss the advantages and disadvantages of some of the approaches that have been suggested to adjust the estimated PPS when developing a model based on information collected in a single RCT.

7.3 Potential limitations associated with the use of PPS when developing a model based on information collected in an RCT only

As only the subset of patients who experienced disease progression within the observed period of the trial are included in the analysis of PPS, this estimated transition may not be generalizable to the overall population. This may be the case if patients who had progressed within the observed period of the trial experience a faster or a slower time to death following progression compared with those patients who had not progressed at the data cut-off for the trial.

In order to illustrate this problem, hypothetical trial data were generated assuming 3 levels of dependence (negative dependence, independence and positive dependence) between TTP and PPS, whereby the truth (i.e. the complete data in the absence of censoring) is known, in addition to the same data assuming censoring reflects early termination in trials.

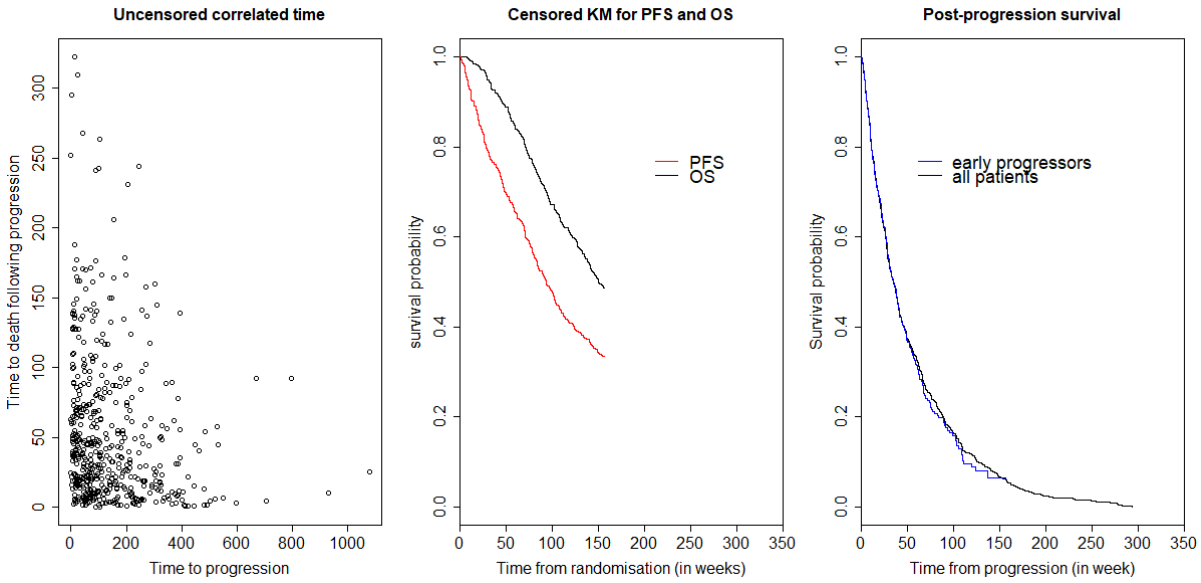
The process used to generate data is described below. Hypothetical uncensored time-to-event data were first generated for PFS and OS using a simple multi-state process for 500 individuals. TTP and PPS

were assumed to follow an exponential distribution, with rate parameters of 0.03019738 and 0.082085, corresponding to mean sojourn times of 144 weeks and 53 weeks, respectively. No death was assumed prior to progression; thus, PFS equals TTP. A Gaussian copula was then used to link TTP and PPS to induce three levels of dependence (moderate negative, independent and moderate positive). A moderate dependence was defined by a Kendall's Tau of 0.5 for illustration.[134] Once uncensored survival times were generated, administrative censoring was introduced at Week 156 for all three scenarios, so that patients with an uncensored TTP time or death time greater or equal to 156 weeks were censored at this time. This was done to reflect early termination in trials and to compare whether the KM for PPS estimated in all randomised patients estimated using uncensored survival time is similar to the KM for PPS estimated only in the subset of patients who had a recorded progression event by the trial cut-off. For simplicity, no random censoring was assumed to occur before the administrative censoring time-point in this illustration.

KM plots for the scenarios of independence between TTP and PPS, negative moderate dependence and positive moderate dependence are shown in Figure 26, Figure 27 and Figure 28, respectively. The simulated uncensored correlated time for TTP and PPS (in all randomised patients – complete dataset) are shown in the left-hand panels, the generated censored PFS and OS KM are shown in the centre panels, and estimates for PPS are shown in the right-hand panels of the figures.

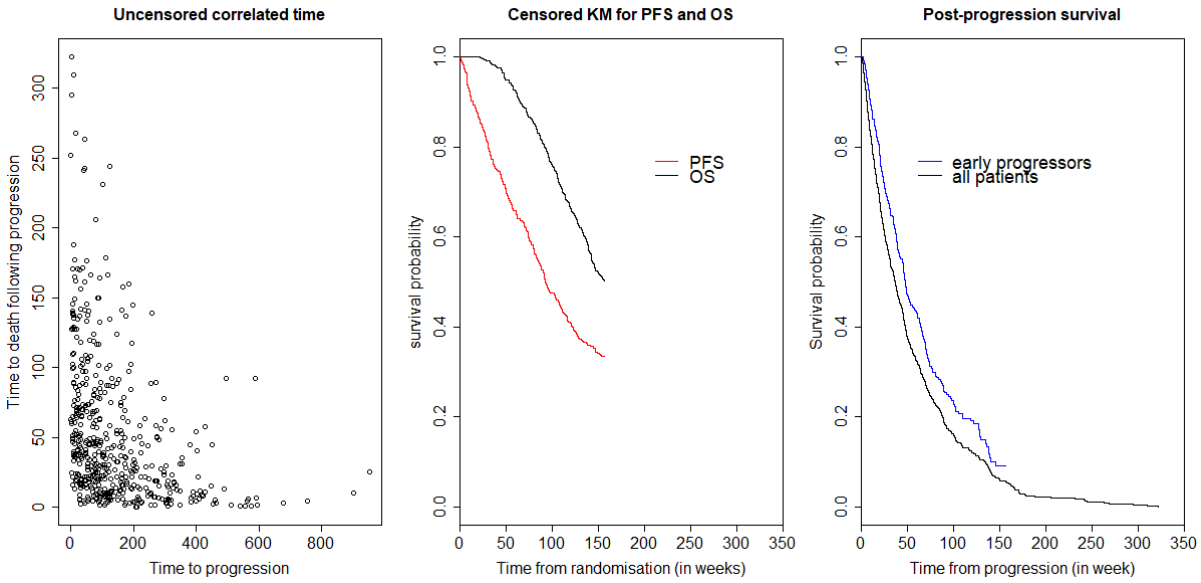
Unsurprisingly, when TTP and PPS duration times are independent / not correlated (Figure 26), the KM for PPS estimated in the censored dataset only in the subset of people who progressed before Week 156 (blue line) is relatively similar to the KM for PPS generated from the respective uncensored dataset (amongst all randomised patients in an uncensored dataset). This could be explained by the fact that the time to progression does not affect the time to death following progression. In this case, it is reasonable to assume that PPS estimated in the subset of patients who progressed early is generalisable to the overall population.

Figure 26: Data generated assuming independence between PFS/TTP and PPS



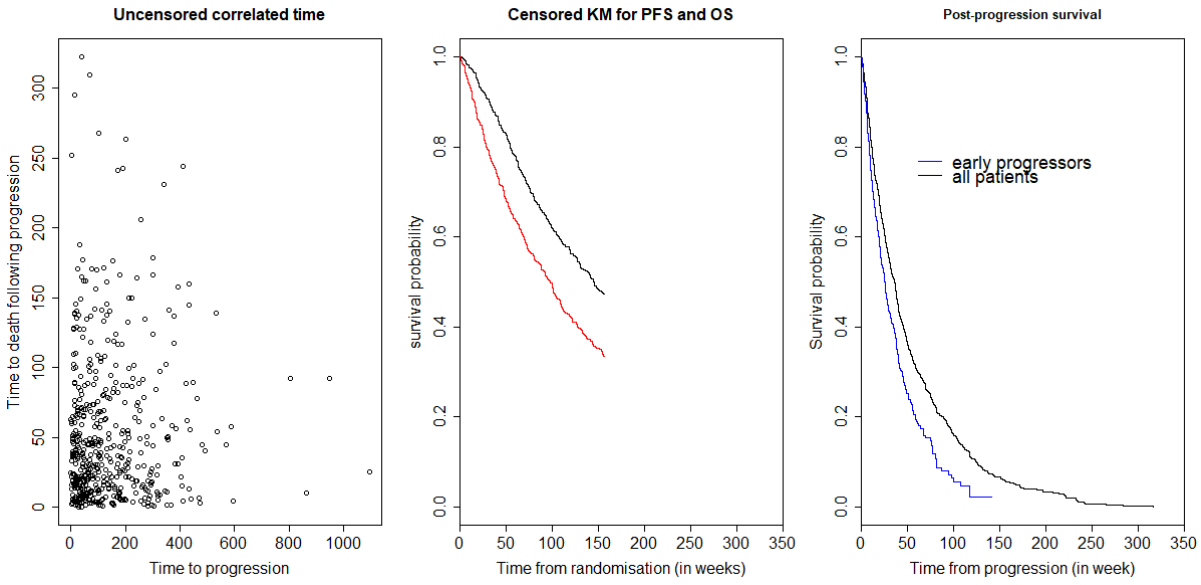
When TTP and PPS duration times are moderately negatively correlated (Figure 27), the KM plot for PPS estimated for early progressors (before Week 156, blue line) is significantly higher compared with the KM for PPS estimated in all randomised patients in the uncensored dataset. This is intuitive because in the case of negative dependence, patients who progress early have a longer time to death following progression.

Figure 27: Data generated assuming moderate negative between PFS/TTP and PPS



Conversely, in case of moderate positive dependence (Figure 28), the KM plot for PPS in the early progressors (before Week 156, blue line) is significantly lower compared than the KM plot for PPS estimated amongst all randomised patients in the complete dataset. This is intuitive because in the case of positive dependence, patients who progress early have a shorter time to death following progression.

Figure 28: Data generated assuming moderate positive between PFS/TTP and PPS



Whilst simple and intuitive, this example using hypothetical data illustrates the general concept that using the PPS estimated in the subset of patients who progress may not be generalizable to the overall population when TTP and PPS are not independent.

Three levels of dependence were examined (negative, independence and positive) to reflect the possible scenario that could possibly be encountered in HTA. In reality, data could exhibit any degree of dependence (either positive or negative). The nature of the dependence is likely to be different between cancers, but also within line of treatment for the same cancer. Negative dependence could happen for instance when the introduction of a new treatment has a positive effect on PFS but not on OS. Positive dependence is relevant when delaying progression is associated with longer PPS.

Furthermore, usually, fewer people receiving the intervention (new treatment) progress by the time the trial ends, compared with the control group. Therefore, potential biases are likely to impact on the two treatments to different degrees and could work in opposite directions.

7.4 Illustration of potential biases associated with the use of PPS estimated in a subset of patient using real trial data

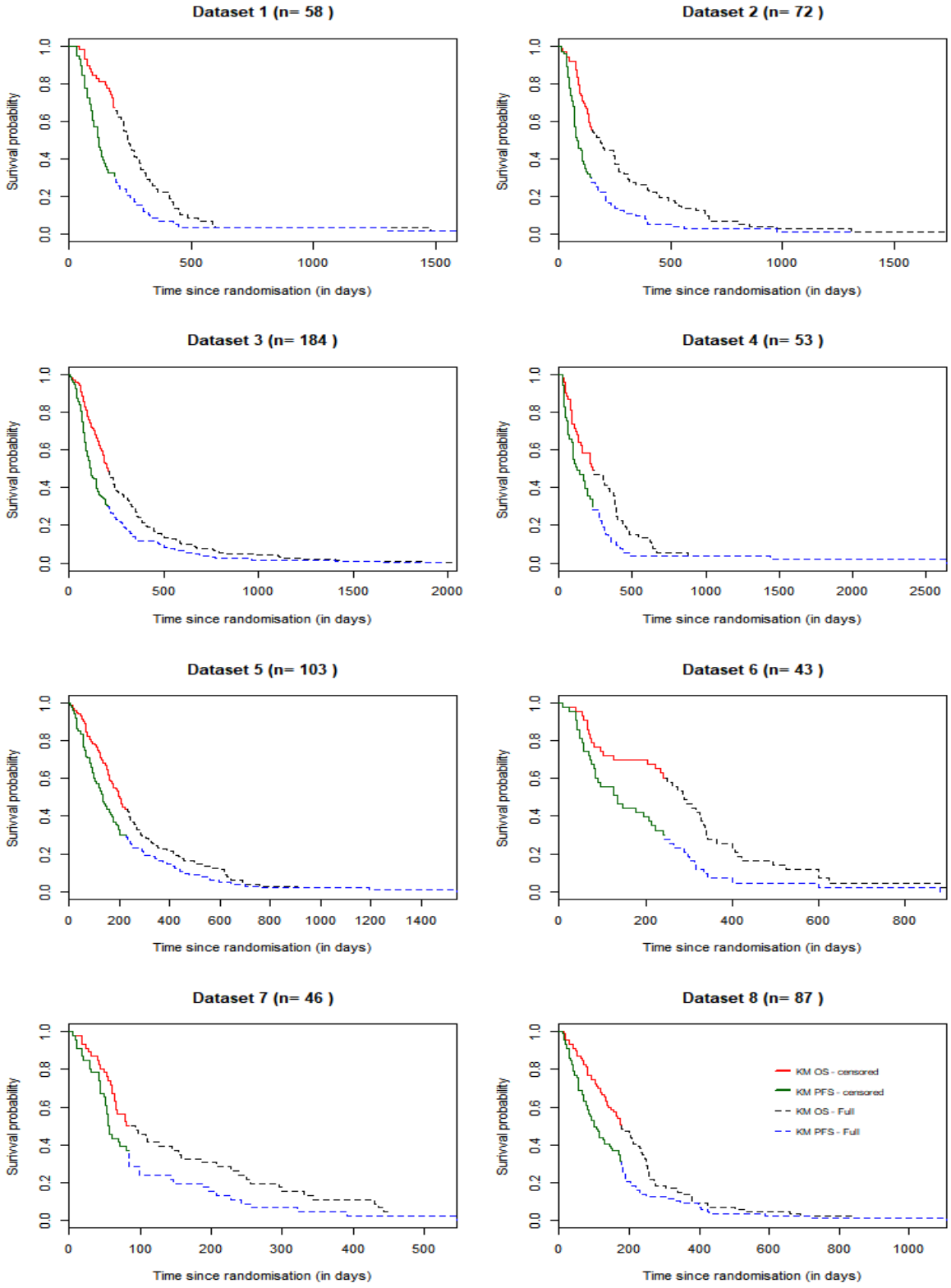
There are limits with using hypothetical data (used in Section 7.3) as the degree of dependence assumed could be deemed unrealistic and these types of dependence may not be observed in real datasets. To further illustrate this issue and to overcome potential limitations associated with the use of hypothetical data, the validity of assuming that PPS estimated in a subset of patients is generalizable to PPS for the overall randomised population was examined in eight trial arms in gastric cancer in which data on both PFS and OS were complete (i.e. all randomised patients had recorded progression and death times).

These trial arms were selected as they had complete information (everyone in the trial progressed and died, and therefore, information is available for all randomised individual rather than a subset). Furthermore, these trial arms were selected as they were part of a broader set of 20 trials in gastric cancer used in a meta-analysis.[135-137] The data are publicly available in the R package `surrosurv` (gastadv dataset) provided by the GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) group.

Compared with using hypothetical data, using real trial data allows an assessment of whether using PPS estimated in the subset of patients who progressed is generalizable to the overall randomised population without having to make potentially arbitrary assumptions regarding the dependence between TTP and PPS.

A total of eight trial arms (4 control and 4 experimental arms) from 5 RCTs included in the meta-analysis[135-137] were used in this analysis; two trials arm were excluded as there was some censoring. The KM plots of PFS and OS from these trial arms is shown in Figure 29. The sample size of patients in each arm varied from 43 to 184. Although the sample size for some trial arms was small, trial arms were not pooled. This is because whilst trials were conducted in gastric cancer, the intervention and control arms assessed were different between trials. Pooling datasets would therefore introduce biases given that different trial arms may exhibit different behaviour in terms of PPS.

Figure 29 : KM for PFS and OS for trial included in the analysis (Gastric cancer)



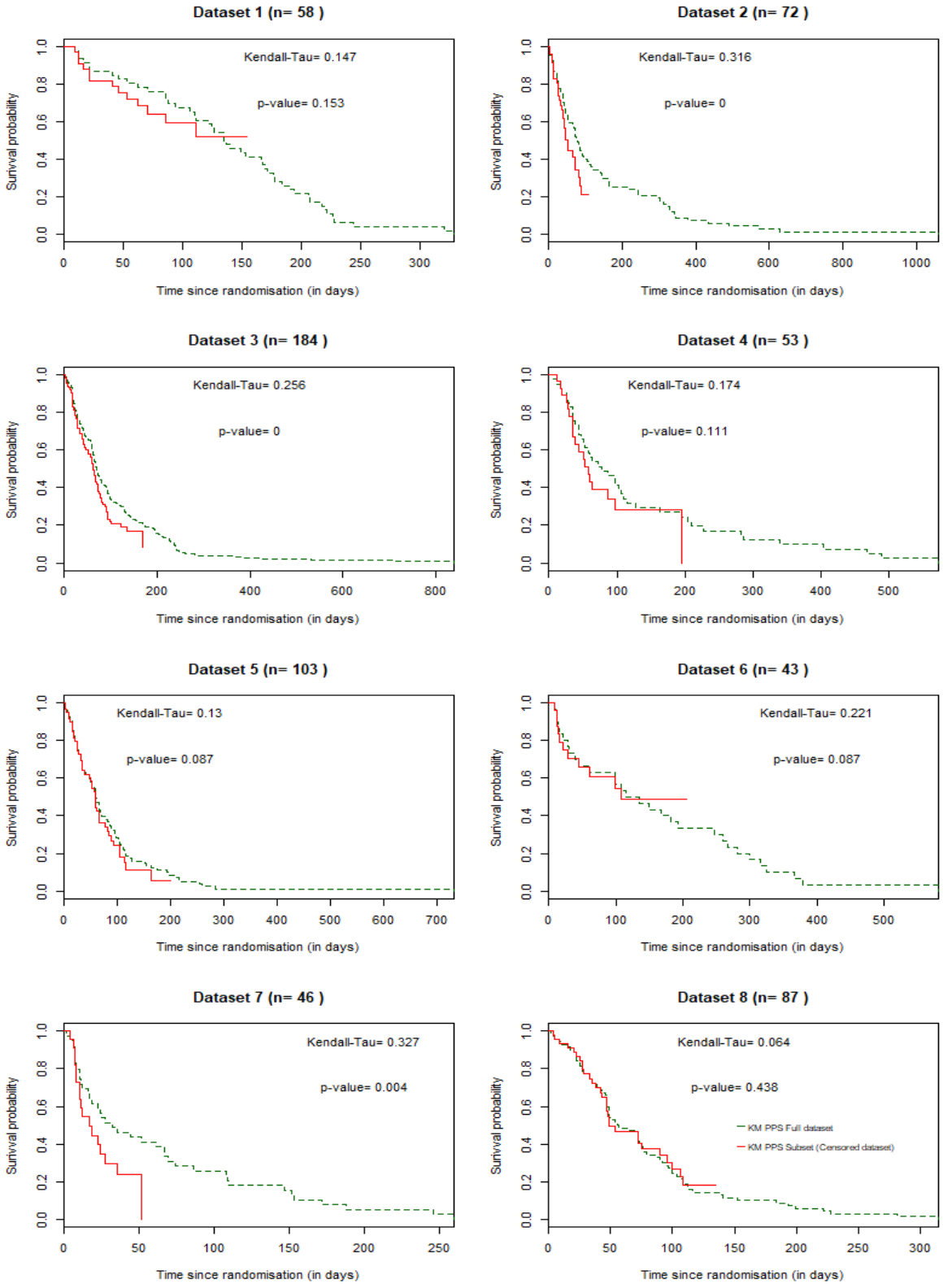
The correlation between PFS/TTP and PPS (excluding patients who had the same recorded times to progression and death) was computed for each trial arm using the Kendall Tau test, in order to understand whether the complete data exhibited any type of dependence (Figure 30). The Kendall's Tau in the trial arms examined ranged between 0.067 (p=0.438) indicating no dependence to 0.327 (p<0.005) indicating some small to moderate positive dependence between TTP and PPS.

In order to examine whether people who progressed early have the same risk of death as all randomised patients, administrative censoring was arbitrarily introduced when 70% of PFS events occurred. Therefore, for each complete dataset, an equivalent dataset was created whereby patients who progressed or died after a specified time-point were censored. The KM plots for PFS and OS including censoring are shown below in Figure 29 alongside the complete KM. A comparison for the PPS estimated in the complete datasets amongst all randomised patients against the KMs estimated in the truncated datasets (censored datasets) is shown in Figure 30 alongside the Kendall Tau (estimated the complete dataset) for the dependence between TTP and PPS.

With the exception of Dataset 5, the median PPS in the censored datasets were consistently under-estimated (to varying degrees) compared with the PPS observed in all randomised patients. Visually, the PPS estimated in the subset of patients who progressed was worse than that estimated in the full randomised population in five of the 8 datasets examined. It can also be seen that the degree of dependence was more pronounced in some trial arms compared with others. This potentially suggests that data could exhibit any degree of dependence, even within the same cancer type. Differences could also be attributable to differences in mechanism of action or differences in sample size. Although not shown here, as expected, differences in PPS between all randomised patients and the subset of patients who progressed were more pronounced when the level of censoring was greater, but less pronounced when the degree of censoring was lower.

This simple illustration confirms that using PPS estimated only in a subset of patients could introduce biases in that it may not be generalisable to the overall population. It should be noted that only trial arms in gastric cancer were used, and that different degrees of variation could be observed if alternative datasets were used or if different censoring assumptions were assumed.

Figure 30 : Comparison of PPS in the complete and censored datasets (Gastric Cancer)



7.5 Adjusting the time to death following progression, by making it conditional on time to progression (covariate in a statistical model)

Because of the selection biases described above and following informal discussions with analysts working in HTA, when only data from the trial are used (e.g. no external evidence is included in the model), analysts may sometimes consider adjusting PPS by making it conditional on TTP. This may be done in order to: (i) reduce possible biases and make PPS more generalizable to the overall population; (ii) account for the trend in the data during the observed period, and; (iii) extrapolate that trend beyond the observed period of the study.

The approach typically considered in HE consists in using TTP as a covariate in the statistical model when estimating PPS. Whilst different formulations are feasible and have been proposed, analysts may typically consider two approaches:

(i) making PPS conditional on TTP on the log scale,
(ii) making PPS conditional on TTP on the normal (non-log) scale.

It should be noted that different formulations are also feasible, for instance using the square of TTP, or more complex formulations. These two formulations were selected here to illustrate the impact of adjusting PPS using the two most common (simple) formulations. In this section, I will re-use the eight trial arms in gastric cancer described in Section 7.4 to explore whether making PPS conditionally dependent on TTP (in the log and non-log scale) reduces selection bias. The following steps were taken:

- first, a statistical model is estimated for PPS using TTP as a covariate (on the log and non-log scale) in each censored dataset.
- second, the statistical model is then used to generate predictions for PPS based on: (a) the number of patients who progressed in each cycle, and (b) the estimated PPS for a given cycle (i.e. patients who progressed in cycle 1 will be assigned a different PPS compared with patients who progressed at cycle 30, based on the covariate in the PPS statistical model). This can be done both at a cohort level (using tunnel states) or at the individual patient level.
- in order to compare the lifetime predictions, parametric distributions were fitted to both TTP/PFS and PPS. This was required in order to generate predictions for PPS which were conditionally dependent on TTP beyond the observed period. The resulting PPS was then generated from the proportion of people who were expected to die over time.

Given that the statistical model for PPS uses TTP as a covariate, any predictions for PPS using this statistical model will be linked to the predictions for TTP/PFS. This is because, dependent on the

prediction for TTP (i.e. when people progress), the effect of the covariate will be different at the individual level when predicting PPS.

Given that the prediction for TTP would impact on the prediction for PPS, TTP/PFS was fitted to the complete datasets without censoring in order to reduce any misspecification of PFS/TTP if curves were fitted to the censored data. This was necessary given that the approaches considered use TTP/PFS as a function of PPS, and therefore any misspecification of TTP/PFS could lead to a misspecification of PPS. Therefore, the cases presented here are likely to represent a best case scenario. Seven parametric distributions were considered for TTP/PFS (exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma). For each dataset, the parametric distribution with the best goodness-of-fit in terms of AIC/BIC was selected for TTP/PFS (as this was fitted to the complete datasets). The fit to PFS is shown in Appendix 6 for each dataset.

Statistical models were estimated for PPS (with or without covariates) to the censored datasets to reflect information typically available from a trial when PPS is fitted to the data. Seven parametric distributions were also considered for PPS (exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma). However, results for the exponential distribution are not reported here as these were similar to those for the Weibull distribution. Results are presented for each arm separately and for each distribution for PPS. This was necessary to avoid misinterpreting results due to the use of an inappropriate parametric distribution.

Results for each trial arm are presented in Figure 31 to Figure 38. For each arm, the KM plot for PPS estimated amongst all randomised patients (black line) and the KM plot for PPS estimated in the subset of patients who progressed early (red dashed line) are plotted against: (a) the unadjusted predicted PPS fitted to the censored dataset (blue line); (b) the adjusted predicted PPS conditional on the log of TTP fitted to the censored dataset (represented by the dark green line), and; (c) the adjusted predicted PPS conditional on TTP (on the non-log scale) fitted to the censored dataset (represented by the orange line). It should be noted that Gompertz distribution could not be fitted in a number of datasets.

Figure 31 : Predictions for PPS using adjusted and unadjusted method in Dataset 1

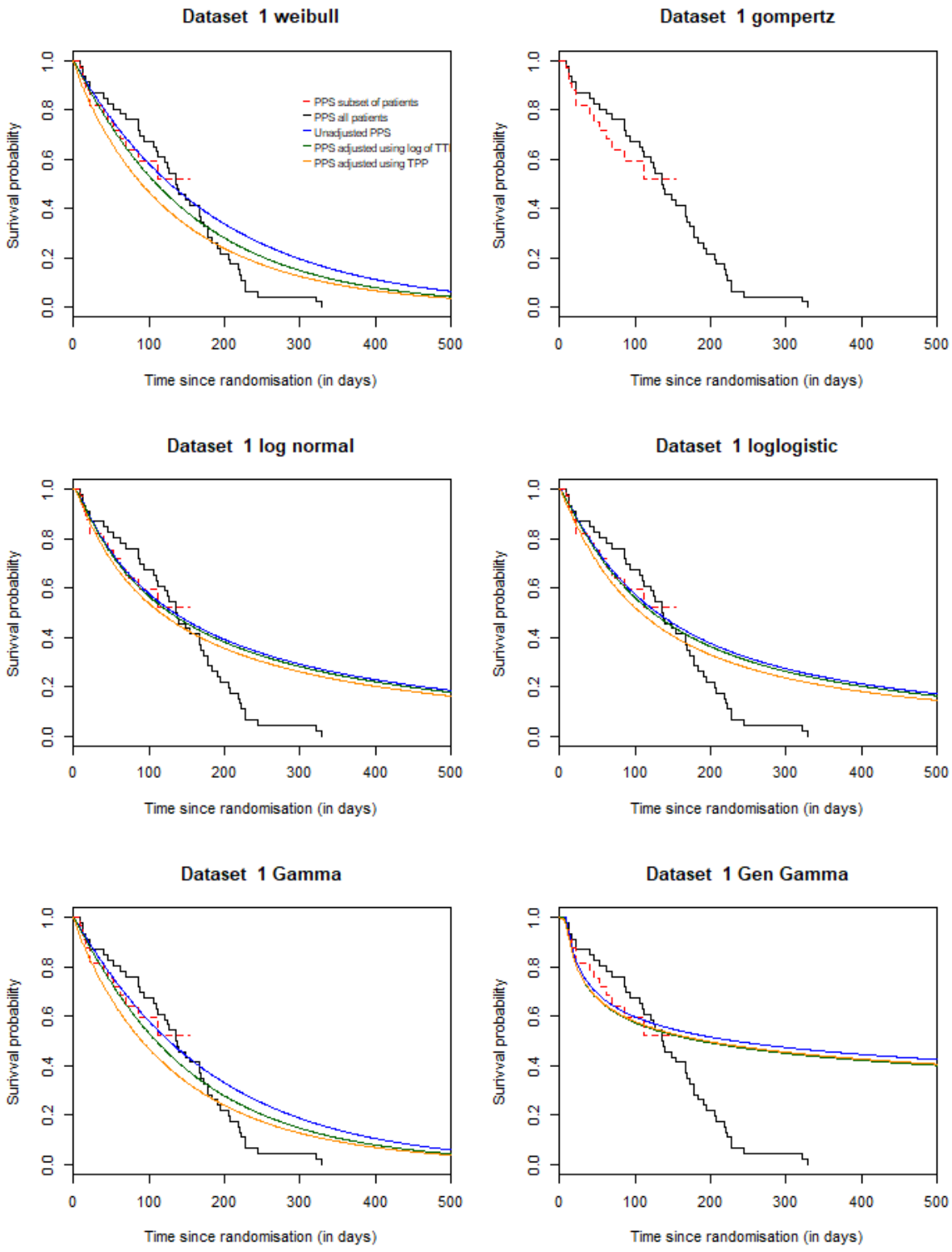


Figure 32 : Predictions for PPS using adjusted and unadjusted method in Dataset 2

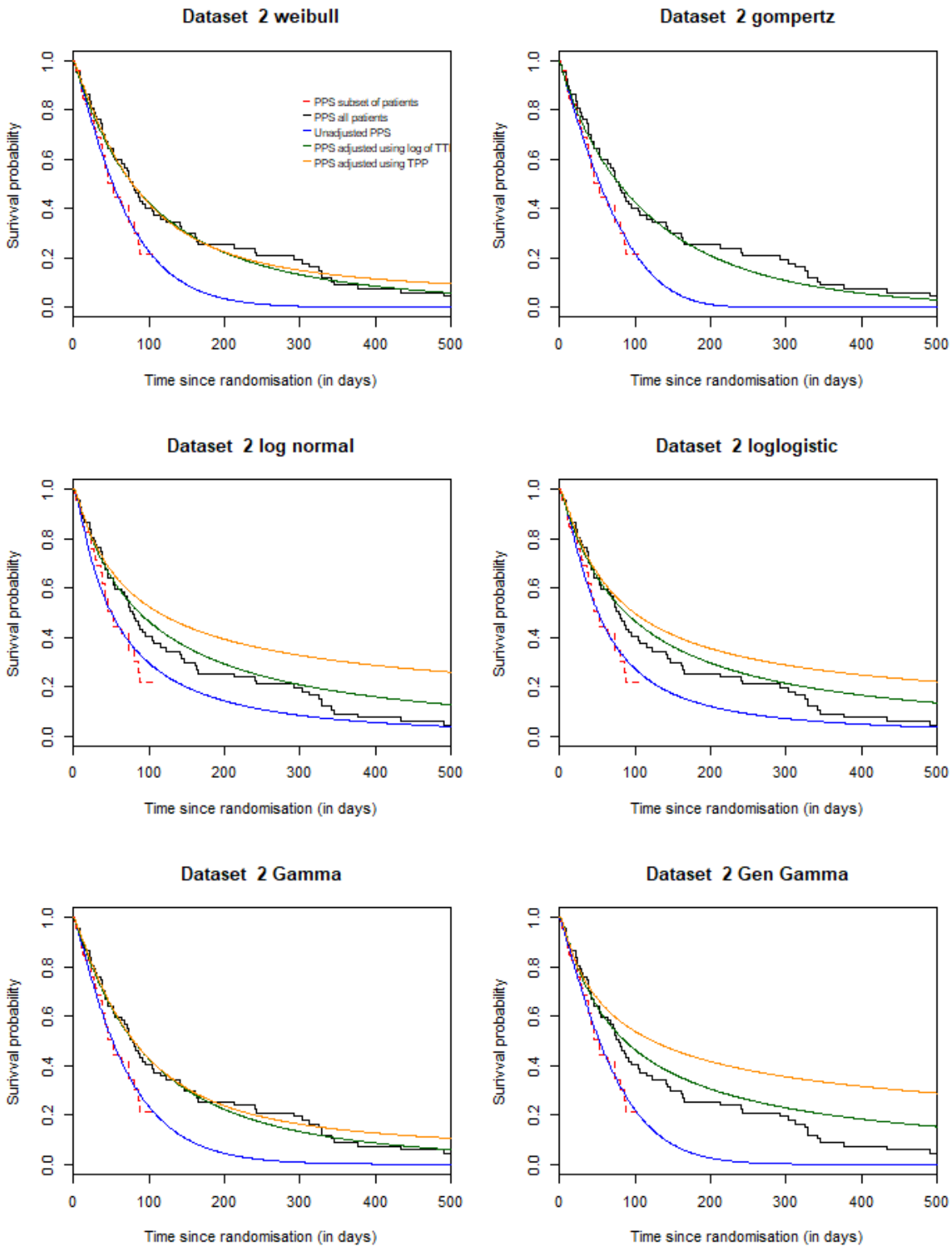


Figure 33 : Predictions for PPS using adjusted and unadjusted method in Dataset 3

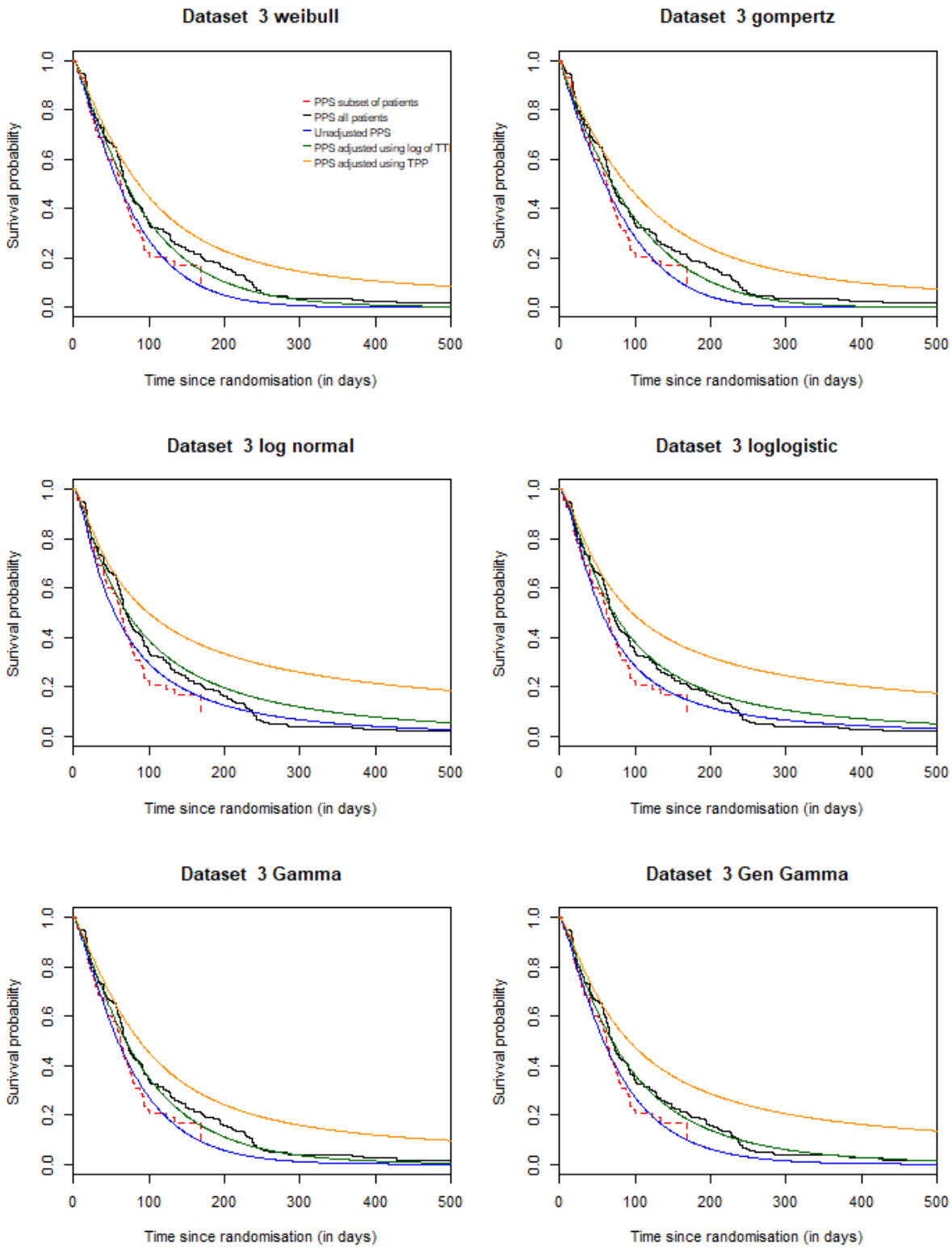


Figure 34 : Predictions for PPS using adjusted and unadjusted method in Dataset 4

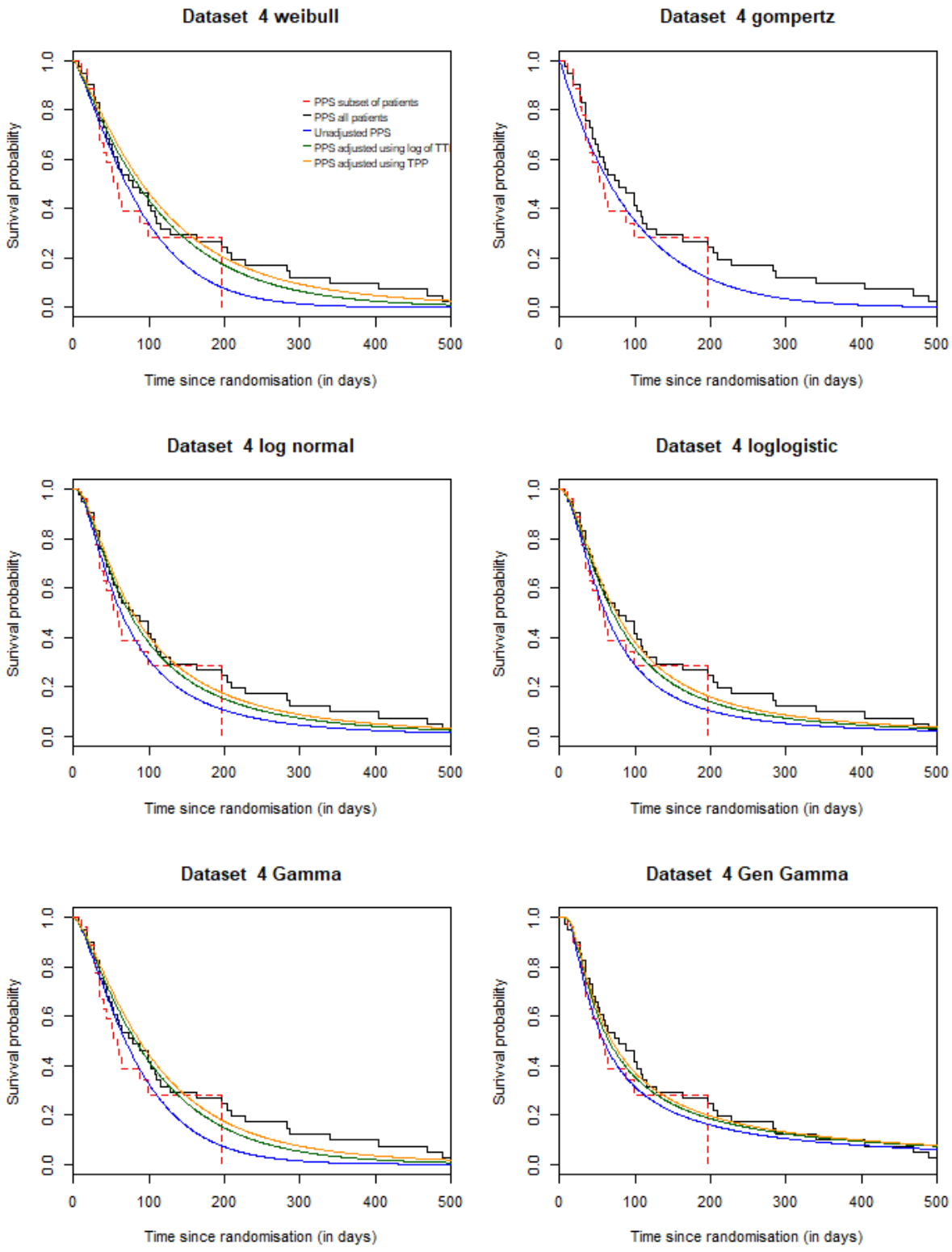


Figure 35 : Predictions for PPS using adjusted and unadjusted method in Dataset 5

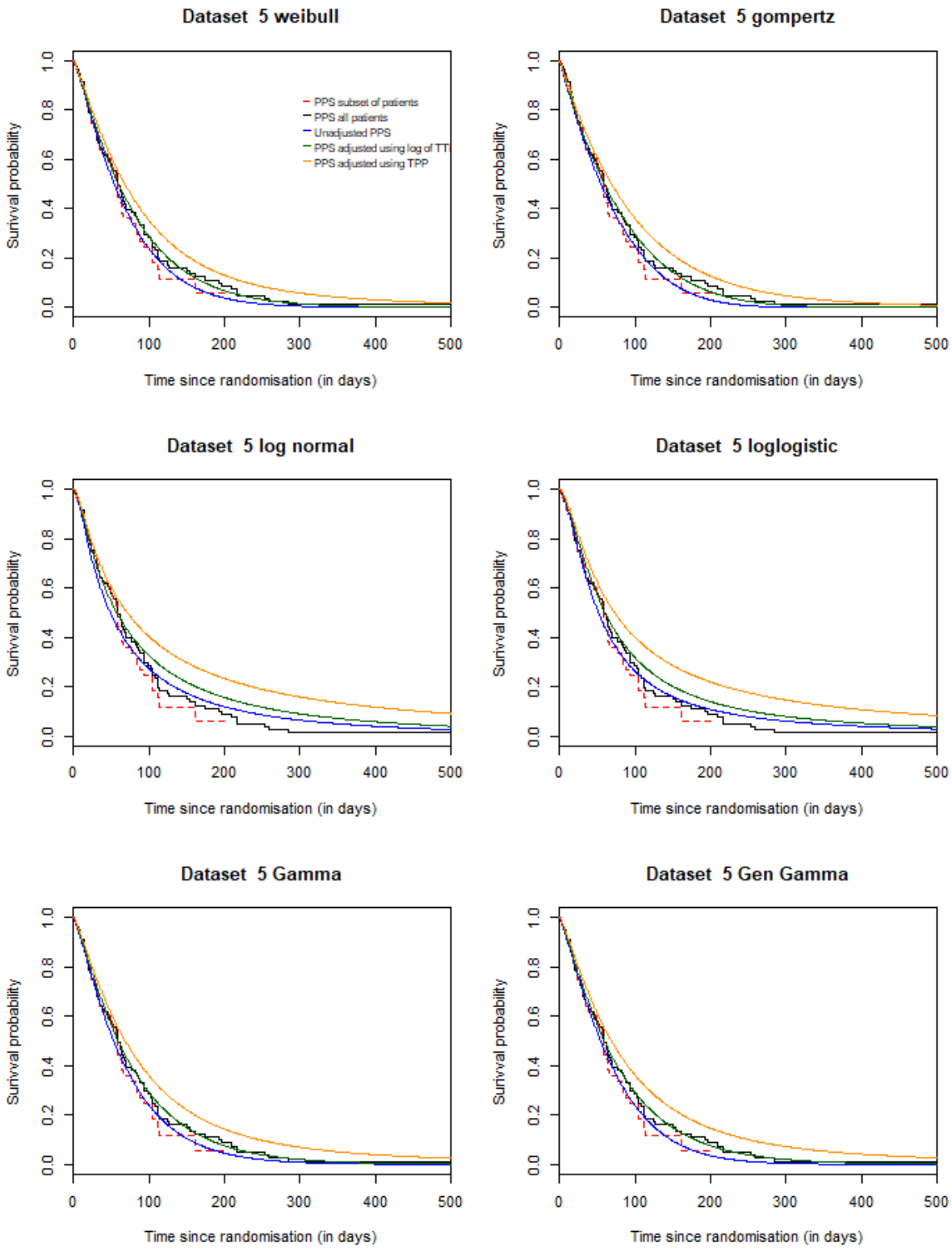


Figure 36 : Predictions for PPS using adjusted and unadjusted method in Dataset 6

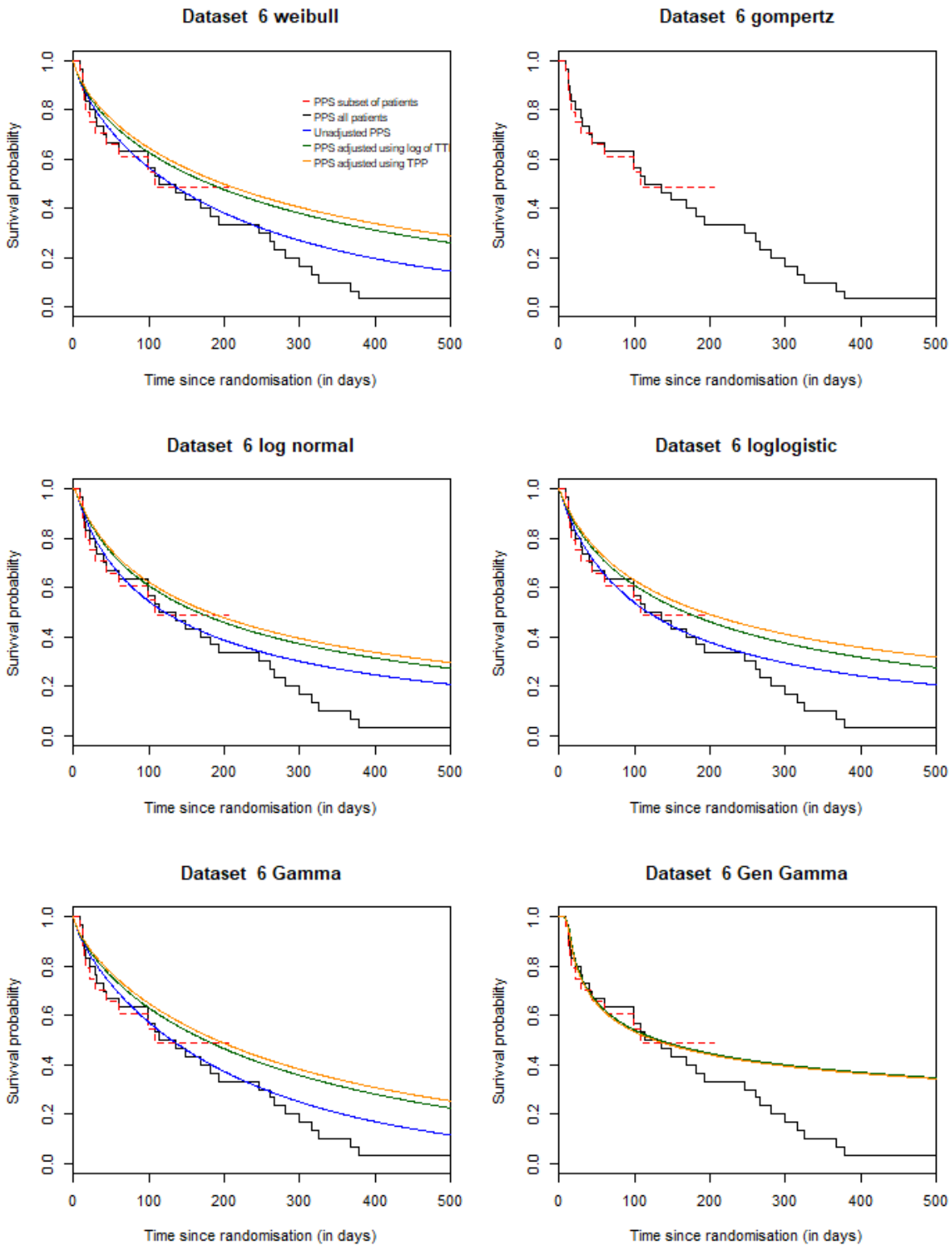


Figure 37 : Predictions for PPS using adjusted and unadjusted method in Dataset 7

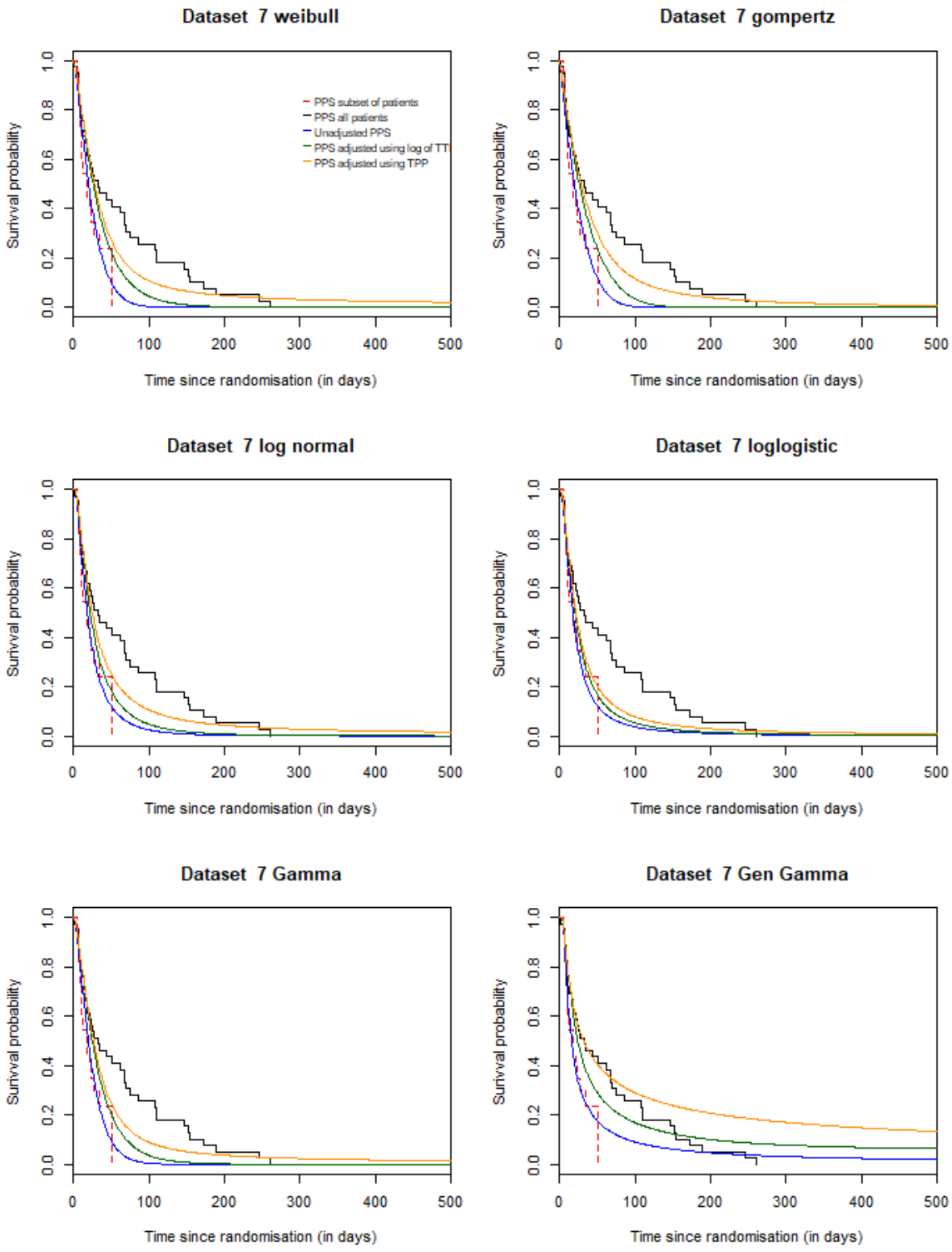
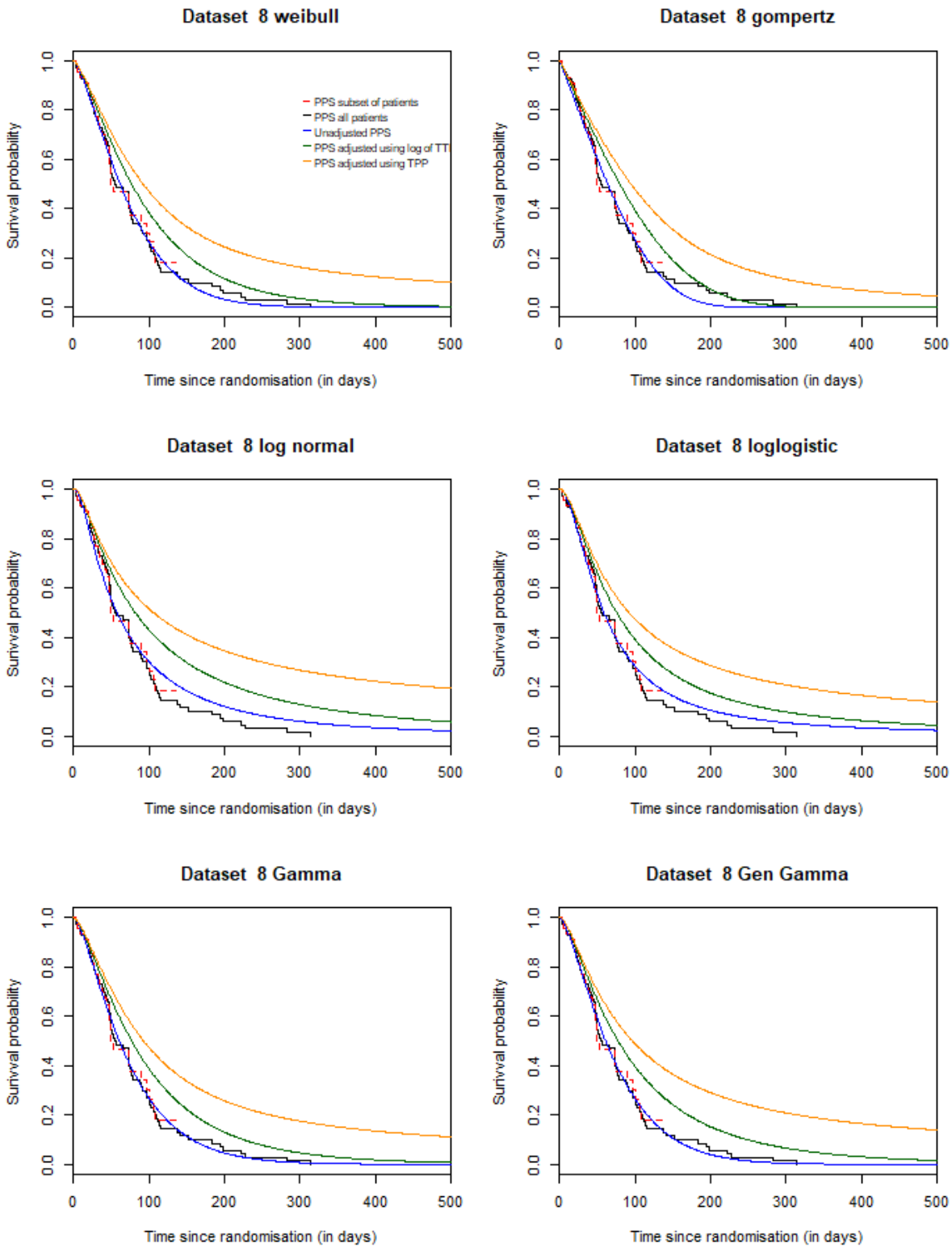


Figure 38 : Predictions for PPS using adjusted and unadjusted method in Dataset 8



As shown in the figures, the results differed between the trial arms.

In summary, under optimal conditions (i.e. reduced misspecification of PFS) I observed that:

- adjusting PPS using TTP as a covariate did not always correct for selection biases irrespective of the parametric distributions assumed (e.g. Datasets 1, 6, 7 and 8),
- adjusting PPS using TTP as a covariate on a non-log scale was generally less appropriate compared with using TTP on the log scale,
- whilst adjusting PPS using the log of TTP appeared to improve predictions in some trial arms (Datasets 2, 3, 4 and 5). This was only the case for selected parametric distributions. Furthermore, whilst predictions were improved, biases were only partly addressed.
- adjusting PPS using TTP may introduce biases when PPS estimated in the subset of progressors is close to the PPS in the complete dataset (Dataset 8).

Unexpectedly, results were consistent when different degree of censoring was assumed (results not shown).

7.6 Discussion and conclusions

This chapter highlights potential issues associated with the use of (unadjusted) PPS when developing a model (STM or MSM) based on incomplete data from an individual RCT.

Whilst the estimate for PPS could be considered unbiased for those patients who progressed, the estimate could be biased when representing the overall population (those who progressed and those who did not yet progress) for two reasons. First, patients who experienced a progression by the time the trial ended (also referred as early progressors) may have a different survival prognosis following progression compared with patients who had not yet progressed by end of the trial (from which data are not observed). Secondly, those who progress later are more likely to be censored in the PPS dataset. This would suggest that there is time-dependent bias in the form of informative censoring.

It is important to recognise these biases in order to understand how this could affect the performance of the STM/MSM whereby OS is estimated indirectly as a function of PPS and other transitions.

Whilst it is possible to adjust PPS by making it conditional on TTP (e.g. on the log or non-log scale), this approach remains flawed because the PPS in people who did not progress in the trial remains unknown and therefore the estimation relies on assumptions (notably that the trend in the data will continue into the unobserved period). The validity of such approach is likely to depend on the amount of data available and whether the trend observed in the data continue beyond the observed period of the study. Furthermore, whilst a constraint could be added in order to reduce the effect of the extrapolation of the trend in the data (e.g. assuming the same PPS after a certain point), there remain limitations as the cut-off is arbitrary.

Similar to using TTP as a covariate for PPS in the statistical model, it is possible to model PPS separately for the subgroups of early and late progressors. However, this approach suffers from similar limitations, including:

- the definition of “early” and “late” progressors is arbitrary.
- the amount of data on PPS available in an RCT relies on a small number of patients and events. Separating PPS into two groups (“early” and “late” progressors) may therefore further increase uncertainty.
- patients who did not progress in the trial are assumed to have the same prognosis as late progressors.

In this study, TTP/PFS was fitted to the complete datasets to avoid any misspecification. However, in practice, when constructing an economic model, TTP/PPS is fitted to the censored dataset. Any misspecification of TTP/PFS is likely to have a knock-on impact on the estimate of PPS.

Alternative approaches could also be considered to adjust PPS, using more complex statistical methods, such as Inverse Probability of Censoring Weighting (IPCW) whereby the PPS will be adjusted according to weights (no example was identified or known). Such an approach could help with the problem of informative censoring (those who progress later are more likely to be censored in the PPS dataset) if there is good information on prognostic characteristics measured over time. However, such an approach is also likely to be limited given that the PPS in people who did not progress remain unobserved; therefore, the PPS is weighted according to what is currently observed in the trial. These approaches are likely to be more valuable when supplemented with data from other trials or clinical opinion.

This study has a number of limitations that need to be recognised. Firstly, a limited number of datasets were examined and therefore conclusions may be different if different datasets were considered. However, using a limited example illustrated that the problem exists. Datasets were also arbitrarily cut when 70% of PFS events had been experienced. Results are likely to be different if alternative cut-offs were used. Indeed, if censoring was introduced later, progression would have been observed for a larger proportion of patients and the PPS for people who had progressed would have been more complete. Furthermore, the same degree of censoring was assumed for each trial arm. Usually, more censoring will be observed for the intervention arm. Biases could affect the control and intervention arm differently. The sample size in trial arms included was generally small, which could increase the uncertainty. It was not appropriate to pool the trial arms as different interventions were assessed. Results were also heavily influenced by the choice of parametric distribution. Finally, this study only looked at making PPS conditional on TTP on the log and non-log scale. Alternative parameterisations are possible. It is unclear whether this would improve or worsen the adjustment. However, whilst different parameterisations are possible, it is unclear how analysts could decide on the most appropriate parameterisation given that the shape of the PPS function for people who did not progress remains unknown.

Despite findings, this does not mean that PPS estimated in the subset of patients who progressed cannot and should not be used, as this is often the only information available to analyst and this is appropriate when the dependence between TTP and PPS is believed to be low. Rather, this exploratory analysis highlights that prior to using PPS naively, analysts should be aware of the limitations of the data and should consider whether the PPS observed in early progressors may be different to PPS in the overall

population. This could be informed by discussion with clinical experts or supplemented using data from registry or other data published in similar conditions; however, it should be noted that this could be different if a new drug has a different mechanism of action. Analysts should also examine whether any signs of dependence between TTP and PPS are exhibited in the data. This could help provide an understanding of whether the PPS estimated from the trial in the subset of patients is likely to suffer from selection biases. The sample size of the data and number of events is also important to consider as this is likely to make any adjustment more challenging and any prediction less reliable. This example also highlights that the impact of biases may differ between treatment arms; hence, what is observed in one [control] arm may not necessarily be transferable to the other [intervention] arm.

In summary, this exploratory analysis suggests that there is no single best approach for adjusting PPS. Analysts should therefore consider reporting both the unadjusted PPS and adjusted PPS (e.g. using log of TTP as a covariate) in order to reflect the uncertainty so that it can be considered in the decision-making process. Decision-makers are more likely to be more confident in their decision-making if predictions using the adjusted and unadjusted PPS are similar. However, as demonstrated in this chapter, adjusting PPS did not consistently improve predictions, therefore there is a risk that neither are correct. This is unknown, and decision-makers need to consider the possibility that all these estimates are uncertain and may be biased. If predictions using the adjusted and unadjusted PPS vary widely, this uncertainty should be reflected in the decision-making process. Whilst there is no clear solution, it may be useful for decision-makers to understand the importance of PPS on model results across a wide variety of alternative scenarios.

The potential impact of biases in PPS is further examined in the next chapter, where I describe the methods and results of a simulation study to examine the performance of methods, including the STM, MSM and PSM in estimating health state sojourn time.

PART IV: HOW DO METHODS PERFORM ACCORDING TO DIFFERENT DATA CHARACTERISTICS?

8 CHAPTER VIII: A SIMULATION STUDY TO EVALUATE THE PERFORMANCE OF METHODS TO ESTIMATE HEALTH STATE SOJOURN TIME IN SINGLE TRIAL ARMS

8.1 Chapter overview

This chapter presents the methods and results of a simulation study undertaken to: (i) evaluate and compare the performance of methods that are currently used in health economics to estimate health state sojourn time in single trial arms, and; (ii) assess whether including the dependence between progression and survival outcomes when constructing such models improves their performance. The performance of seven methods are examined in this simulation study; these are described in Section 8.3.4.

Section 8.3 describes the methods for the simulation study. This includes a description of the aim, details on how data for the simulation study were generated, the underlying assumptions of the methods examined, and the targets and performance measures considered within the study. Results are presented in Section 8.4, and a discussion of the limitations of the simulation study is presented in Section 8.5.

8.2 Introduction

As described in Chapter 2, a number of approaches are currently used in health economics to estimate health state sojourn times, and by extension, QALYs. However, these methods are used inconsistently between appraisals, largely because their relative performance is unknown. Furthermore, a key simplification on how these methods are currently implemented in health economics, which could affect their performance, is that PFS and OS are not jointly modelled.

In health economics, at present, whilst consideration is made on the long-term predictions, the performance of a method is typically judged upon its prediction against the observed data, e.g. does a PSM fit the observed data better than an STM? Whilst this can be evaluated easily during the observed period of a trial, follow-up within clinical trials is frequently limited and time-to-event data, particularly OS, are often immature. Furthermore, a lifetime horizon is often the standard in health economics. Whilst external longer-term evidence can be used in order to validate predictions from an approach, there are often limitations and these data are also often incomplete or unavailable.

A number of studies have examined how well models fitted to early data-cuts compared with later data-cuts.[27, 138] However, these were limited to single case studies and the later cut-off used to compare predictions was not fully complete. The overall performance of methods used in health economics is unknown because the presence of censoring (unobserved data) makes it impossible to know for certain what the data would have looked like had they been fully observed. As such, the performance of alternative methods for estimating OS and health state sojourn time is unknown.

Consequently, a simulation study is proposed in this thesis to attempt to evaluate the overall performance of alternative methods, accounting for their performance during both the observed and unobserved period in single trial arms. This is possible in a simulation study as hypothetical data can be generated that reflects the type of information that is typically available from a clinical trial, as well as the underlying truth (in the absence of censoring).

In this simulation study, contrasting approaches are examined in single arm trials; the PSM whereby OS is estimated directly from a parametric function fitted to the data with the hazard extrapolated over time and the STM which models the underlying disease process. A key debate in health economics is whether fitting OS directly to the data is more or less appropriate than modelling of the underlying process when using data from a RCT only (based on the information available), provided the most appropriate available extrapolation is used. The data is simulated and methods in this study are applied in a way that allows this question to be explored comprehensively.

8.3 Methods for the simulation study

A simulation study was chosen to evaluate the performance of methods currently used in health economic evaluation. This is because it allows for the evaluation of the overall performance of methods during both the observed and unobserved period, which would have otherwise not been possible using published trial data. In a simulation study, data can be generated in such a way that reflects both the restricted follow-up times and the small sample sizes typically encountered in clinical trials of oncology treatments, in addition to the unbiased times if the follow-up was not restricted and if a larger sample size was considered. Another key strength is the flexibility to assess a large number of defined scenarios, which would be challenging using published trial data.

Simulation studies are an accepted and widely used tool when evaluating the performance of statistical methods. Morris *et al* (2019)[139] define simulation studies as "*computer experiments that involve creating data by pseudo random sampling*" that help "*to understand the behaviour of statistical methods*". There are a large number of published simulation studies. However, they often differ from each other in terms of design, data generating mechanisms as well as the methods evaluated. This is because rather than following a strictly defined process, simulation is a general tool which needs to be adapted to its aim.

Morris *et al* (2019) published a recent review of 100 simulation studies that were published in Volume 34 of *Statistics in Medicine* (2015).[139] The authors highlight the variation between the included studies but also, the inconsistency in terms of terminology and study design. Findings from this comprehensive review were then used by the authors to formulate practical guidance, to develop a structured approach for the design, execution, analysis, reporting and presentation of simulation studies and to formulate coherent terminology.

In particular, Morris *et al* (2019) propose the planning of a simulation study to follow the ADEMP (Aims, Data-generating mechanisms, Estimands, Methods, Performance measures) structured approach.[139] The ADEMP approach proposed by the authors covers the key steps that need to be considered when designing a simulation study. This structured ADEMP approach is followed in this thesis for the planning of the simulation study.

The aim of the simulation study is described in Section 8.3.1. The data-generating mechanism is described in Section 8.3.2. Estimands are described in Section 8.3.2. Methods included in the simulation study are described in Section 8.3.3. Finally, the performance measures are described in Section8.3.5.

8.3.1 [A]DEMP: Defining the aims of the simulation study

Perhaps the most important aspect when designing a simulation study is a clear definition of its aim. This influences its scope, how data are generated and in turn, the level of complexity required, but also how reliable conclusions from the simulation study are and how results should be interpreted.

In this thesis, the primary aim of the simulation study is to evaluate the performance of methods that are currently used in health economics to estimate health state sojourn time (PSM, STM, MSM) in a single treatment group. A secondary aim is to evaluate whether including the dependence between progression and survival outcomes when constructing such models could improve their performance (Li's model[91] and Fu's model[89]). In particular, a key debate in health economics is whether fitting OS directly to the data is more or less appropriate than modelling of the underlying process when using data from a RCT only (based on the information available), provided the most appropriate available extrapolation is used.

In health economics, the accurate prediction of the incremental QALYs gained between treatment arms is important. However, a single trial arms approach was adopted in this simulation study to avoid the potential for spurious conclusions arising from apparently appropriate incremental outcomes despite the presence of a poor model fit in both treatment groups. Furthermore, when considering a control and an intervention arm, different choices are possible (e.g. assuming PH between arms for the whole model duration or until after a certain time point, pooling PPS data across arms, or fitting curves separately to data for each arm). These choices may lead to different model results which would introduce further complexities in interpreting the performance of alternative economic model approaches in the simulation study. An exploratory analysis is presented in Chapter 9, whereby the performance of methods in estimating incremental QALYs is assessed for transparency using real cancer trial datasets.

Whilst a simulation study is used to evaluate the performance of analytic methods, there are a number of challenges which need to be considered upfront when interpreting results. In particular,

(i) a large component of the performance of a method stems from subjective judgement on how transitions/survival endpoints are chosen and extrapolated, rather than the analytical approach itself. Consequently, results will need to be interpreted in the context of how parametric models were selected in the simulation study,

(ii) the key target is the estimation of the mean lifetime health state sojourn time. This makes assessing the performance of methods very challenging as a method may generate appropriate estimates of appropriate health state sojourn time despite poor model fit,

(iii) the hazard of death is likely to vary between cancer sites, interventions and lines of treatment. It is therefore challenging to draw generalizable conclusions.

8.3.2 A[D]EMP: The data-generating mechanism

A key objective of this simulation study is to generate data that are reflective of trials encountered in HTA used to estimate the health state sojourn time. Clinical trials typically include a number of variables, including time-to-event data (e.g. PFS and OS) and response outcomes (e.g. complete response [CR] and partial response [PR]) as well as other prognostic information (measured at baseline or over time). In theory, data could be generated to reflect all of this information using complex processes which account for the correlation between different prognostic factors and/or the inclusion of time-dependent covariates. It is equally important to consider what information from clinical trials is currently used when estimating health state sojourn time, in addition to information that could affect the performance of a method. As described in Chapter 2, current modelling methods (PSMs, STMs, and MSMs) use information on two key events; progression and death. The survival outcomes (PFS, OS, PPS) are generally, but not always, estimated without consideration of covariates (with the possible exception of treatment arm) or prognosis factors (e.g. survival outcomes are used on their own). However, there are instances, where survival outcomes may be estimated using baseline characteristics, notably when there are important stratification factors.

Generating data using complex methods by modelling a set of unnecessary covariates that are time-dependent (for instance, a link between change in white blood cell and progression) and linking them to the time to death or progression would rely on a number of unsupported assumptions that could reduce the realism of the data generated and reduce transparency. However, ignoring prognosis entirely (notably at baseline) may reduce realism in the data-generating mechanism.

Consequently, in this simulation study, a "semi-complex" process was used to generate data, whereby:

- The general process underlying the natural history of cancer is modelled to estimate PFS and OS, rather than assuming that the times to progression and death are independent,
- Time to death is made dependent on TTP (depending on the scenario of interest),
- A prognostic factor is considered at baseline to reflect patients with a good, moderate or bad prognosis in terms of speed of progression,
- Random, administrative and interval censoring are introduced to reflect the types of censoring typically observed in cancer trials.

A key aspect of the data-generating mechanism is the conceptual representation of the natural history of advanced/metastatic cancer, as summarised in Section 8.3.2.2. The subsequent sections detail the mathematical model used to simulate the data. Prior to describing the data-generating mechanism, it is important to define the scenarios that will be investigated. Given the aim of the simulation study (Section 8.3.1), the scenarios to be investigated need to be realistic to reflect characteristics of data typically encountered within the HTA context when estimating health state sojourn time.

8.3.2.1 Scenarios investigated

The scenarios investigated in this simulation study focus on four key characteristics of trial data that are likely to impact on the performance of methods under investigation. Characteristics for each scenario are assumed to follow three levels (high, medium, low) or two levels (high/low) for each category, resulting in 54 unique possible scenarios (summarised in Table 12 in Section 8.4). These include:

- The proportion of observed PFS events amongst all enrolled/randomised patients, which is defined as either progression or death (85%, 65%, 45%), informed by findings from a rapid review of previous NICE TAs in advanced/metastatic cancer (Appendix 7).
- The ratio of OS to PFS events (75%, 55%, 35%), informed by findings from a rapid review of previous NICE TAs in advanced/metastatic cancer (Appendix 7).
- The dependence/correlation between the TTP and PPS (negative moderate, independent and positive moderate), based on an assumption to reflect the possible form of dependence that could be seen in a trial. This dependence/correlation was induced using a Gaussian copula (bivariate normal) as the correlation coefficient equals the copula. Further details on how the dependence is included is provided in Section 8.3.2.4.2.1.
- The proportion of deaths that occur prior to a progression event (low or moderate/high), to reflect scenarios that have a lower or higher proportion of deaths pre-progression (low is assumed to be approximately two to three times lower than high). Further details are provided in Section 8.3.2.4.2.2.

As previously described, data relating to single trial arms only are generated. Scenarios are also limited to non-cure processes given: (a) the challenges in generating realistic data given the uncertainty in the underlying process for such treatments/conditions; (b) only non-cure survival models are considered when fitting data to survival outcomes and (c) to keep the number of scenarios manageable.

It should be noted that whilst 54 scenarios are examined in this simulation study, some scenarios might be considered more relevant or plausible than others. In particular, whilst negative dependence is

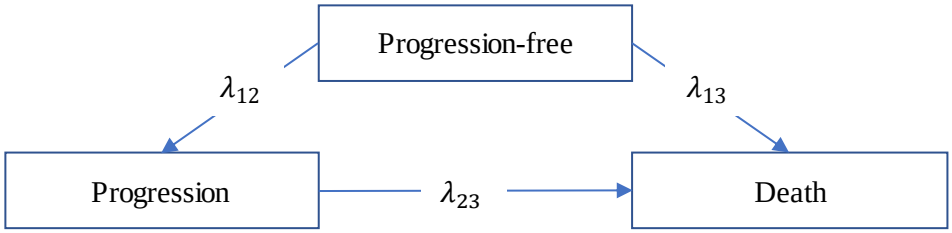
possible, the level assumed could be considered extreme. The level of positive dependence could also be considered extreme. However, it is useful to consider more extreme scenarios in order to examine trends, and also to assess whether the conclusions regarding the performance of methods are robust even under less plausible assumptions about the underlying distribution from which the data are drawn.

8.3.2.2 Conceptual representation of the natural history of cancer

As described in Section 8.3.1, a progressive multi-state process is used to represent the natural history of advanced/metastatic cancer (Figure 39) whereby patients could either (1) progress, (2) die without progression or (3) die following progression.

Figure 39: Conceptual representation of the natural history of cancer in the advanced/metastatic setting (transition between health states)

λ_{12} and λ_{13} represent the transition for the progression-free to progression and death respectively; λ_{23} represent the transition from progression to death



Within a trial, PFS is defined as the time to progression or death without progression and is therefore represented by two transitions; the transitions from progression-free to progression (λ_{12}) or death without progression (λ_{13}).

OS is a function all transitions within the multi-state with:

- Individuals dying following progression are represented as a function of two transitions: (a) the transition from progression-free to progression (λ_{12}) followed by (b) the transition from progression to death (λ_{23}),
- Individuals dying in the absence progression are represented by the transition from progression-free to death without progression (λ_{13}).

8.3.2.3 Generating censored data reflective of scenarios observed within the HTA context

A key challenge in this simulation study is to generate realistic data that reflect the different scenarios encountered within the HTA context, whereby data on PFS and OS are typically not fully observed for all patients. Previous simulation studies,[139] typically first generate the uncensored survival times (using the natural history model), with censoring introduced subsequently. Furthermore, performance is often assessed according to the level of censoring for OS alone, PFS alone, with scenarios rarely defined to account for the degree of censoring for both outcomes.

Scenarios to be investigated in this simulation study are defined according to the degree of censoring for both PFS and OS. Therefore, in addition to the natural history model, the censoring mechanism for both PFS and OS needs to be part of the data-generating mechanism.

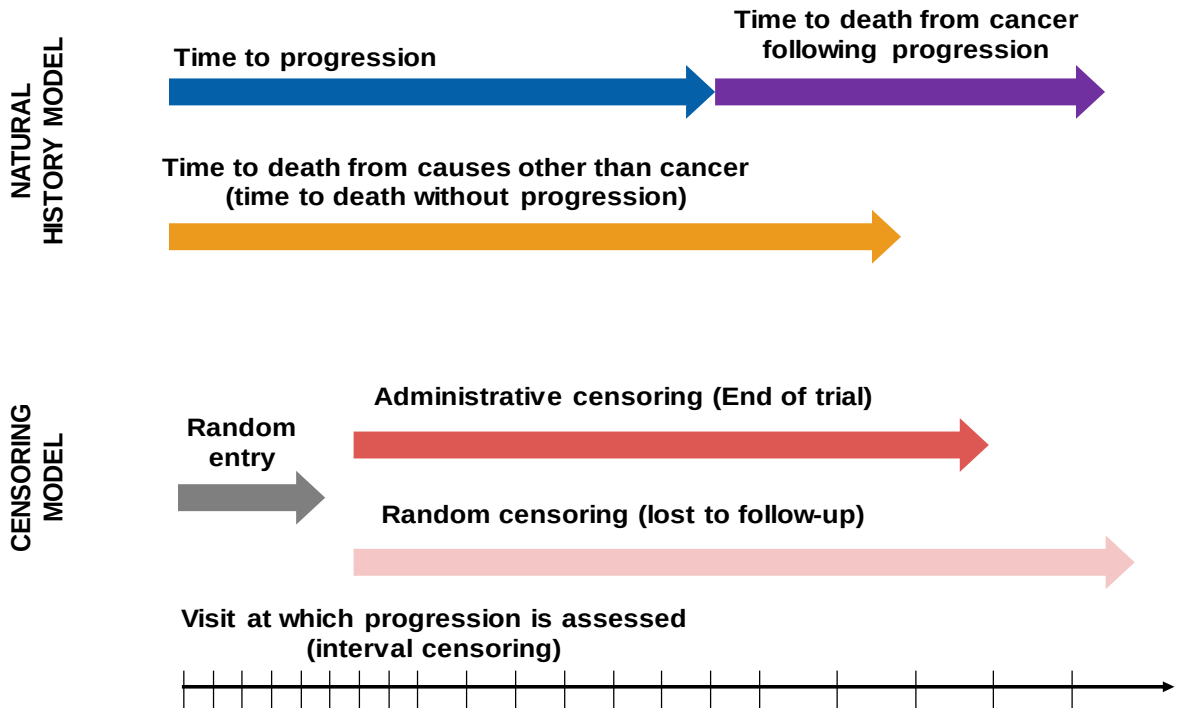
A multi-state, patient level-simulation was built in R software (R studio version 3.6.3). The model has two parts: (i) a natural history model of the progression of cancer which generates uncensored survival times, and (ii) a censoring model that generates censoring (see Figure 40). In this model it is assumed that the patient would die from causes other than cancer, and the time to death without progression is shorter than the time to death following progression.

The model simulates the life histories of a sample of patients with cancer. The simulation of the individual event histories uses Monte Carlo sampling. This means that each uncertain event within the individual's modelled lifetime can occur randomly, but overall the events conform to a pattern which is specified by the scenario identified. Briefly, uncensored survival times are first generated using the natural history component of the model. For each simulated individual (assumed to be free of progression at the start of the simulation), TTP, prePS and PPS are generated based on the relevant distribution (described in Section 8.3.2.4). The generated uncensored times are then compared with:

- PFS, which is defined as the minimum of the TTP and PrePS
- OS, which is defined as death before or after progression (PrePS or TTP+PPS)

In parallel to the process of generating uncensored survival times, a censoring mechanism is introduced (shown in the bottom half of Figure 40), which includes administrative censoring (the end of the trial period), random censoring (patients who are lost to follow-up) and interval censoring (progression assessed during scheduled visits). Uncensored survival times are then compared against the censoring mechanism to define whether censoring (either administrative or random) occurs before the event of interest. The model further considers random entry into the trial in order to add realism to the generated data.

Figure 40: Representation of the model used to generate data



In order to match the scenarios of interest (in terms of censoring), some parameters need to be calibrated. Consequently, predictions from the MSM in terms of the proportion of PFS and OS events, given a set of initial parameters, are compared against the target values (the proportion of PFS and OS events for the scenario of interest). Parameters from both the natural history and censoring models are then varied simultaneously using an iterative approach until the characteristics of the generated data match the scenario of interest in terms of the level of censoring for PFS and OS.

A number of optimisation functions such as the `optim()` function in R (including different algorithms such as Nelder Mead [NM] or Broyden–Fletcher–Goldfarb–Shanno [BFGS] methods) or other algorithm approaches exist such as the Metropolis Hastings calibration method.[140, 141] For the sake of simplicity, I used the iterative approach taken by Rotolo *et al* (2013) to calibrate the parameters from the MSM to match the target values.[129] In brief, the authors developed a criterion function which compares the target values versus predictions and varies parameters from the MSM such that the differences between the two is minimised according to a threshold set by the user. Parameters are varied according to an automated heuristic function with acceptance criteria until the logarithm of the ratio between the targets and predictions reaches a value close to zero. Further details on the algorithm and heuristic function are available in Rotolo *et al* (2013).[129] It should be noted that the choice of optimisation function would not affect findings, as this is only used to calibrate parameters for the natural history and censoring models that match (broadly) the scenario of interest.

8.3.2.4 Parameters included in the mathematical model used to generate data

The mathematical model used to generate the data for this simulation study for each scenario is composed of three categories of parameters; (i) those that are fixed and common to all scenarios; (ii) those that are fixed but specific to the scenario investigated, and (iii) those that need to be calibrated so that the generated data match the characteristics of the scenarios investigated. Parameters are summarised in Table 11. Further description is provided in turn below.

It should be noted that several key inputs in the natural history MSM are a function of a number of parameter values. For instance, in the natural history model, the time to death in the absence of progression (general population mortality) is a function of: (a) the age of the patient, and (b) the rate of mortality from the general population (mortality is higher as patient age increases). Similarly, TTP is a function of: (a) the patient's prognostic group, and (b) TTP by prognostic group. Finally, administrative censoring is a function of: (a) random entry in the model, and (b) the follow-up time for the last patient who enters the trial.

8.3.2.4.1 <u>Parameters that are fixed and common to all scenarios</u>

8.3.2.4.1.1 Age and time to random entry

Patient age is included in the natural history model when generating the data to estimate the time to death in the absence of progression (used as a proxy for death other than cancer). For simplicity, patient age is sampled from a triangular distribution assuming a median age of 65 years (range: 45-90 years) to reflect the age of patients typically observed in people with advanced/metastatic cancer enrolled in clinical trials. Other distributions such as a Gaussian distribution with upper and lower truncation could have been more realistic; however, a triangular distribution was used for simplicity.

Figure 41 : Distribution used for the time to random entry and distribution of age

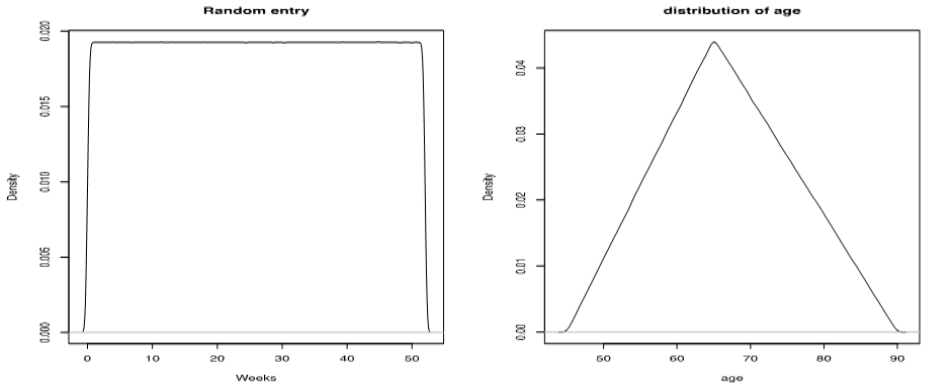


Table 11: Summary of parameters used in the mathematical model used to generate data

Parameter	Distribution	Parameter value	Source
Parameters that are fixed and common to all scenarios			
Age	Triangular	median: 65 (range 45-90)	Assumption
Random entry	Uniform	0 - 52 weeks	Assumption
Rate of non-cancer mortality by age	Fixed		ONS
Prognostic group	Fixed	High: 40% Moderate: 30% Low: 30%	Assumption
Time to progression by prognostic group	Spline with two knots models		Project Data Sphere
Random censoring rate	Exponential	0.0009	
Interval censoring	Fixed	Week 0-52: every 4 weeks Week 52-124: every 6 weeks Week 124-260: every 8 weeks Week 260+: every 12 weeks	Assumption
Parameters that are fixed and specific to each scenario			
Dependence (copula function)	Gaussian	Independence: Kendall's $\tau=0$ moderate positive: Kendall's $\tau=0.4$ moderate negative: Kendall's $\tau=-0.4$	Assumption
Excess mortality (SMR) – general causes	Fixed	SMR = 1.5 (low mortality) vs SMR = 7 (high mortality)	Assumption
Additional background mortality due to cancer	Exponential	0.0002 for scenarios with low mortality 0.00055 for scenarios with high mortality	Assumption
Parameters that are calibrated			
Time to death following progression	Exponential		Calibrated
administrative time for censoring	Fixed		Calibrated

Abbreviations: ONS: office of national statistic; SMR: standard mortality ratio

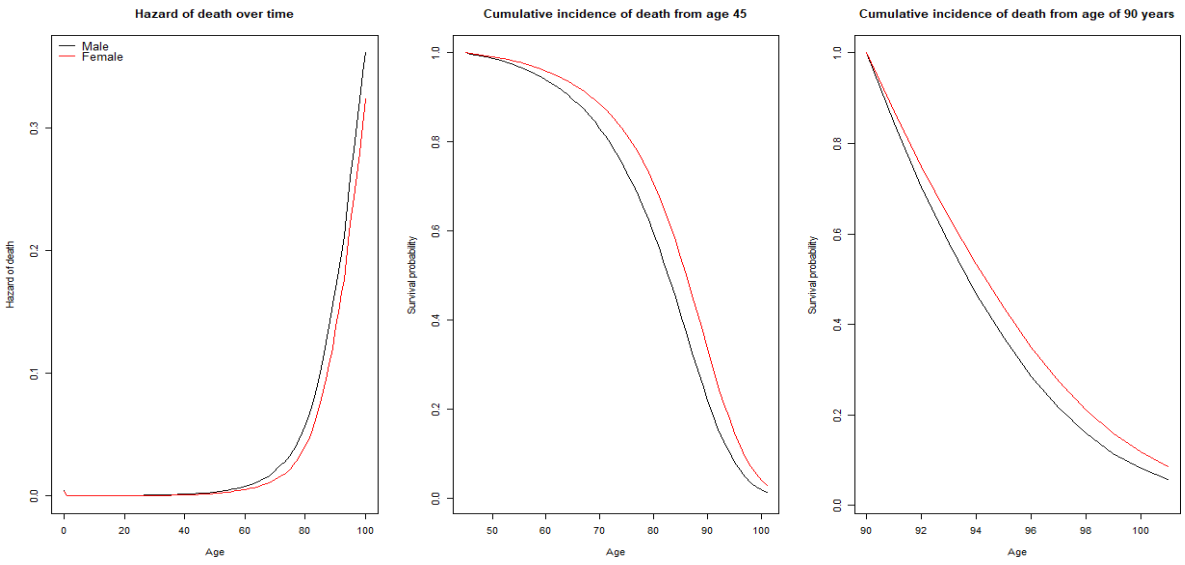
Random entry is included in the censoring model to reflect that patients do not all enter the trial at the same time, but within a time interval, with the trial ending when the last patient who entered the trial has been followed for a specified minimum duration or when a certain number of events have occurred (also called administrative censoring). In this simulation study, patients are assumed to randomly enter the trial within a 52-week period, assuming a uniform distribution. This was informed by the recruitment period observed in two trials for the first-line treatment for advanced BC, which was 60 weeks and 78 weeks, respectively.[142, 143]

8.3.2.4.1.2 General mortality – time to death in the absence of progression

The hazard of mortality by age in the general population is taken directly from life tables.[144] The hazard of death in males and females is summarised in Figure 42. For simplicity, to avoid including gender in the data-generating mechanism, the hazard for males and females was pooled assuming a 50:50 split. For each sampled individual, the hazard of death by age is transformed into probabilities and then transformed into the cumulative incidence of events from the age at which they enter the model; this was done to reflect the higher probability of dying from non-cancer-related causes in older individuals.

This is illustrated in Figure 42 (for males and females separately) where the cumulative incidence of death is presented for an individual entering the trial at the age of 45 or an individual entering the trial at the age of 90 years.

Figure 42 : Hazard of death by age and sex and example of cumulative incidence for a patient aged 45 years and a patient aged 90 years



8.3.2.4.1.3 Prognostic group

To add further realism to the data-generating mechanism and to avoid the use of a simple function for TTP, sampled individuals were classified into three prognostic group: good prognosis (40%), moderate prognosis (30%) and bad prognosis (30%). TTP is assumed to be dependent on prognostic group.

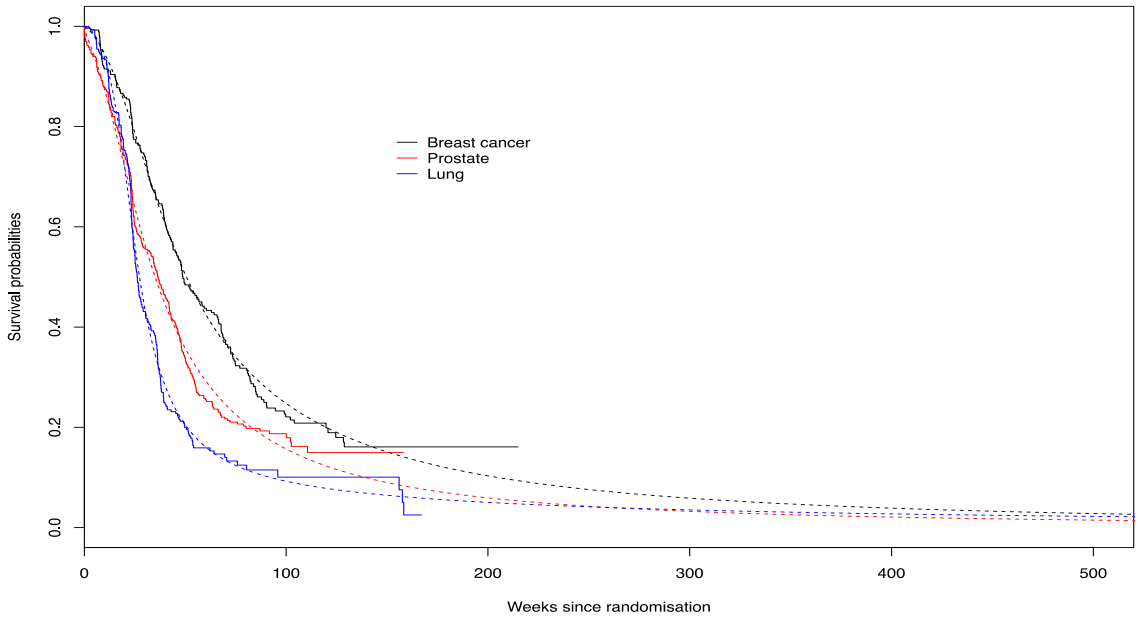
8.3.2.4.1.4 Time to progression by prognostic group

TTP by prognostic group was taken directly from the TTP observed in trials for some cancers.

Individual patient data from three trials in breast, prostate or lung cancer were obtained from the Project Data Sphere[74] with TTP calculated by censoring death in the absence of the progression. These datasets were described in Section 5.4.1. In these datasets, the median TTP was estimated to be 48.8 weeks, 36.6 weeks and 26.1 weeks in individuals with breast, prostate and lung cancer, respectively.

Data on TTP were not fully observed in either the breast, prostate or lung cancer datasets (Figure 43). Consequently, parametric extrapolation was required to extrapolate the TTP over a lifetime horizon. A range of parametric distributions was explored (exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma, and a spline hazard model with one to four knots). Following assessment of the visual fit and statistical fit (AIC), a spline model with two knots was selected for each dataset (Appendix 8).

Figure 43: KM for TTP observed in breast, prostate and lung cancer and fitted curve



The TTP in breast cancer is used as a proxy for TTP in individuals considered to have a good prognosis. The TTP in prostate cancer is used as proxy for TTP in people with a moderate prognosis. Finally, the TTP in lung cancer is used a proxy for TTP in people with a bad prognosis. It should be noted that using data related to cancer site as a proxy for prognosis is an assumption. It could have been possible to use an arbitrary HR to define patients with good, bad or moderate prognosis for example. Data related to cancer were used as proxy in order to assume realistic shapes of the TTP functions. This assumption is unlikely to affect results from this simulation study.

8.3.2.4.1.5 Rate of random censoring (for time to progression only)

Whilst the majority of censoring is likely to be attributable to administrative censoring, a small degree of random censoring was introduced to add realism to the data-generating mechanism, defined as the time to loss of follow-up. In the model, only the TTP could be randomly censored due to loss of follow-up as the time to death was assumed to be known with certainty.

The time to loss of follow was assumed to follow an exponential distribution with a rate of 0.0009 (weekly). This was selected (using trial error) so that only a small number of progression events (approximately 5-10%) would be censored for this reason.

8.3.2.4.1.6 Interval censoring (for time to progression only)

In addition to random censoring, interval censoring was also introduced to add further realism for the TTP given that progression is often only recorded during scheduled visits rather than at the time of the progression event. This was done by assuming that sampled individuals attend regular scheduled visits and that progression would only be recorded during the visit following the real time to progression. This is a simplification, as in reality, some patients may present earlier if they develop symptoms, hence leading to an earlier time of documented disease progression. I assumed that sampled patients had a visit every 4 weeks the first 52 weeks, every 6 weeks between Week 52 and Week 124, every 8 weeks between Week 124 and Week 260 and every 12 weeks thereafter. This is an unsupported assumption. Consequently, if progression occurs before a scheduled visit, progression is assumed to be recorded at the next scheduled visit, rather than at the time of progression.

8.3.2.4.2 Parameters that are fixed but specific to the scenario investigated

8.3.2.4.2.1 Dependence parameter between TTP and PPS

Scenarios were defined according to three levels of dependence between: (a) TTP, and (b) PPS (independence, negative moderate dependence, positive moderate dependence) using a copula (θ). Correlated random numbers were generated from the specified copula, which were then linked to the marginal function for TTP and PPS.

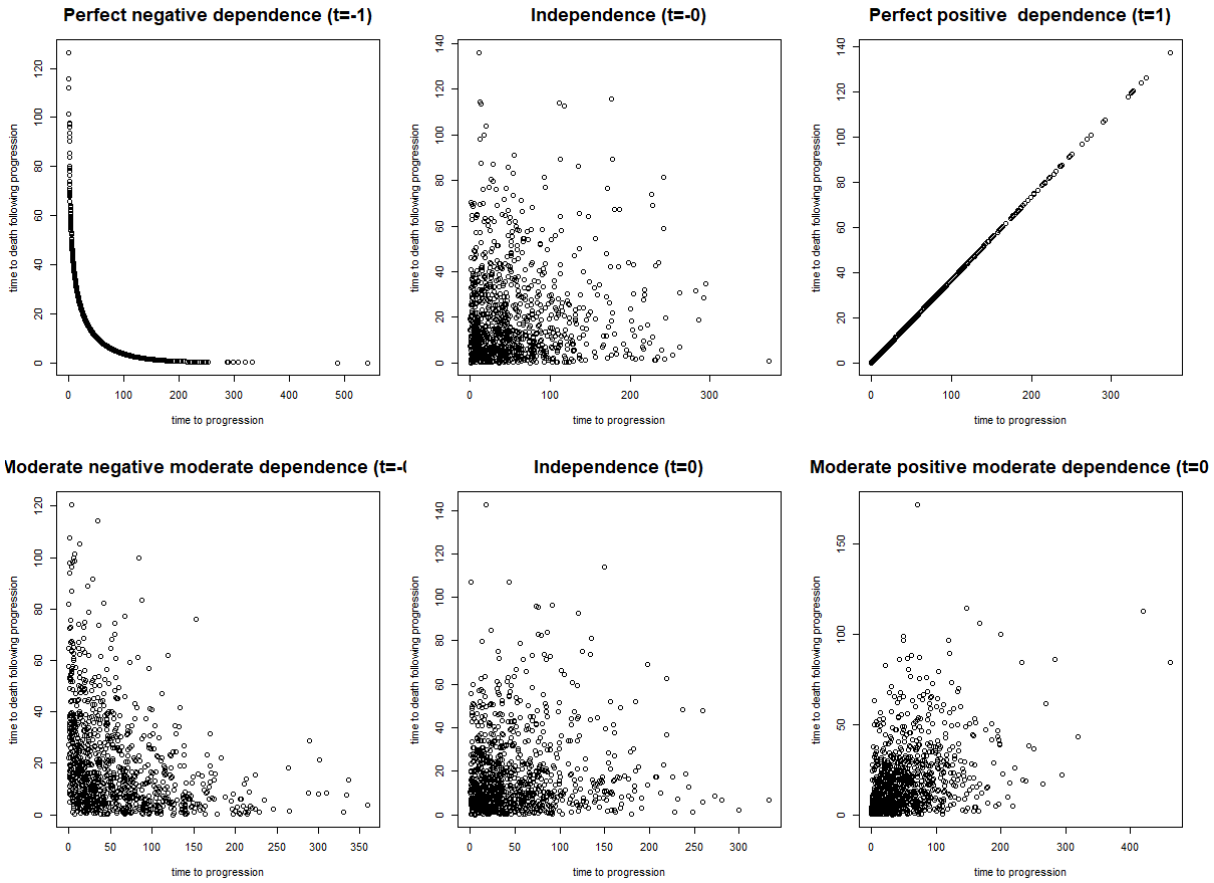
As described in Chapter 6.9.3, a number of copulas exist, with different tails and distributions. For the sake of simplicity, the dependence between TPP and PPS was induced using a Gaussian copula (bivariate normal). The Gaussian copula is attractive as it is easier to interpret and the level of dependence can be determined directly by the Kendall's Tau correlation coefficient (theta = Kendall's Tau). A moderate dependence was therefore defined in this study by a Kendall Tau of 0.4 [-0.4].[145] It should be noted that greater or lower degree of dependence could have been explored. Whilst a moderate dependence could be considered extreme, such level was chosen in order to be able to distinguish any possible trend in the performance of methods.

The value for the copula parameter is defined by the scenario of interest and therefore was fixed for each scenario. Scenarios are separated into three categories:

- Negative moderate dependence between TTP and PPS = Kendall's τ =-0.4
- Independence between TTP and PPS = Kendall's τ=0
- Positive moderate dependence between TTP and PPS = Kendall's τ=0.4

An illustration of how the dependence between TTP and PPS is provided in Figure 44, assuming independence (Kendall's τ=0) and both perfect dependence (Kendall's τ=-1 and τ=1) in the top row of Figure 44 and moderate dependence (Kendall τ =-0.4 and τ=0.4) in the bottom row of Figure 44. On the left-hand side of Figure 44, TTP and PPS are negatively correlated and therefore, a longer time to progression (shown on the x-axis) is associated with a shorter time to death following progression (shown on the y-axis). The middle figure in Figure 44 shows an example of independence (Kendall τ=0), whereby the PPS is independent of TTP. Finally, on the right-hand side of Figure 44, positive dependence is assumed, whereby a longer time to progression (x-axis) is associated with a longer time to death following progression (y-axis).

Figure 44: Illustration of dependence between the time to progression and time to death assuming perfect and moderate dependence



8.3.2.4.2.2 Mortality

Scenarios were classified according to the rate of death in the absence of progression. This was included applying a SMR to general mortality risk, supplemented by an additional background mortality to reflect scenarios that have a lower or higher proportion of deaths pre-progression (low is assumed to be approximately two to three times lower than high).

- SMR for general mortality

The SMR against general population mortality was partly informed by Zaorsky *et al* (2017).[146] This study reports risks of non-cancer deaths amongst cancer patients against the risk of non-cancer death in the general population using US death certificate data in Surveillance, Epidemiology, and End Results (SEER). The study reported large variations in the risk of death from non-cancer causes according to the type of cancer, patient age, calendar year and time after diagnosis. The authors reported the SMR by age for all cancers, but also specific for each cancer. For instance, considering all cancers, the authors

reported a SMR ranging between 0.66 to 20.21 for different causes of death in patients aged between 60-65. The SMR in patients aged 85+ according to the cause of death ranged from 1.45 to 2.78.

For the scenarios where the PrePS was assumed to be low, the rate of mortality from non-cancer causes from the general population was used and slightly uplifted assuming an SMR of 1.5. For the scenarios where the rate of death in the absence of progression is assumed to be high, I assumed an SMR of 7, meaning that that the rate of death from non-cancer causes in individuals with cancer is 7 times higher compared with the rate observed in the general population.

- Additional background mortality

In addition to the mortality associated from general causes (applied to general population mortality), additional background mortality was included for prePS to ensure enough deaths occurred by the time data were censored and to ensure that the difference in terms of proportion of PFS events that are death between the scenarios within the low level and moderate to high level was between two-fold to three-fold. The additional background mortality was assumed to follow an exponential distribution with a rate of 0.0002 for the scenarios assuming a low mortality rate in PFS. An exponential distribution with a rate of 0.00055 was used for the scenarios assuming higher mortality rate in PFS. These values were selected following trial-error experimentation to ensure that scenarios that have a higher proportion of deaths pre-progression generate approximately two to three times higher number of death compared with scenarios that have lower proportion of death.

8.3.2.4.3 Parameters that are calibrated so that the data generated match the scenario investigated

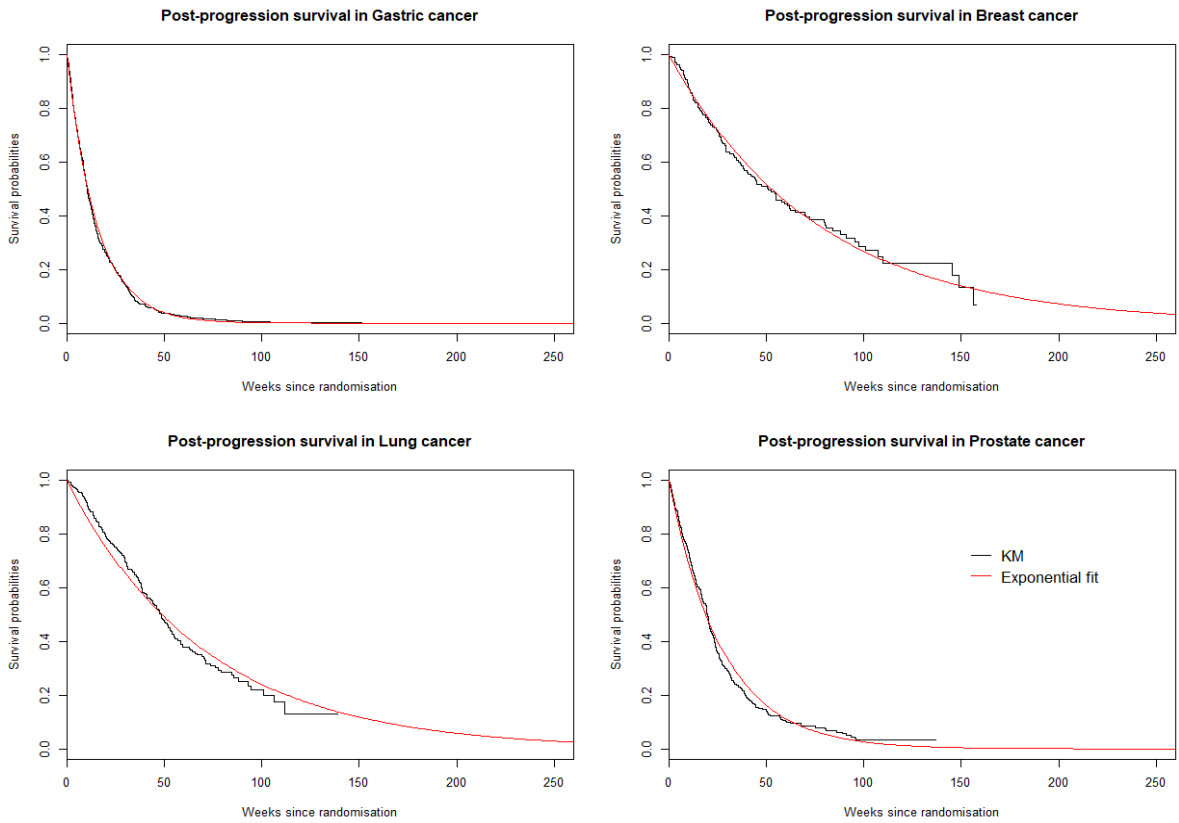
Only two parameters (PPS and time to administrative censoring) were calibrated using the optimisation mechanism reported by Rotolo *et al* (2013).[129]

8.3.2.4.3.1 Time to death following progression (PPS)

PPS was assumed to follow an exponential distribution. The rate parameter for the exponential distribution for PPS was calibrated to match the proportion of OS events for the scenario of interest. PPS was assumed to follow an exponential distribution as this is a single distribution parameter. If a two-parameter distribution was selected, such as a Weibull distribution, additional assumptions (such as the shape of the function) would be required given the large number of possible solutions. The selection of the exponential distribution was also supported by the shape of the PPS function observed in several cancers such as the BC, prostate, and lung cancers datasets (taken from the DataSphere project[30, 32, 33]) used in Chapter 5.4.1 and Gastric cancer[135, 136, 147] (Figure 45).

It should be noted that with the exception of the gastric cancer dataset (whereby most patients progressed), data in the breast, prostate and lung cancer datasets were less complete. Nevertheless, this supports the choice for the exponential as a reasonably realistic distribution for PPS for this simulation study.

Figure 45: Exponential models fitted to PPS data in the gastric, breast, prostate and lung cancer datasets



8.3.2.4.3.2 Duration of follow-up for the last patient who is entering the trial (administrative censoring)

Administrative censoring is included in the model to account for the fact that trials terminate early. This is included by stopping the trial after the last patient who entered the trial has reached a specified follow-up duration (or after a pre-specified number of events occur).

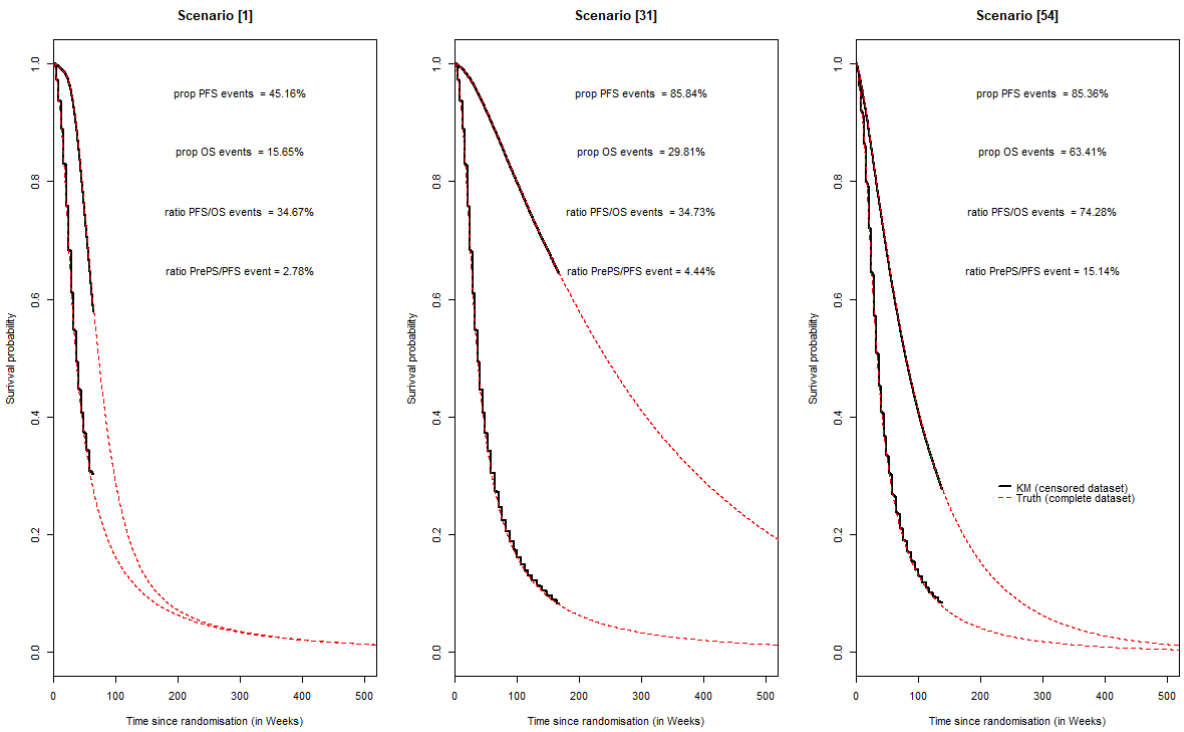
This parameter is calibrated so that the model predicts both the proportion of PFS and OS events to match the scenario of interest, given other parameters in the model.

8.3.2.5 Calibration of parameters

Parameters were calibrated using the algorithm reported by Rotolo *et al* (2013)[129] assuming 10,000 patients enter the simulation. The log of target values is compared against the log of the predictions. Parameter values are varied until the ratio of the log is minimised until reaching a specified acceptance criterion (set to 0.0001). Therefore, for each scenario, we have a set of fixed/calibrated parameters that are specific to the scenario of interest and therefore one underlying truth.

The calibrated parameters are shown in Appendix 9. The resulting PFS and OS for three scenarios (1, 31, 54) are shown in Figure 46 assuming a large number of patients (assumed to be 240,000 individuals – this is justified in Section 8.3.3) in order to represent the truth. The resulting PFS and OS KM for all 54 scenarios are presented in Appendix 10. It can also be seen than some scenarios could be considered to be more plausible than others, but overall, the resulting PFS and OS reflect a range of what could be observed within a trial. As a consequence of sampling variation and the use of only 10,000 patients when calibrating the parameters (larger number of patients would increase the calibration time), the value predicted for the truth when assuming 240,000 patients may not exactly match the target value; however, this should remain very close (e.g. for Scenario 31, the target for the proportion which were progression or death [PFS] events was 85% compared with 85.84% assuming 240,000 patients).

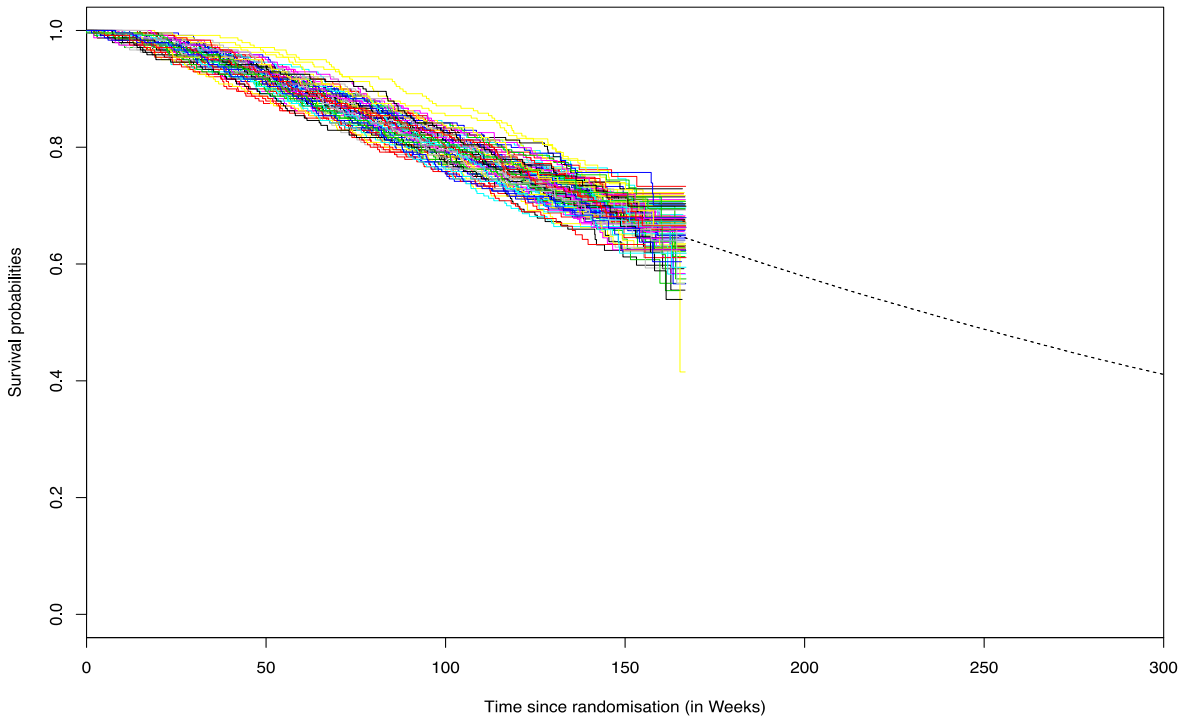
Figure 46: Examples of generated PFS and OS for selected scenarios (Scenario 1, 31 and 54)



8.3.2.6 Generating the simulated datasets to test the performance of the methods

The parameters previously calibrated/defined (in Section 8.3.2.4) represent the true parameters for the underlying natural history. However, trials typically include a small number of patients; hence, if a trial was to be replicated, there would inevitably be some variation. For each scenario, using the same set of parameters, 1,000 unique datasets are generated, assuming a sample size of 240 patients based on a typical median number of patients recruited in trials seen in HTA (Appendix 7) to account for sampling variation. This is illustrated in Figure 47 for Scenario 31, whereby the uncensored true survival time (generated assuming 240,000 patients) is plotted against the censored survival times observed in the first 100 datasets.

Figure 47: Example of datasets generated for Scenario 31



8.3.3 AD[E]MP: Selection of the estimand/target

As highlighted by Morris *et al* (2019), simulation studies typically compare methods for estimating a parameter of the data-generating model, also termed an estimand. However, the authors highlight that the target can be other quantities, such as a fitted outcome mean.

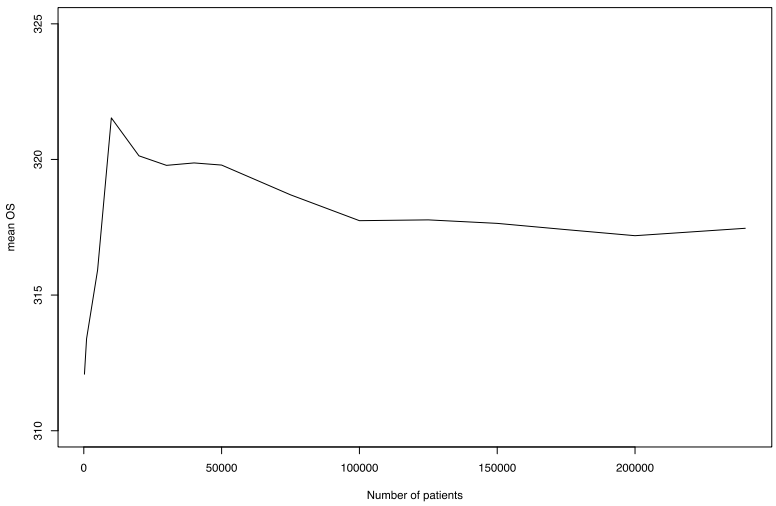
Given the aim of the simulation study - to evaluate methods to estimate health state sojourn time and QALYs - mean lifetime OS was selected as the key target (denoted θ). Total QALY gain was chosen as a secondary target in sensitivity analysis as its interpretation can be more challenging given that QALYs in PFS and PD contribute to the overall QALYs. Indeed, an approach may be considered to have low bias when looking at the overall QALYs, but the contribution of QALYs in PFS and PD may be biased. Mean PFS was also evaluated as a sensitivity analysis.

Using the mean OS as a key target is further justified by the expectation of similar PFS predictions between approaches. PFS is used directly within the PSM and simplified STM. In the MSM, PFS is estimated by combining TTP and PrePS under a competing risk framework. However, this is unlikely to lead to a significant variation in PFS given the small number of deaths typically observed prior to disease progression.

It should be noted that whilst the restricted mean can be considered more convenient and is less influenced by the choice of extrapolation curve, in health economics, methods are used to calculate the LYGs and QALYs over a lifetime rather than for some finite time interval.

For a given scenario, the target (θ) is calculated by generating data from the multi-state natural history model (without censoring) using parameters estimated in Section 8.3.2.4 using 240,000 patients to provide a robust approximation of the population mean OS, QALYs and PFS. This is illustrated in Figure 48 where the true population mean OS was calculated for Scenario 31 assuming different numbers of patients. It can be seen that the mean estimate OS is stable (little variation) after generating approximately 100,000 patients.

Figure 48 : Predicted mean OS using different population size



In health economic models, analysts calculate the mean OS, QALYs, PFS using both (a) the point estimate parameter value (usually referred to as the deterministic estimate) and (b) using the distribution around the point estimate (referred to as the probabilistic estimate) to both account for the uncertainty around the point estimate but also to account for possible non-linearity (to get the expectation of the mean).

Consequently, two analyses are presented when estimating the mean OS, QALYs and PFS, calculated using:

1. The point estimate parameters for each distributions
2. A set of input parameter values drawn by random sampling for each distribution repeated a 1,000 times with the mean OS, QALYs and PFS estimated as the mean of the 1,000 iterations. Parameters for the exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma distributions were drawn by random sampling using the `normboot.flexsurvreg` function in R which simulates alternative parameters from the asymptotic normal distribution of parameter estimates under sampling uncertainty (multivariate normal distribution).

8.3.4 ADE[M]P: Modelling methods included in the simulation study

The performance of seven methods are examined in this simulation study;

1 The PSM model as implemented in health economics, whereby PFS and OS are extrapolated independently of one another,
2 The Simplified STM as implemented in health economics:
    - Assuming a constant probability of death in PFS, with the post-progression survival time unadjusted
    - Assuming a constant probability of death in PFS, with the post-progression survival time estimated as a function of the time to progression on a log scale (TTP in log scale used as a covariate in the regression model for PPS),
    - Assuming a constant probability of death in PFS with the post-progression survival time estimated as a function of the time to progression on the normal (non-log) scale (TTP in normal scale [non-log] used as a covariate in the regression model for PPS)
3. The MSM (fitted using the `mssample` function),[43, 44, 50]
4. The model developed by Li *et al* (2015)[148] described in Section 6.9.1.1
5. Modelling the dependence between PFS and OS under a semi-competing risk framework (using a copula) – the model developed by Fu *et al* (2013),[89]

The methods chosen for evaluation within the simulation study include those methods which are commonly used in health economic models to estimate health state sojourn time, as well as additional methods identified from the review of methods which could be easily adopted in health economics. Further details regarding the implementation of some these approaches included in this simulation study are given in Chapter 2, 4, 5, 6 and 7.

The performance of the MSM using the `msm` function (Chapter 4.3) was not evaluated in this chapter to avoid misleading conclusions. Whilst a key assumption of the MSM implemented using the `msm` function is that the model is a Markov process, this assumption could be relaxed by assuming that transition rates follow piecewise exponential distributions (i.e. the hazards varies across a defined number of intervals). The performance of the MSM using the `msm` function is therefore likely to be highly dependent on assumptions made about the number of time intervals selected. Whilst it is possible to assume an arbitrary cut-off, conclusions are likely to be misleading, as the performance of the method could be different if alternative time intervals are selected. Consequently, I considered that it would be challenging to interpret any results from the MSM using the `msm` function given the infinite number of

possible implementations/combinations (number of intervals). Furthermore, when fitting the MSM and predicting health state occupancy over the patient's lifetime, the long-term extrapolation beyond the observed period of the trial will be based on the constant hazard observed in the last time interval considered – this assumption may not be considered appropriate.

Similarly, as described in Section 6.10 , alternative methods identified during the review of the methods to jointly model PFS and OS (such as the Meller model,[93] the use of frailty[87] and the first passage method[98]) were not taken forward for the simulation study in the absence of a clear tutorial on how these should be implemented. Given the need for transparency and technical skills of analysts typically in charge of building or reviewing cost-effectiveness models, a method is unlikely to be adopted if there are no examples/tutorials on how to implement the method in a suitable software package. Even with a tutorial (and examples available in a suitable package),[45, 50] adoption is not guaranteed as illustrated by the slow adoption of the MSM. Including only the most relevant approaches either currently used in HE, or well described was therefore a pragmatic reason.

The code use for this simulation study is provided in Appendix 11.

A weekly cycle length was assumed for all methods under consideration. Methods were implemented in R software. Further details regarding the implementation of methods, in particular how the MSM, the Li and Fu model are implemented are provided in Appendix 12.

The approach for survival extrapolation and selection of the parametric survival distributions is described in Section 8.3.4.1. This is because survival extrapolation and the model selection process plays an important role in the implementation of the methods investigated and their performance. Indeed, the prediction associated with each method is conditional on the parametric distribution assumed for each endpoint. Finally, Section 8.3.4.1 describes how general population mortality was included for each approach.

8.3.4.1 Selection of the preferred survival distribution/prediction

A large component of the performance of a method stems from subjective judgement on how transitions/survival endpoints are chosen and extrapolated, rather than the analytical approach itself. For instance, within the PSM approach, OS is estimated from the direct fit to the OS data. Different distributions will provide different predictions for the mean OS. Similarly, for the STM, OS predictions are a function of the time in PFS and the time in PPS. The choice of parametric distribution for PFS and PPS will therefore lead to different estimates of OS. Consequently, irrespective of the approach

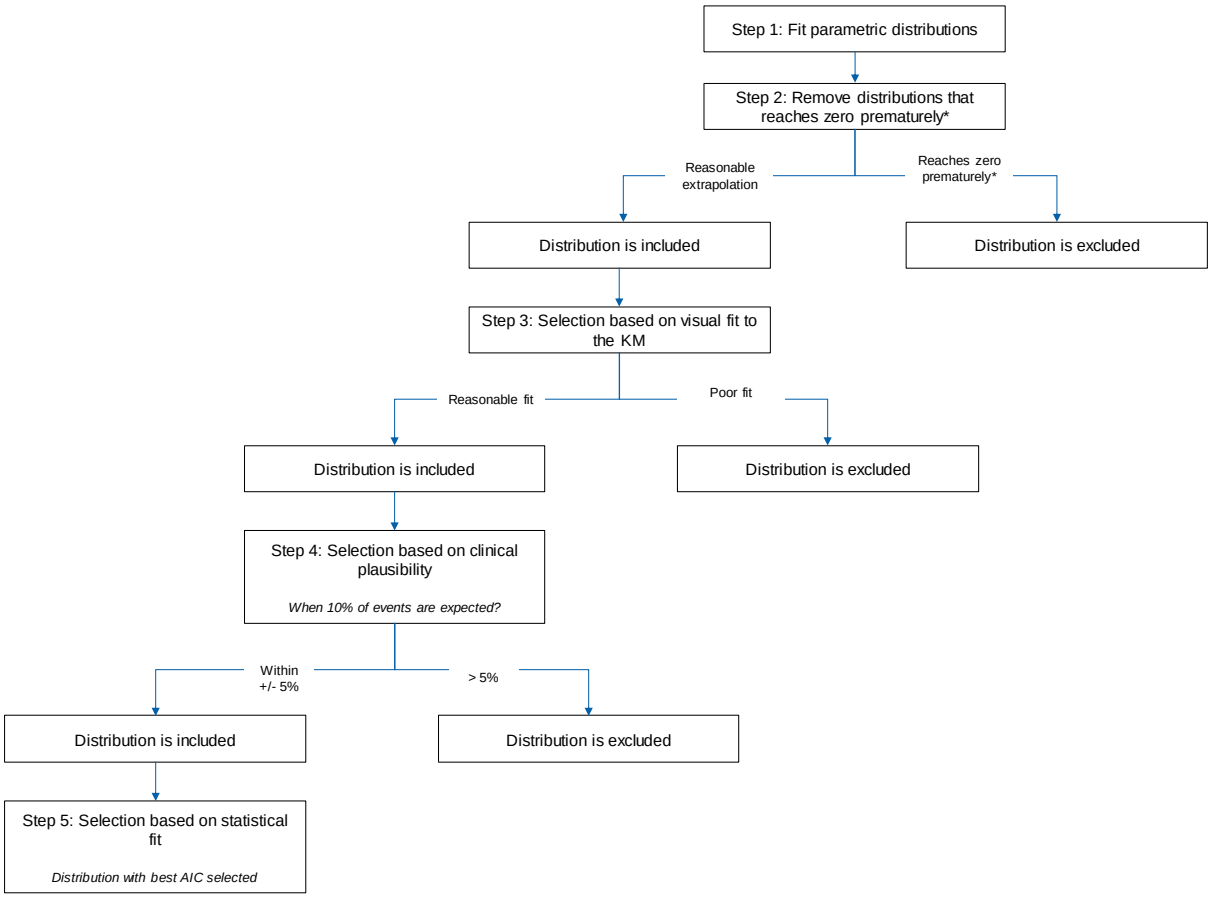
considered, analysts must make a decision about the most plausible parametric distributions. When looking at the performance of methods, it is therefore important that the selection process reflects the process typically followed by HTA analysts, rather than a best- or worst-case scenario.

Selecting the most appropriate parametric model for extrapolation is challenging due to incomplete information (censoring). As such, subjective judgement is required. NICE TSD 14[12] recommends that the most appropriate distribution is selected based on consideration of: (a) the visual fit of the predicted models to the observed time-to-event data; (b) the statistical goodness-of-fit of the model relative to all other fitted models (measured using the AIC and the Bayesian Information Criterion [BIC]); (c) an assessment of the observed hazards, and (d) the plausibility of the long-term extrapolation. However, despite these recommendations, analysts sometimes consider only statistical fit without giving sufficient attention to plausibility.

The base case for this simulation study uses a similar process to select the preferred survival distribution/prediction based on the TSD14[12] recommendations. The preferred distribution is selected based on the visual fit, the clinical plausibility and its statistical fit. This is possible because in addition to the censored data, data in the absence of censoring are also generated for each dataset. Assessment of the nature of the hazard function is not considered as it was not possible to automatize criteria related to this.

Whilst an attempt was made in this simulation study to use a similar process to select the preferred survival distribution based on the TSD14[12] recommendations, assessment of the long-term plausibility remains a subjective judgement and there is no one size fit all. A set of rules was defined to inform the selection of the most appropriate survival distribution. It should be noted that predictions for the PSM are from the direct fit to PFS and OS, whilst prediction for PFS and OS for the STM/MSM are a function of different transitions. Consequently, the term transition/prediction is used here to reflect the different inputs used to generate PFS and OS. The overall process used to select the most appropriate survival distribution in this simulation study based on the statistical fit, visual fit and clinical plausibility is depicted in Figure 49 with further description provided in the text below.

Figure 49: Process used to select the most appropriate survival distribution



* reaches zero when 90% of events are expected

- Fit parametric distributions to the transition/survival endpoint of interest (Step 1):

As described in Chapter 3.8, a number of parametric distributions exist (standard, spline, polynomials or mixture-cure models). Seven standard parametric distributions are considered in this thesis for the possible survival endpoints depending on the approach examined: exponential, Weibull, Gompertz, Log-Normal, Log-Logistic, Gamma and Generalised Gamma models. These distributions were chosen as they are commonly used to extrapolate time-to-event data in health economic models. These distributions were fitted to each survival outcome. Whilst not investigating cure models in this simulation study is a possible limitation, the scenarios examined do not include the possibility of a cure, reducing the relevance of cure models. Spline models were also not considered, as this would have increased model run-time, and many options are available in terms of the number of knots and the scale on which the model is fitted.

- Remove distribution/prediction that reaches zero prematurely when 10% of events are still expected to be observed (Step 2):

Cumulative survival probabilities for a number of distributions could reach zero prematurely and therefore these were not considered to be appropriate and were removed outright. This was defined in this simulation study by the time at which 10% of patients had not yet experienced the event of interest.

- Assessment of visual fit (Step 3):

The predictions for each model are compared (Step 2) against the KM plots for OS and PFS. Only those distributions for the transitions/prediction of interest that provided a reasonable fit to the KM were retained. Distributions that did not provide a reasonable fit for OS and PFS were excluded. A reasonable fit was defined as models that provided in most cases (85% of the time) predictions within the 95% confidence interval for the KM for OS and PFS. The threshold was increased gradually (by 10%) if none of the predictions were deemed to provide a good visual fit.

- Assessment of the long-term plausibility (Step 4):

It is assumed that the analyst possesses some (imperfect) information about the expected trend of PFS and OS in the long-term. In HTA, analysts consult clinicians to understand the expected long-term trend or look at external data in order to understand the expected long-term survival. However, this information is imperfect. Of the distributions that provide a good visual fit to the KM for OS and PFS (within 95% CI), predictions for PFS and OS for the remaining models are generated when 10% of events are remaining in the uncensored dataset. Distributions that predict a survival probability between +/-5% (i.e 5%-15%) of the truth (from the uncensored dataset) are considered to provide a reasonable long-term extrapolation (Step 4). As mentioned previously, the margin was increased if no predictions were within this margin. An arbitrary margin of +/-5% was selected in order to represent the plausible base-case scenario. A greater margin of error would represent a case where the analysis was less sure about what the long-term survival would be. Conversely, a narrower margin of error would represent a case where the analyst knows the truth with more certainty. Assuming an arbitrary margin of error of +/-5% was considered to be more reflective of reality, compared with assuming a best-case in which the truth is known with certainty, or a worst-base whereby there is high uncertainty about long-term outcomes. Plausibility was defined as follows:

- For the PSM model, whereby survival distributions are fitted directly to OS and PFS (and therefore the fitted functions are equal to the prediction for OS and PFS), plausibility is based on extrapolation associated with the direct fit to the data.

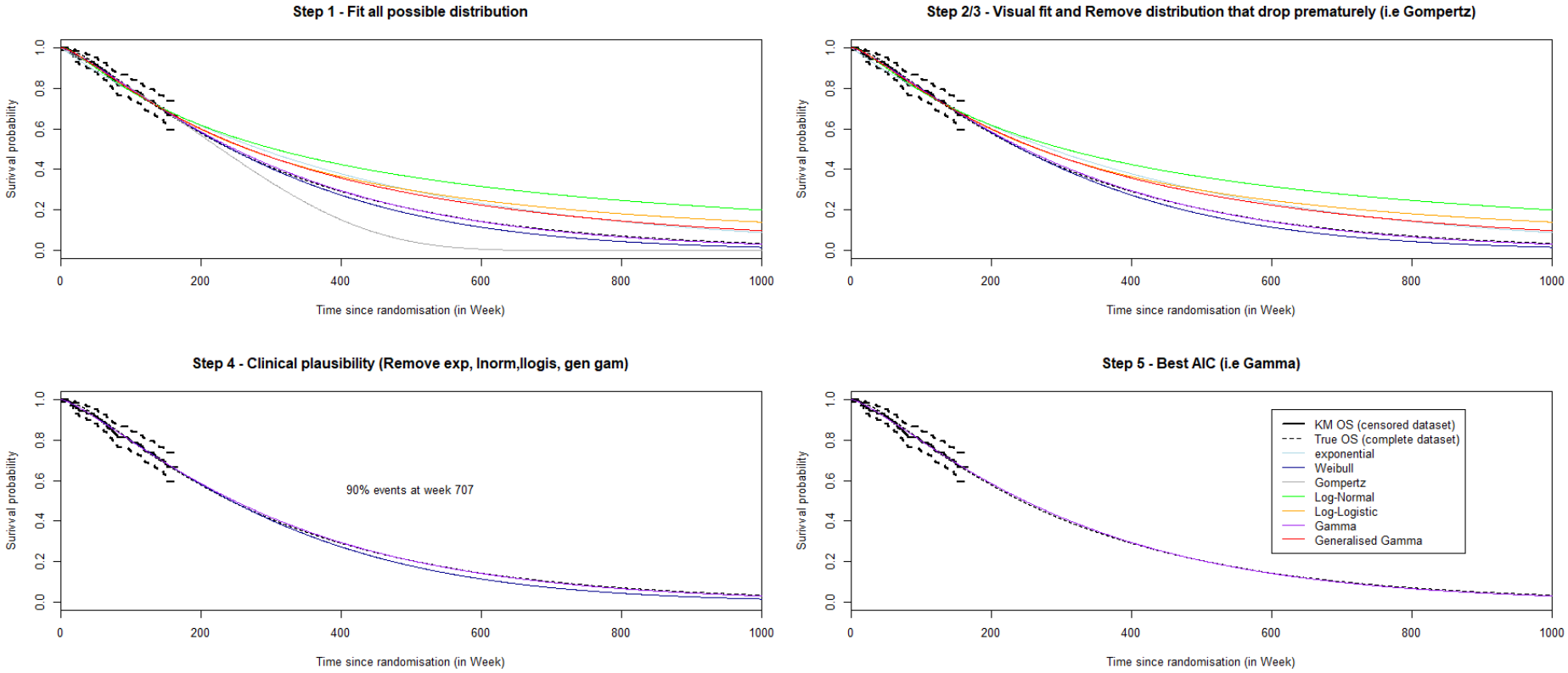
- For the variations of the STM assessed in this simulation study (whereby PPS is adjusted or unadjusted), the selection process for the PFS curve was the same as that used for the PSM for PFS as the distribution is fitted directly to the data. The selection process for the PPS curve, however, is based on the OS predictions from the STM (i.e. the resulting OS after combining all transitions) rather than the expected PPS. This approach was adopted to: (a) reflect the information that is likely to be provided by clinical experts advising on long-term OS projections, and (b) allow a fair assessment against the PSM. As described in Chapter 7, PPS is typically estimated in a subset of people who progressed, who may have different outcomes compared with the people who have not yet progressed. Therefore, basing the plausibility on the estimated OS is more likely to be fairer to the STM when compared with the PSM. Consequently, following the selection of the PFS distribution, all combinations of OS for the seven PPS distribution examined are run, with only PPS distributions that provide a plausible OS estimate being included next stage.
- Similarly, for the MSM, the choice of survival distributions for TTP, PrePS and PPS is based on the predictions of PFS and OS after combining all transitions. A two-step approach is used. First, the combination of curves for TTP and PrePS that provide a reasonable PFS projection is selected (i.e. 49 combinations). Secondly, OS is generated for the seven possible PPS distributions (given the selected PFS), with only PPS distributions that provide a plausible long-term prediction for OS, kept for the next stage.
- The Li model uses transitions based on Weibull distributions, therefore there was no requirement to select alternative distributions within this method.
- Finally, for the copula model, the distributional form for OS was assumed to be same as that for the PSM (independent) model. Consequently, only the choice of distribution for TTP was necessary, based on the expected prediction for PFS (Further detail in Appendix 12).

➢ Statistical fit (Step 5):

Finally, of the distributions included based on visual fit and long-term plausibility, the distribution with the best AIC is selected. The AIC was used instead of the BIC as it was easier to compute for use in combination with the Fu model. The choice between AIC and BIC as a measure for statistical fit is not expected to impact results from this simulation study.

An example of the model selection process used in the simulation study is provided below, based on hypothetical OS data (see Figure 50) for the PSM model (given OS is fitted directly to the data). In this example, the Gompertz distribution (grey curve) is considered inappropriate as it reaches zero prematurely and is therefore excluded. OS is expected to reach 10% at around Week 707 (based on the "truth") with a margin of error of +/- 5%.

Figure 50: Illustration of the selection process for the most appropriate extrapolation distribution



Therefore, in this example, the exponential, Log-Normal, Log-Logistic and Generalised Gamma distributions are excluded as these distributions predict an OS which are inconsistent with the truth (<5% or >15% at week 707). Of the remaining two distributions (the Weibull and Gamma), the Gamma distribution (purple curve) has the lowest AIC and is therefore selected.

8.3.4.1 Adjustment for general population mortality

General population mortality is included when estimating health state sojourn time to prevent logically inconsistent situations whereby the model predictions indicate a better survival prognosis for patients who have the disease compared with people who do not have the disease.

The expected weekly hazard of death from the general mortality is taken from UK life tables, assuming the median starting age of patients observed in each dataset. Predictions of OS and PFS are then adjusted accordingly, if necessary, such that the weekly unadjusted hazard for PFS and OS predicted remains higher than or equal to the expected hazard of death for the general population.

For the STM approaches, both PrePS and PPS were adjusted for general mortality. However, for the MSM, only PrePS was adjusted for general mortality. It was not possible to easily adjust PPS for general mortality as the hazard for PPS was not dependent on TTP. Consequently, the final estimates for PFS and OS in the MSM were adjusted for general mortality, rather than transitions used to reach this estimate. This is a possible limitation that needs to be considered when interpreting results and comparing results for the MSM vs. the STM in this simulation study.

8.3.5 ADEM[P]: Performance measures

Morris *et al* (2019) identified a number of performance measures in previous published simulation studies that have been used to assess the performance of a method. The authors recommend that the performance measure should depend on the aim of the study.

Four key performance measures are considered in this simulation study, in addition to their Monte Carlo (MC) standard error;

- Biases,
- Empirical SE / Coefficient of variation
- (Root) mean square error (MSE),
- Coverage

The MC SE provides an estimate of the uncertainty in the Monte Carlo sample. The model-based SE is also calculated and reported for transparency and completeness.

It is also important to inspect distributions and identify outliers.[139] Given the number of scenarios and methods assessed, the process was automatized and 5% of samples at the extreme were removed. Results are therefore presented for 950 samples for each scenario (rather than 1,000). This is a simplification.

The `rsimsum` function in R was used to both extract key performance measure and generate nested loop plots. [149] Nested loop plots are a type of diagram that allow the presentation of all information about the results of a simulation study with respect to chosen criterion in a single picture and therefore make it easier to look at results across the range of scenarios examined.[150]

8.3.5.1 Biases

Bias (denoted β) is calculated as the difference between the true mean population OS (θ) and the mean OS estimated by methods examined to estimate health state sojourn time ($\hat{\theta}$) so that:

$$\text{Bias} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \widehat{\theta_i} - \theta$$

The average biases and relative biases are computed in this simulation study. The relative bias is convenient to facilitate comparison between approaches. Absolute bias is also computed to provide an order of magnitude of biases either side of the true population mean.

The MC SE of the estimate is computed as:

$$\sqrt{\frac{1}{n_{sim}(n_{sim}-1)}\sum_{i=1}^{n_{sim}}(\hat{\theta}_i - \bar{\theta})^2}$$

8.3.5.2 Empirical SE / coefficient of variation

Whilst biases are an important measure, biases cannot be interpreted on their own, and it is important to also consider precision. Consequently, the empirical SE is also described (defined as the SE calculated in the probabilistic results).

In order to help interpreting the empirical SE, the relative empirical SE to the predicted mean (otherwise named as coefficient of variation) was generated (empirical SE / mean OS) for each scenario.

8.3.5.3 MSE

This is the sum of the squared bias and variance.

$$\text{MSE} = \frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}(\hat{\theta} - \theta)^2$$

Morris et al (2019) suggest that when MSE is a performance measure, the data-generating mechanism should include a range of values for the number of observations (sample size in the trial) given that the MSE will be sensitive to this. This was not considered necessary in this simulation study given that the number of observations was selected to reflect the number of patients typically observed in clinical trials and therefore reflects the expected performance of the methods in practice.

The MSE is a convenient measure as it integrates both biases and variance into a single measure. The empirical SE in contrast, gives precision but if a method is biased we need to interpret the empirical SE carefully. The Root of MSE is reported for simplicity. The MC SE of the MSE is also computed with the formula available in Morris *et al* (2019).[139]

8.3.5.4 Coverage

An approximation of the coverage of each method is also computed. The coverage typically refers to the coverage of confidence intervals, defined as the probability that a confidence interval contains θ. A

method is therefore considered good when the coverage is approximately 95% reflecting that the true value is contained within the 95% CI.

The methods under consideration combine information from different transitions rather than estimating a 95% CI directly, and therefore the coverage is given by the combination of transitions varied within their 95% CI. Furthermore, as previously described, input parameter values are drawn by random sampling for each distribution repeated 1,000 times using the `normboot.flexsurvreg` function in R which simulates alternative parameters from the asymptotic normal distribution of parameter estimates under sampling uncertainty. Therefore, this may not fully reflect the 95% CI. The mean OS over the lifetime is also calculated as the AUC estimated by a method and is therefore highly dependent on the extrapolation. Consequently, whilst the 'true' mean OS over the lifetime could fall within 95% of the predicted OS for a particular method, prediction during the observed period could fall outside the 95% CI for the OS KM. The coverage estimated in this simulation study only represents an approximation of the true coverage. Despite being an approximation in this simulation study, assessing the coverage provides an insight into the performance of a method. The coverage is defined as:

$$\text{Coverage} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} 1(\hat{\theta}_{low,i} \leq \theta \leq \hat{\theta}_{upp,i})$$

Two approaches are considered in this simulation when estimating the coverage;

1. **Method 1**: measured by considering whether the true value fell within the PSA predictions. Over-coverage is expected. This is likely to happen because using the mean time in health state is challenging when judging the performance of a method given that the mean may be unbiased, despite a poor visual fit.
2. **Method 2**: The true mean falls within the calculated 95% CI. Under-coverage is expected when the underlying data are non-normal, and therefore using the calculated SE to estimate the 95% may not be reliable.

In this simulation, I defined over-coverage as a coverage of >95%. Under-coverage was defined using 80%. It should be noted that other thresholds could have been selected

The Monte Carlo SE of the coverage is also computed with the formula available in Morris *et al* (2019).[139]

8.4 Results

Scenarios in this simulation study are defined according to four key observable (or at least partially observed) data characteristics: (1) the dependence (moderate negative, independence, moderate positive) between TTP and PPS, (2) the proportion of recorded PFS events (low, moderate, high) amongst all patient, (3) the ratio of OS to PFS events (low, moderate, high) and (4) the proportion of pre-progression mortality (low, high) in patients with a PFS event. Using a 3x3x3x2 factorial design resulted in a total of 54 scenarios. While the same underlying hazard of progression (TTP) is assumed for all scenarios; the simulated hazard of death (OS) generated for each scenario was not. A summary of the characteristics for each of the 54 simulated scenarios is provided in Table 12, with the KM plots reported in Appendix 10.

The performance of methods investigated is likely to vary depending upon the simulated scenarios. Due to the large number of methods and scenarios assessed, detailed results are not presented for every method, simulated scenarios and targets. Instead, rather than focusing on individual results, key trends for the performance of methods are summarised in this Section. Furthermore, results presented within the main body of this thesis focus on the predicted mean lifetime OS, with reference to results for prediction to the time in the progression-free health state, when necessary (given similarities between approaches). Results using QALYs are not discussed in the body of this thesis, as they followed generally the same trend as for OS. For transparency, and completeness, full results are presented in Appendix 13 for each of the 54 simulated scenarios, for every method and for every target.

The relative performance for all methods (PSM, STM, STM with PPS adjusted using log of TTP, STM with PPS adjusted using TTP, MSM, Li's model, Fu's model) is described here. However, rather than comparing all methods directly against each other, for ease of reading, results are separated into the following subsections. In Section 8.4.1, I present details of the performance of the PSM and STM (See Section 8.3.4.1). The unadjusted STM is compared with the adjusted STMs, the MSM and the Li model in Section 8.4.2, 8.4.3 and 8.4.4 respectively. The comparative performance of the PSM (independent model) against the Fu model is presented in Section 8.4.5. A short summary describing the key points is provided in Section 8.4.6.

Although results are summarised in this Section, some interpretation is provided alongside to help the reader understand why a particular trend is observed and whether it is solely attributable to the observed data characteristics considered, other unobserved data characteristics, external factors like the simulated underlying hazard or curve selection process used.

Table 12 : Summary of characteristics for the scenarios investigated

Sc n	Dep	PFS %	PFS / OS	prePS / PFS	Sc n	Dep	PFS %	PFS / OS	prePS / PFS	Sc n	PFS %	PFS / OS	prePS / PFS	prePS / PFS
1	Neg	Low (45%)	Low (35%)	Low	19	No	Low (45%)	Low (35%)	Low	37	Pos	Low (45%)	Low (35%)	Low
2	Neg	Low (45%)	Mod (55%)	Low	20	No	Low (45%)	Mod (55%)	Low	38	Pos	Low (45%)	Mod (55%)	Low
3	Neg	Low (45%)	High (75%)	Low	21	No	Low (45%)	High (75%)	Low	39	Pos	Low (45%)	High (75%)	Low
4	Neg	Low (45%)	Low (35%)	High	22	No	Low (45%)	Low (35%)	High	40	Pos	Low (45%)	Low (35%)	High
5	Neg	Low (45%)	Mod (55%)	High	23	No	Low (45%)	Mod (55%)	High	41	Pos	Low (45%)	Mod (55%)	High
6	Neg	Low (45%)	High (75%)	High	24	No	Low (45%)	High (75%)	High	42	Pos	Low (45%)	High (75%)	High
7	Neg	Mod (65%)	Low (35%)	Low	25	No	Mod (65%)	Low (35%)	Low	43	Pos	Mod (65%)	Low (35%)	Low
8	Neg	Mod (65%)	Mod (55%)	Low	26	No	Mod (65%)	Mod (55%)	Low	44	Pos	Mod (65%)	Mod (55%)	Low
9	Neg	Mod (65%)	High (75%)	Low	27	No	Mod (65%)	High (75%)	Low	45	Pos	Mod (65%)	High (75%)	Low
10	Neg	Mod (65%)	Low (35%)	High	28	No	Mod (65%)	Low (35%)	High	46	Pos	Mod (65%)	Low (35%)	High
11	Neg	Mod (65%)	Mod (55%)	High	29	No	Mod (65%)	Mod (55%)	High	47	Pos	Mod (65%)	Mod (55%)	High
12	Neg	Mod (65%)	High (75%)	High	30	No	Mod (65%)	High (75%)	High	48	Pos	Mod (65%)	High (75%)	High
13	Neg	High (85%)	Low (35%)	Low	31	No	High (85%)	Low (35%)	Low	49	Pos	High (85%)	Low (35%)	Low
14	Neg	High (85%)	Mod (55%)	Low	32	No	High (85%)	Mod (55%)	Low	50	Pos	High (85%)	Mod (55%)	Low
15	Neg	High (85%)	High (75%)	Low	33	No	High (85%)	High (75%)	Low	51	Pos	High (85%)	High (75%)	Low
16	Neg	High (85%)	Low (35%)	High	34	No	High (85%)	Low (35%)	High	52	Pos	High (85%)	Low (35%)	High
17	Neg	High (85%)	Mod (55%)	High	35	No	High (85%)	Mod (55%)	High	53	Pos	High (85%)	Mod (55%)	High
18	Neg	High (85%)	High (75%)	High	36	No	High (85%)	High (75%)	High	54	Pos	High (85%)	High (75%)	High

Abbreviations: Dep = dependence; gen pop = general population; Mod = moderate; Neg = Negative; PFS = progression-free survival; PFS % = Proportion of PFS (progression or death) events; PFS / OS = Ratio of death to PFS events (i.e. PFS events / OS events); Pos = Positive; prePS / PFS = Proportion of pre-progression mortality; Sc n = Scenario number.

8.4.1 Performance of the PSM and (unadjusted) STM when predicting the mean lifetime OS.

In summary, while both the PSM and STM were prone to some biases (notably in case of moderate dependence), both approaches were generally reasonable at predicting the mean health state sojourn time for single trial arms in this particular simulation study; with the PSM generally associated with less biases than with the STM, but with less precision.

8.4.1.1 Individual performance of the PSM and STM

Nested loop plots for the percentage bias for the prediction of the mean lifetime OS for each of the 54 scenarios for the PSM and (unadjusted) STM are displayed in Figure 51.

The mean lifetime OS predicted using the PSM (i.e. by fitting a curve directly to the trial data) or unadjusted STM (i.e. estimating OS as a function of PFS and unadjusted PPS) was considered unbiased (variation of less than 0.5% from the true mean) in six of the 54 simulated scenarios examined; Scenarios 25, 28, 41, 49, 50, and 53 for the PSM and Scenarios 30, 31, 34, 50, 51, 54 for the STM. This means that, on average, when the PSM or STM was replicated 1,000 times, the average of mean predicted lifetime OS was similar or very close (less than 0.5%) to the true mean OS.

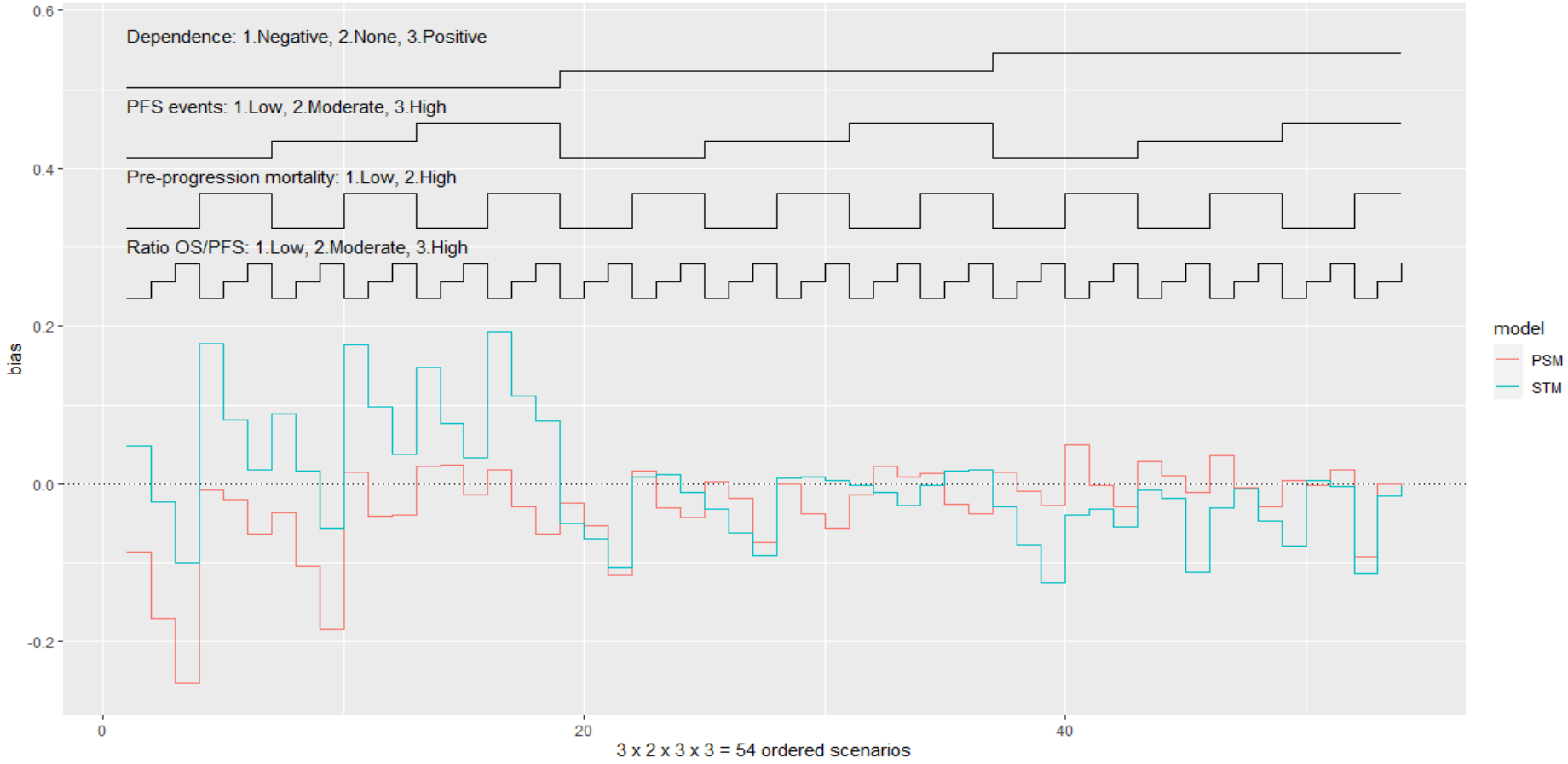
Nevertheless, bias remained generally low in the majority of simulated scenarios examined. The average bias was less or equal to 5% in 41 simulated scenarios using the PSM indicating that fitting a parametric function directly to OS to represent the hazard of death was generally reasonable (under the curve selection criteria used in this simulation study) in the scenarios investigated in this simulation study. Biases were less or equal than 10% in 49 scenarios. Using the unadjusted STM, biases were low ($\leq$5%) in 32 scenarios and less or equal than 10% in 44 scenarios. While reasonable, the PSM and STM were therefore prone to some biases.

The coefficient of variation (empirical SE as a percentage of the true mean) was greater than 5% in 52 scenarios using the PSM and in all of the simulated scenarios using the STM (n=54), and above 10% in 17 and 18 scenarios respectively, indicating some lack of precision for both approaches.

The coverage (the ratio of times the PSA predictions overlaped the true mean) was reasonable. The coverage was high (over-coverage; defined in this simulation study as >95%) for both the PSM and STM when measured by considering whether the true value fell within the PSA range predictions (method 1: within the PSA range) but was lower under method 2, when the true mean falls within the calculated 95% CI (based on SE). Under method 1, the coverage was greater than 95% for 51 scenarios using the PSM and 52 scenarios using the STM. Under method 2, the number of scenarios drops to 29 and 18, respectively. However, the coverage remained reasonable.

Figure 51 : Percentage bias in estimation for the mean lifetime survival (OS) using the PSM and STM

Please note the scale used for this graph - In percentage (0.2 represent 20%)



A possible explanation for the high coverage when using the PSA range is the fact that the mean time in health state is used in this simulation study, rather than whether the true KM is within the PSA predictions (visual fit). When estimating coverage under method 2, when the true mean fell within the calculated 95% CI, as expected, the coverage was generally lower and there was under-coverage in some occasion; but the coverage remain generally reasonable. This was expected given that the underlying data were generally non-normally distributed, and therefore using the calculated SE to estimate the 95% may not have ben reliable.

Key trends are presented/described in tabulated format, with a focus on only the key messages for ease of reading (given the wealth of information in Appendix 13). Key trends for the performance of the PSM according to data characteristics are summarised in Table 13, alongside some interpretation for particular results to help the reader understand the reason for these trends. Trends for the STM are summarised in Table 14; alongside some interpretation.

In summary, while the performance of the PSM and unadjusted STM in predicting the mean OS appear to be affected by both the dependence in the data, and the proportion of pre-progression death (Figure 51); a key factor for their performance was unobserved data characteristics; in particular the shape of the hazard during the unobserved period and whether there was any turning point. Indeed, the poor performance of the PSM in the case of negative dependence was largely explained by the problems of standard models in representing more complex hazard of death, notably when PrePS was low (hazard was decelerating after the observed period, with the long-terms hazard plateauing). For the STM, the poor performance in some of the scenarios was largely explained by: (i) biases associated with the use of PPS (generalisability to the overall population for PPS when estimated in a subset of patients who progressed), as well as (ii) limitations of the models in representing the hazard for progression (the simulated hazard for progression was associated with a tail, which makes estimation of PFS challenging and therefore impacted estimate for OS for the STM).

For both the PSM and STM, there was a trend toward reduced biases when data on OS were more mature. Indeed, the range for predicted biases tended to reduce as more OS events were reported. However, the trends observed here are weaker than expected due to the design of the simulation study. I would expect the trends to have been plainer if the simulation was designed differently and scenarios used the same hazard of death, but assumed different levels of censoring. In this simulation study, the simulated hazards of death were different between scenarios, and therefore the ability for the selected parametric function to represent those hazards is not directly comparable. Similarly, the range of coefficients of variation was reduced with larger OS events recorded and indicated less precision when data were less mature.

Table 13: Summary and interpretation of key trends for the performance of the PSM observed in this simulation study

Sensitivity of results to	Key trends	Interpretation
the dependence in the data	• Larger biases in scenarios with negative dependence. Biases ranged between -25.18% and 2.43%, -11.53% and 2.31% and -9.23% and 5.01% in simulated scenarios with negative, no or positive dependence. • OS predictions were generally under-estimated in the majority of scenarios with negative (14 out of 18), no dependence (13 out of 18) or positive dependence (10 out of 18). • Estimates were more precise in scenarios with negative dependence (when considering the coefficient of variation and RMSE). • The coverage was generally worse in simulated scenarios with negative dependence.	• Larger biases were explained by clearer change in the hazard of death beyond the observed period for scenarios with negative dependence. The simulated hazard was typically decelerating beyond or close to the end of the observed period of the trial, resulting a long tail in the underlying simulated true survival curve. It was difficult for the fitted curve to reflect the change in hazard beyond the observed period. Therefore bias for PSM werenot really related to the dependence, but more about whether hazards changed post observed period • Under-estimation explained by the presence of tail in the simulated hazard of death. This was more pronounced for the scenarios with negative dependence as the hazard of death was more accelerating beyond the observed period, but decelerating beyond the observed period. The fitted curve in cases of no or positive dependence allowed a better reflection of the hazard (as the turning point was less sharp). • Better precision for scenarios with negative dependence as scope for extrapolation to become erratic is reduced due to the accelerating hazard of death during the observed period • Worse coverage for scenarios with negative dependence as biases are higher but estimates are also more precise.
pre-progression mortality	• Smaller biases in scenarios with higher pre-progression mortality	• In simulated scenarios with higher pre-progression mortality, simulated individuals are more likely to die prior to progression, and therefore the simulated tail for the survival function becomes less pronounced for those scenarios, allowing the fitted parametric function to better reflect the hazard of death.

Table 14: Summary and interpretation of key trends for the performance of the unadjusted STM observed in this simulation study

Sensitivity of results to	Key trends	Interpretation
dependence in the data	• Larger biases in scenarios with negative dependence. Biases ranged between -10.03% and 19.37%, -10.57% and 1.74% and -12.47% and 0.37% in scenarios with negative, no or positive dependence. • OS predictions were generally over-estimated in case of negative dependence (15 of 18) and under-estimated in case of positive dependence (17 of 18) • Estimates were also in general less precise in simulated scenarios with negative dependence compared with those with no or positive dependence. • Better coverage in simulated scenarios with negative dependence	• In the case of negative dependence, because patients who progress early have a longer time to death following progression, PPS in the trial in the subset of patients who progress will be significantly higher compared with PPS estimated in all randomised patients; this is likely to lead to over-estimation of OS. • In the case of positive dependence, because patients who progress early have a shorter time to death following progression, PPS in the trial in the subset of patients who progress will be significantly lower compared with PPS estimated in all randomised patients; this is likely to lead to under-estimation of OS. • There was therefore more scope for bias when survival predictions are long (as in the negative scenarios) than when survival predictions are short (as in the positive scenarios). • Better coverage for scenarios with negative dependence due to the lack of precision.
pre-progression mortality	• Smaller biases in scenarios with lower pre-progression mortality	• Combination of two factors; (1) PFS was more biased (and under-estimated) in scenarios with lower pre-progression mortality compared with those with higher pre-progression mortality and (2) generalisability of PPS to randomised population • Also, in scenario with lower pre-progression deaths, more patients and events contributed to the PPS transition

8.4.1.2 Comparative performance of the PSM vs. STM

While the PSM performed slightly better compared with the STM in this particular study, notably in case of positive dependence, predictions were relatively similar for the majority of scenarios. Differences between the approaches were generally smaller when data on OS were more mature. As expected, the range for predicted differences tended to reduce as more OS events were reported.

There was a noticeable difference between the PSM and STM (defined as a difference in percentage bias of 5% or more) in a third of the simulated scenarios; most were in scenarios with negative dependence (n=13 out of 18), although four were in scenarios with positive dependence and one was a scenario with no dependence.

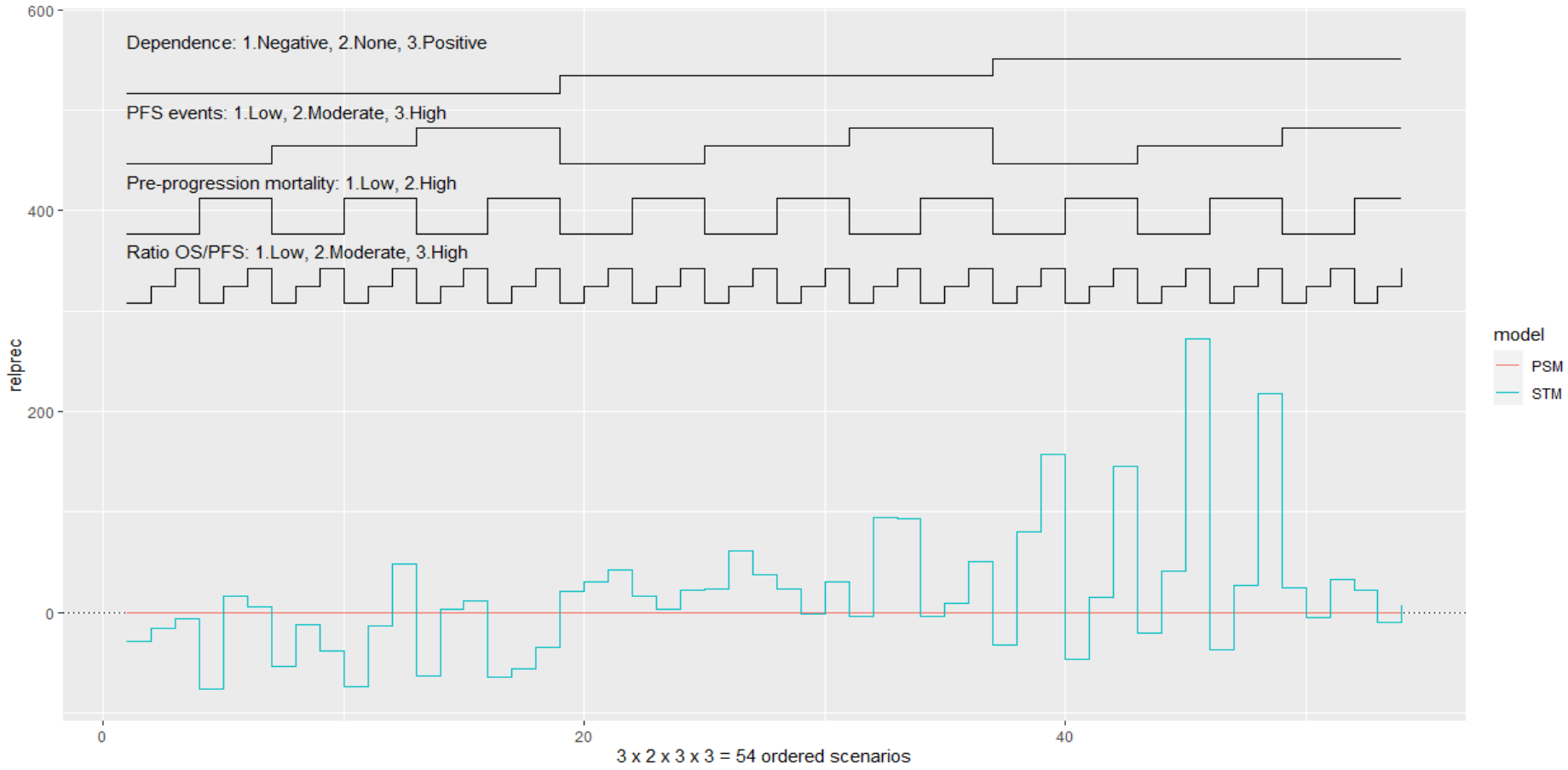
Of those scenarios with negative dependence where there was some noticeable difference (n=13), the PSM performed better compared with the STM in most scenarios (n=9), notably in those where the number of pre-progression deaths was higher. As previously highlighted, biases for the PSM were large in scenarios with negative dependence and low PrePS; this is attributable to the simulated hazard.

Of the four scenarios with positive dependence where there were noticeable differences between the PSM and STM, the PSM had less bias in all scenarios.

Ideally a method is both unbiased and precise. When looking at the precision, the STM was generally more precise compared with the PSM in the absence of dependence or when positive dependence is simulated (Figure 52). This is because OS is estimated as a function of two transitions, estimated with perhaps more certainty due to the number of events contributing to these transitions (while the total number of event is the same between the PSM and STM, the relative number of events for each transition is larger, reducing uncertainty around the estimate). In the simulated scenarios with negative dependence, the STM was generally less precise compared with the PSM. This could be explained by the high precision of the PSM due to the accelerating hazard with limited large variation in the extrapolation.

Figure 52 : Relative precision for the STM compared with the PSM (positive values indicate better precision)

Please note the scale used for this graph - In percentage (200 represent 200%)



8.4.2 Performance of the STM where PPS is adjusted on the log or normal scale

The previous section discussed the performance of the unadjusted STM (where PPS is taken from the trial directly without any adjustment). In this simulation study, the performance for the STM was also examined for two variations of the STM whereby the PPS is adjusted by TTP on either (a) the log scale (referred to hereafter as "STM with PPS on the log scale") or (b) normal (non-log) scale ("STM with PPS on the normal scale"). Nested loop plots for the percentage bias for the prediction of the mean OS for each of the 54 scenarios for the STM adjusted or not (for comparison) are displayed in Figure 53 (Please note the differences in scale when comparing with Figure 51).

The average bias was less or equal to 5% in 29 simulated scenarios using the STM with PPS on the log scale; most scenarios with no dependence (n=15/18) and more than half of scenarios with negative dependence (n=10/18). In contrast, biases were low only in a minority of scenarios with positive dependence (n=4/18). Using the STM with PPS on the normal scale, the average bias was less or equal to 5% in less than of half of the simulated scenarios (n=22); none with positive dependence and more than half of scenarios with negative (n=11/18) or no dependence (n=11/18).

A summary for the comparison against the unadjusted STM and the key trends are presented in Boxes 1 and 2, respectively.

In summary, predictions for the mean OS were relatively similar between the unadjusted and adjusted STM in the absence of dependence. As predictions were already reasonable with the unadjusted STM, the scope for improvement was reduced in those scenarios. These methods however did not worsen predictions. Predictions were however less precise.

In contrast, both the STM with PPS adjusted on log and normal scale typically improved predictions in scenarios with negative dependence (where the unadjusted STM did not perform well), but worsened predictions in scenarios with positive dependence. The adjusted STMs were also less precise in case of positive dependence.

The STM with PPS in the normal scale was less precise compared with the STM with PPS in the log scale and was associated with significantly large biases in case of positive dependence.

The coverage for both the STM with PPS adjusted on log and normal scale typically larger compared with the unadjusted STM.

Box 1: Key trends for the performance of the STM with PPS on the log scale

- Similar performance (in terms of percentage bias) to the unadjusted STM in case of no dependence (no noticeable differences)
- Noticeable differences (difference in percentage bias of 5% or more) for 13 of the 18 scenarios with negative dependence. Predictions were improved in most scenarios (n=10/13)
- Noticeable differences for 13 of the 18 scenarios with positive dependence. Predictions were worsened in most scenarios (n=10/13)
- The STM with PPS on the log scale was less precise (Figure 54) compared with the unadjusted model in most simulated scenarios with no or positive dependence. The adjusted STM on the log scale was more precise compared with the unadjusted model in about half of the scenarios with negative dependence.

Box 2: Key trends for the performance of the STM with PPS on the normal scale

- Noticeable difference in 6 scenarios with no dependence (predictions improved in half cases)
- Noticeable difference in 8 scenarios with negative dependence (predictions improved in most cases [n=7/8])
- Noticeable difference in almost all scenarios with positive dependence (predictions worsen in all cases [n=17/17])
- The adjusted STM on the normal scale was consistently worse than the adjusted STM on the log scale in all simulated scenarios with positive dependence
- The STM with PPS on the normal scale was less precise compared with the unadjusted model (Figure 54), in all but 4 scenarios (all with negative dependence)
- Compared with the STM with PPS on the log scale, the STM with PPS on the normal scale was also generally less precise in the majority of scenarios examined (n=49/54).

Figure 53 : Percentage bias in estimation for the mean lifetime survival (OS) using the STM unadjusted vs. adjusted STMs

Please note the scale used for this graph - In percentage (0.5 represent 50%)

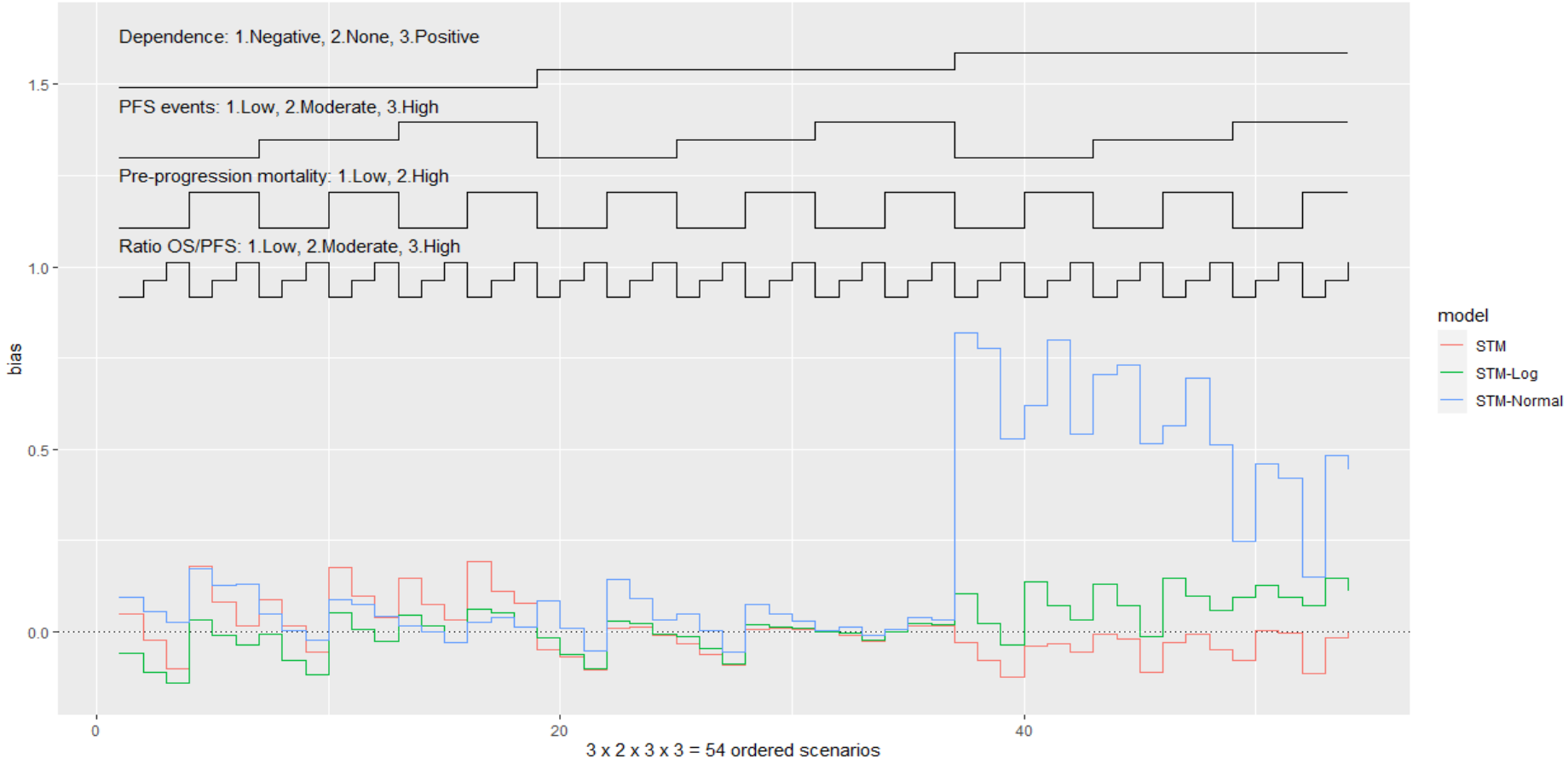
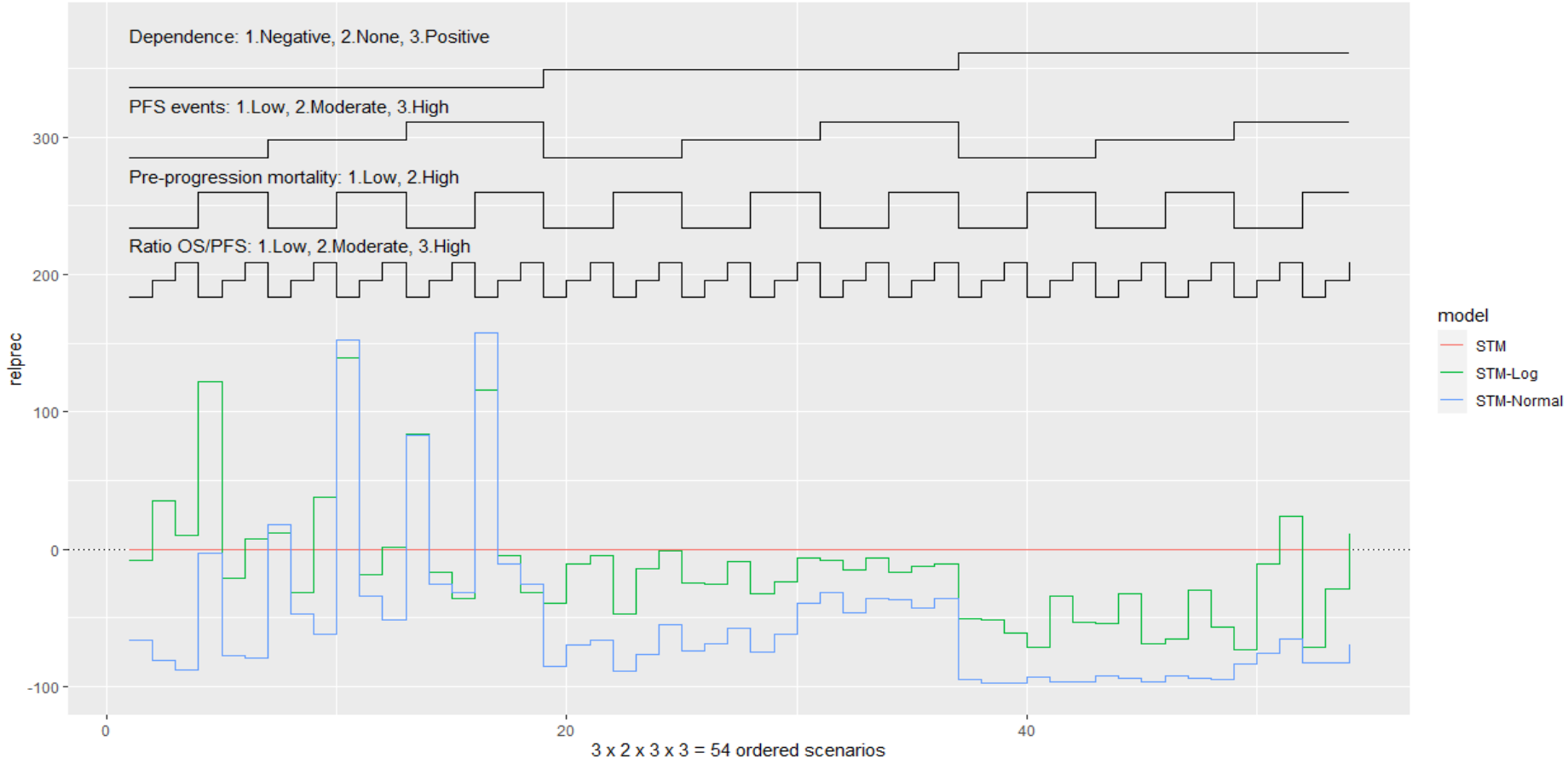


Figure 54 : Relative precision for STM adjusted vs. unadjusted (positive values indicate better precision) for the prediction for the mean lifetime OS

Please note the scale used for this graph - In percentage (200 represent 200%)



8.4.3 Comparative performance of the MSM vs. STM

In summary, as expected, the performance for the MSM and unadjusted STM was relatively similar (Figure 55 and Figure 56). Therefore, the simple approach (unadjusted STM) when curves are selected appropriately (same process) did not fare worse compared with the competing risk approach. Whilst small, there was some differences for a limited number of scenarios for the prediction of the mean lifetime OS between the MSM and STM; in particular in the simulated scenarios with higher pre-progression mortality.

The direction and level of biases were mostly attributable to differences in PFS predictions. The MSM (whereby PFS is estimated under a competing-risk framework) consistently generated lower OS predictions compared with the STM (whereby PFS is fitted directly to the data) in this simulation study, with the exception of 3 Scenarios (Scenarios 40, 46 and 53), but the differences were minimal. This is because PFS estimated using the MSM was typically lower (under-estimated) compared with PFS estimated using the STM in this simulation study which led to an under-estimation of OS. This is likely to be the result of the combination of two factors: (i) the simulated hazard for progression in this particular simulation study which had a tail, and (ii) the choice of curve for PrePS with the MSM which led to more death pre-progression compared with the direct modelling of PFS and therefore under-estimated PFS.

8.4.4 Comparative performance of the Li model vs. STM

The Li model consistently under-estimated OS and had more biases compared with the unadjusted STM in the simulated scenarios with no or positive dependence. The poorer performance of the Li model in predicting the mean OS is not surprising given the restriction placed on the Weibull distribution, and the resulting complex simulated hazards used in this simulation study.

The Li model, however, was noticeably less biased for OS compared with the unadjusted STM in seven scenarios (Scenarios 4, 10, 13, 14, 16, 17, 18), all of which had negative dependence, of which 5 had more complete PFS data.

Whilst the Li model was typically more biased (given the complexity of the simulated hazard in this simulation), it was more precise compared with the unadjusted STM in all scenarios. This led to poor coverage.

Figure 55 : Percentage bias in estimation for the mean lifetime survival (OS) using the STM vs. the MSM and the Li model

Please note the scale used for this graph - In percentage (0.5 represent 50%)

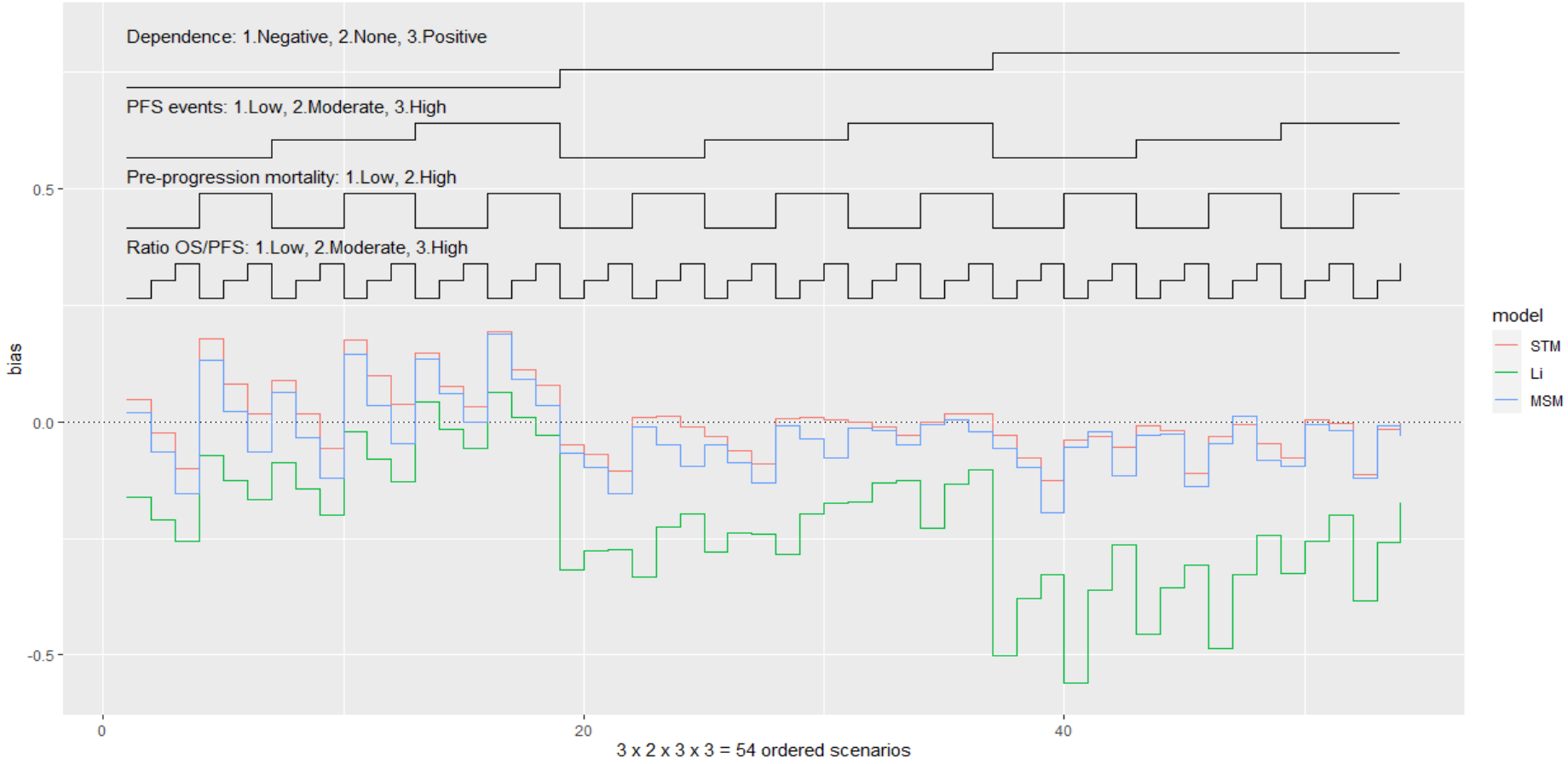
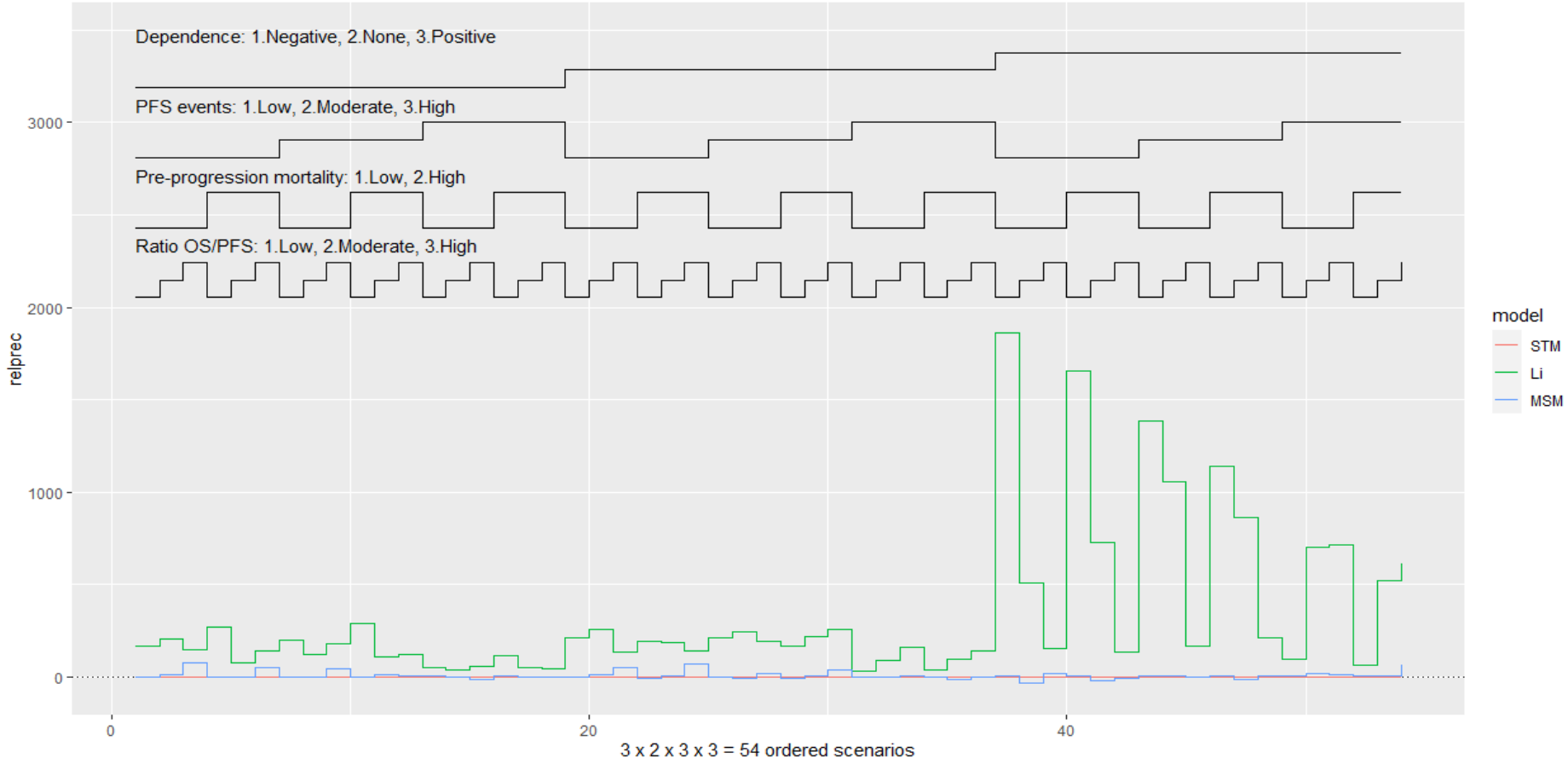


Figure 56 : Relative precision for the MSM and the Li model compared with the STM (positive value indicates better precision)

Please note the scale used for this graph - In percentage (1000 represent 1000%)



8.4.5 Comparative performance of the Fu model against the PSM (independent model)

In general, as expected, differences were minimal and predictions for OS were relatively similar between the Fu model and independent fitted models (Figure 57), as the same parametric distribution was selected for OS as part of the curve selection criterion defined in Section 8.3.4.1.

The same trends as those described for the PSM in Section 8.4.1 therefore applies (e.g. larger biases for scenarios with lower pre-progression mortality and negative dependence).

The coverage was slightly better than the PSM. Estimates were also generally more precise in cases of positive dependence.

A particular benefit of the copula is the joint modelling of OS and PFS under a semi-competing risk framework, and therefore, accounting for the structural relationship that exists between PFS and OS when estimating PFS, rather than the addition of an arbitrary constraint, as is done in the PSM to prevent PFS being higher than OS.

In this particular study, the Fu model did not perform well for estimating PFS and was more biased for the prediction of the time in the progression-free health state (Figure 59). This finding needs to be interpreted in relation with the simulated hazard considered in this simulation study and also the curve selection criteria. The simulation study showed that when the hazard (for both progression or death) is complex, the Fu model could lead to more biases. As previously discussed, standard models are often inappropriate in representing more complex hazards and struggle to reflect the turning point in the hazard and associated tail beyond the observed period. Because OS is acting as a semi-competing risk in the Fu model this will affect the estimation for PFS (as TTP is truncated by OS). The same hazard of progression (TTP) was also assumed for all scenarios and was complex in that it was already difficult for the fitted curve to reflect the change in hazard beyond the observed period. The curve selection process was also simplified to allow automation; which could have contributed to the poor performance of the Fu model in the scenarios examined. The combination of these issuesmay explain why the Fu model exhibited poorer performance in the simulations.

Figure 57 : Percentage bias in estimation for the mean lifetime survival (OS) using the PSM and the Fu model

Please note the scale used for this graph - In percentage (0.2 represent 20%)

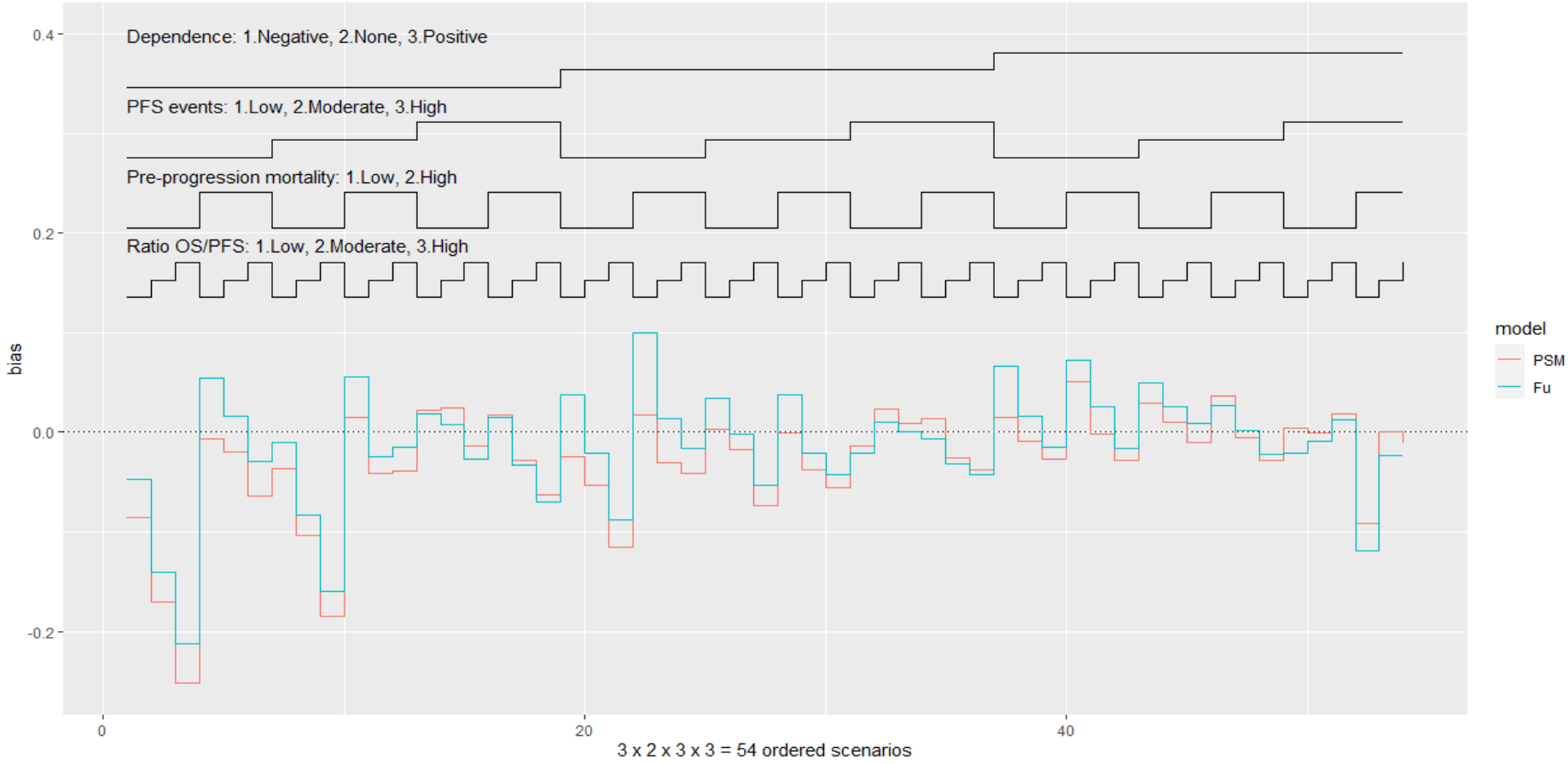


Figure 58 : Relative precision for the Fu model compared with the PSM (positive values indicate better precision)

Please note the scale used for this graph - In percentage (25 represent 25%)

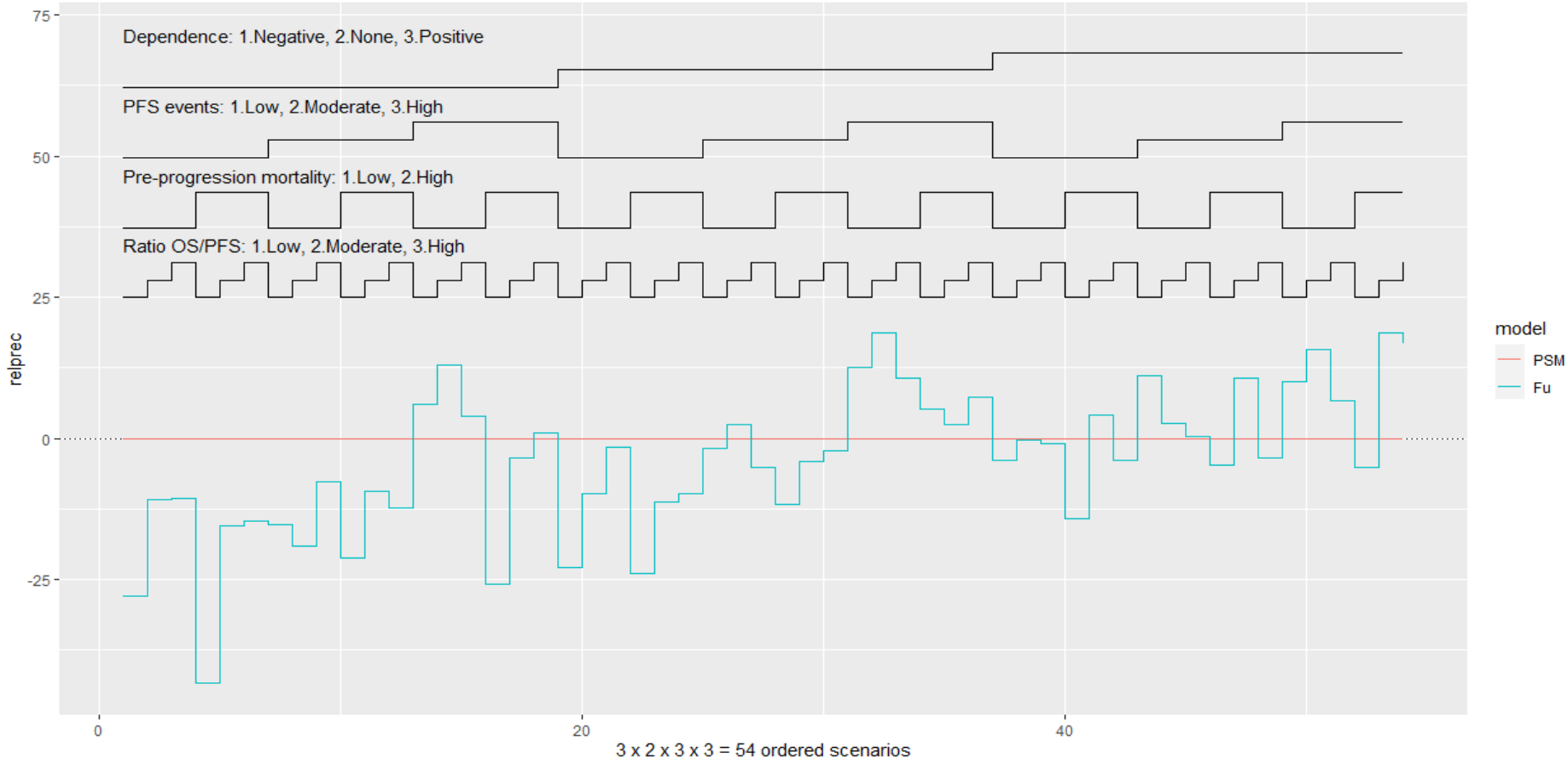
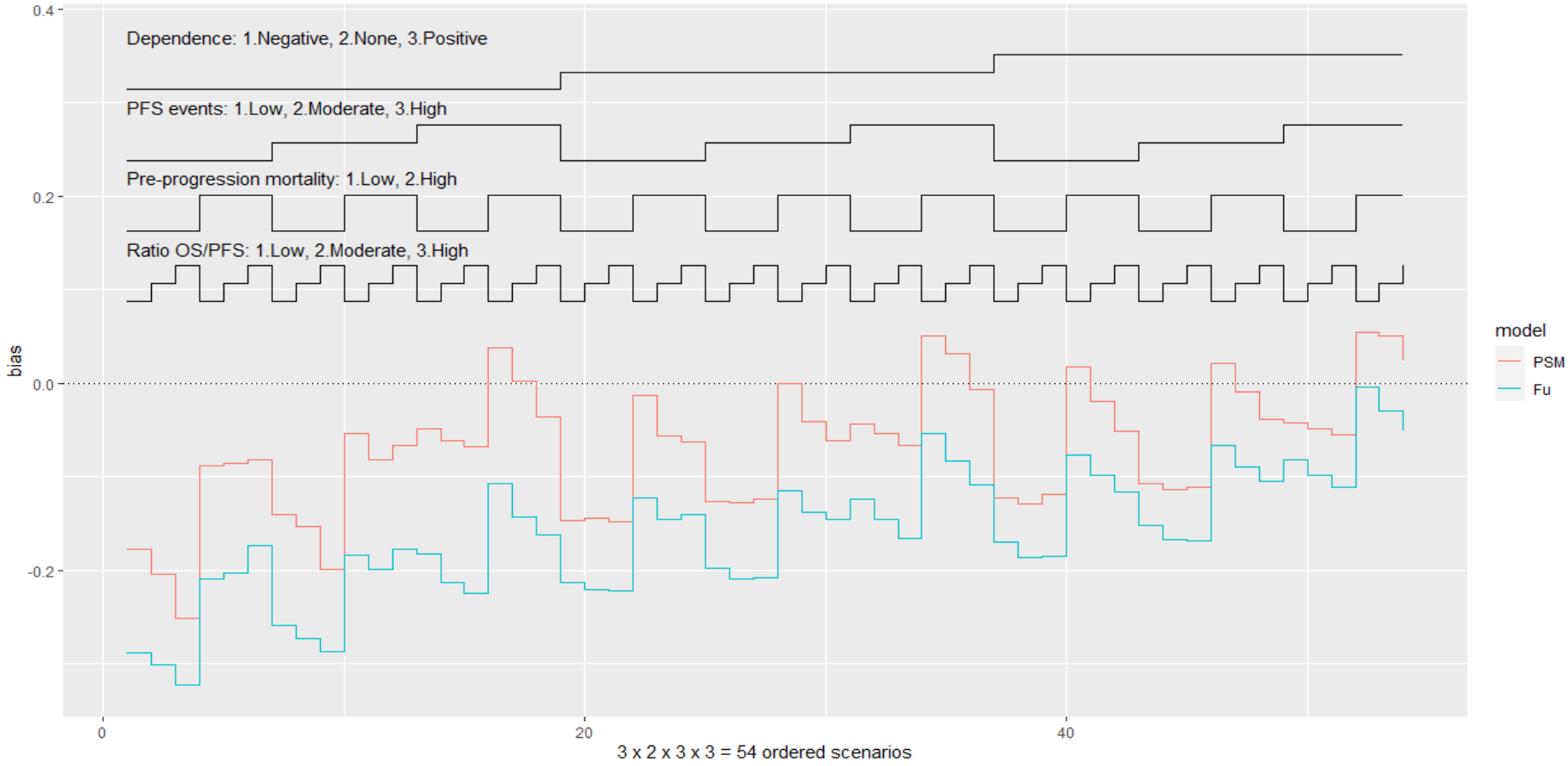


Figure 59 : Percentage bias in estimation for the mean lifetime time in progression-free (PFS) using the PSM and the Fu model

Please note the scale used for this graph - In percentage (0.2 represent 20%)



8.4.6 Comparison of methods - Summary

In summary, while all approaches were prone to biases, the PSM and STM were generally reasonable at predicting health state sojourn time in single trial arms. The PSM was associated with less bias compared with the STM in this particular simulation study; but was less precise.

While the data characteristics assessed in this simulation study explained some of the trends in results, unobserved data characteristics such as the underlying simulated hazard were a larger contributor for particular results, in addition to the curve selection method. There was difficulty in representing the hazard of progression using the standard parametric survival models explored in this simulation study, leading to PFS under-estimation. This had implications for the estimation of OS in the STM, which affected performance. The complexity of the simulated hazard and ability to fit standard parametric models (explored in this study) also affected the performance of the Fu's model.

In this simulation study, all approaches were more prone to biases in the simulated scenarios that included negative (moderate) dependence between TTP and PPS. The PSM, under-estimated OS as it was not able to reflect the shape of the simulated hazard of death which was quickly accelerating during the observed period, but decelerating beyond the observed period. The STM, whereby OS is estimated as a function of PFS and PPS, similarly, could not accurately reflect the simulated hazard of death for the underlying truth, due to: (i) similar challenges with the extrapolation of PFS which was under-estimated, and (ii) the generalisability of PPS estimated only in a subset of patients who progress. The PSM was more appropriate than the STM in the simulated scenarios that included higher pre-progression mortality. This is because in these scenarios, the underlying truth had a shorter tail. In contrast, the STM was more appropriate in those with lower pre-progression mortality due to the large biases observed for the PSM, but also the counteractive effect of lower PFS prediction together with PPS over-estimation.

In the simulated scenarios assuming independence of TTP and PPS, both the PSM and STM were appropriate and had similar performance and therefore neither method is favoured over the other.

In the simulated scenarios in which positive (moderate) dependence was assumed, the PSM was generally appropriate and the unadjusted STM was more prone to biases compared with the PSM in a limited number of scenarios, notably those with lower pre-progression mortality. The PSM appeared to fare better in those scenarios.

There was a general trend for the performance of methods to be improved when data were more mature, despite the simulated hazard being different between scenarios.

To account for dependence in the data, adjusting PPS using TTP as a covariate on the log scale, typically improved biases or did not lead to noticeable differences compared with the unadjusted STM in the majority of the simulated scenarios where there was a negative dependence, no dependence or positive dependence with higher pre-progression mortality. Adjusting PPS on the log scale was more prone to bias in scenarios with positive dependence with higher pre-progression mortality. Adjusting PPS on the normal scale was typically more prone to biases (compared with the unadjusted STM or adjustment in log scale), in particular for scenarios with positive dependence.

The performance of the unadjusted STM (where PFS is used directly) was relatively similar to the performance of the MSM using the `mstate` function. The simple unadjusted STM approach did not fare worse than the competing risk approach (MSM) when curves were selected appropriately (same process). The MSM, however, consistently generated lower OS predictions in this simulation study. This is attributable to the fact that the estimated PFS with the MSM was consistently lower, due to the effect of pre-progression mortality. Some of those differences were also attributable to the choice of curve (based on an automated process), rather than the method itself.

While the performance of the Li model was limited in this simulation study, the approach had low biases in a number of scenarios indicating that this approach was sometimes reasonable, despite the assumption that all transition follow a Weibull distribution (with the same shape parameter).

In this particular study, the Fu model was more prone to biases for the estimation of the mean time in progression-free compared with the direct PFS fit. This is explained by the complexity of the hazard considered (all scenarios also used the same hazard of progression; and therefore, it is challenging to make generalisable conclusions) and also the simplicity of the curve selection process used in this simulation study. While the Fu model remains a potentially appropriate approach for jointly modelling OS and PFS in a way which preserves the structural relationship between the outcomes, this study showed that when the hazard of progression and death are complex, the approach could lead to more biases (as OS is acting as a semi-competing risk).

Whilst biases are an important measure of performance, the precision of methods also needs to be considered. Whilst biases for the PSM and STM were small in a number of simulated scenarios, both approaches were generally imprecise when looking at the RMSE and empirical SE indicating that there was a large variation in the OS prediction when the models were replicated 1,000 times. The STM was generally more precise compared with the PSM in the case of no or positive dependence. Precision was lost when adjusting PPS on the log and non-log scales. Both the PSM and STM had a good coverage.

Although not shown in the main body of this thesis, in health economics, QALYs are usually the key clinical outcome of interest. Similar conclusions to those for OS were observed (appendix 13).

8.5 Discussion and conclusions

This chapter assessed the comparative performance of methods currently used in health economics to estimate health state sojourn time, as well as other methods identified in Chapter 6, in a simulation study to identify whether particular methods fare better than others in particular situations within single trial arms. In summary, while approaches were more prone to biases in some of the simulated scenarios, in particular those with negative dependence, the two commonly used approach in health economics, the PSM and STM were generally reasonable at predicting health state sojourn time in this particular simulation study when applied to single trial arms. The PSM fared better compared with the STM in this simulation study; but was less precise.

Approaches could not be selected based on the observed data characteristics alone as their performance was mostly affected by unobserved data characteristics such as the complexity of the underlying hazard (for both OS and progression) and the curve selection process. As with any study, there are strengths and limitations which need to be recognised, and findings need to be interpreted in light of these.

8.5.1 Strengths

This study has a number of strengths:

- the study follows the ADEMP structure developed by Morris et al[139] and therefore conforms to the latest recommendations for the reporting and conduct of simulation studies.
- the performance of methods was evaluated during both the observed and full unobserved period using simulated data. This is not possible using published trial data. Other studies compared the fit of methods to either the observed period only of a trial, or only part of the unobserved period by comparing the fit to a later data cut (where data are not complete either).
- simulated scenarios were generated to be realistic, using data from real trials to inform the simulation, but also characteristics observed in trials used in HTA.
- the curve selection process was also automatized taking into account the recommended steps in identifying appropriate survival curves. This allowed a fairer comparison between methods. Studies typically compared approaches using different curve selection processes; rendering any comparison challenging to interpret as differences in predictions are mostly the results of the extrapolation method, rather than the analytical approach itself.
- although the natural history model uses a multi-state process, the STM and MSM assessed in this simulation study used a different process and inputs, and therefore results are not likely to be biased in favour of any approach examined.
- while not explicitly discussed in the main body of this thesis, predictions of QALYs were also considered, as well as PSA results to account for the uncertainty around the point estimate.

8.5.2 Limitations

In addition to strengths, this study has a number of limitations including:

- Data-generating mechanism

First, as with any study, there are challenges with generating data and scenarios that are realistic. Although every attempt was made in this simulation study for the data to be generated to be realistic, biases cannot be entirely avoided. The data were generated using a multi-state process as this was considered to provide an accurate representation of the natural history of cancer. There are also challenges in defining scenarios that are realistic. Fifty-four scenarios were included in this simulation study, defined according to the degree of censoring for PFS and OS, proportion of pre-progression death and dependence between TTP and PPS. Despite this, the generated scenarios do not cover all possible situations arising in practice. In order to identify patterns, some of the defined scenarios could be considered extreme, for instance the level of dependence assumed. In addition to scenarios assuming independence between TTP and PPS, scenarios with moderate positive or moderate negative dependence were defined (Kendall Tau = [0.4]). This level of dependence may be unrealistic. Furthermore, although different degrees of censoring were examined, the scenarios examined typically represented those with a high degree of censoring. The performance of all methods is likely to be improved when the degree of censoring is lower, as the need for extrapolation beyond the observed period becomes less and therefore differences between approaches are likely to be reduced. Perhaps more importantly, given the chosen data-generating mechanism, when examining different censoring levels, scenarios were not directly comparable (e.g. I did not compare the same trial with different censoring levels). Instead, each scenario represented different hazards of death, with different censoring characteristics, rather than the same underlying hazard with censoring introduced subsequently. This approach was chosen in order to ensure that results are not influenced by the selected form of the underlying hazard of death. Although the scenarios examined reflected different hazards of death, more complex hazards could be observed in practice.

- Simulated hazard of progression or death

The simulated underlying hazard of progression or death may have disadvantaged some approaches. This was evident when looking at the comparison of the STM against the MSM or the PSM against the Fu model. Consequently, findings from this simulation cannot be fully generalisable, as the simulated hazard was generally tail dependent, and it is not surprising for these approaches to be more prone to biases in those cases. It should be noted that the complexity of the hazard of progression or death also had some impact on the performance of the PSM and the STM. This was apparent when comparing

scenarios with lower or higher pre-progression mortality, where it was more challenging to identify a parametric function which reflected the underlying hazard in scenarios with lower pre-progression mortality due to the longer tail in the true underlying distribution.

- Mechanism of action

Related to the data-generating mechanism and simulated hazard, the simulated scenarios did not explicitly include the possibility of a cure, and therefore results may not be generalisable to new treatments where a cure is expected. However, as highlighted there was a tail in the underlying simulated hazard, but this only affected a limited number of patients, and therefore cannot be interpreted as the effect of a curative treatment. Only a limited number of standard distributions were also considered. This is a possible limitation, as more complex models could have provided more appropriate extrapolations.

- Curve selection mechanism

Another key challenge in this simulation study is the extrapolation approach. Irrespective of the approach considered, analysts need to make a decision about the most plausible parametric distribution for the possible transitions/survival endpoints, with the performance of the method more linked to whether an appropriate extrapolation is selected, rather than the validity of the approach itself. This therefore makes the conduct of a robust simulation study very challenging, as a large component of the performance of those methods relies on subjective judgements, of which decisions need to be made in the absence of evidence. In this simulation study, the process for curve selection was automatised to reflect the process that is typically followed in health economics. Automatising the process was necessary as curves had to be selected for each approach, for the 1,000 datasets for each of the 54 scenarios. Whilst this had the advantage of potentially reducing any unconscious biases when selecting curves, strict criteria had to be employed, and therefore, perhaps more appropriate curves would have been selected, should curves have been examined manually. This was apparent when comparing results between the STM and MSM and the PSM and the Fu model, where some of the differences in OS prediction were attributable to the curve selection process rather than the analytic approach itself. The curve selection criteria used in this simulation study, although believed to be realistic, could be deemed simplistic as only one time-point was used to determine clinical plausibility of the selected function. In practice, clinical experts may be able to provide more information to aid model selection. The process for curve selection may also differ between analysts, and there is no single agreed approach. The base-case curve selection criteria was also perhaps not adequate to select models that reflected the tail in the true underlying data (reliance on a single time point when 10% of events are left), explaining the poor performance of some of the approaches.

- Assessment of clinical plausibility

There is also an assumption that clinical opinion is available, which may not be possible for instance for a new intervention which is not part of clinical practice. It is possible that analysts may have no information on the long-term trend for OS for instance.

- Judging the performance of methods based on the mean lifetime OS

The mean time in health state (lifetime) was used in the base-case to assess the performance of methods, rather than evaluating the performance of method during the observed period. This introduced further challenges, as although the mean time may appear unbiased, the method under consideration may still provide a poor fit compared with the truth, as shown in Chapter 9. It is also difficult to interpret QALYs as this is the result of both utility values but also the mean time in two health states (therefore QALYs may appear unbiased, despite time in health state biased).

- Single trial arms approach

Exploring only one arm could also be considered as a possible limitation. This single trial arms approach was adopted in order to avoid the potential for spurious conclusions arising from apparently appropriate incremental outcomes despite the presence of a poor model fit in both treatment groups. However, in health economics we are interested in the incremental outcomes between treatments. Different approaches make different assumption about the treatment effect. Whilst an approach may be appropriate to model outcomes for one arm, the approach may be less appropriate when considering outcomes for an alternative treatment arm. For instance, a key rationale for using the STM is that it allows the model to capture the underlying disease process and therefore, data from the control arm (where more events typically occur) could be used to inform the process for the intervention arm.

8.5.3 Conclusion

This simulation study shows that that all approaches are prone to biases, but that fitting survival models to OS and PFS may be no better, or worse, than modelling the underlying disease process. All approaches have limitations which need to be recognised. With the PSM, it is unknown in a trial setting whether a parametric function fitted to OS would always represent the underlying hazard and what is the best extrapolation method. With the STM, similarly the best extrapolation method for PFS is unknown and there may be uncertainty regarding the generalisability of using PPS estimated only in the subset of patients who have progressed.

It is challenging to draw any definitive conclusions from this simulation study regarding which modelling method should be used given the characteristics of time-to-event data. Whilst some trends were observed, the performance of methods was primarily explained by unobservable data characteristics such as by the shape of the simulated hazard and the subjective curve selection mechanism. It is important to consider the clinical/biological rationale for the shape of the hazard and also presence of dependence. However, while it is informative to look at the hazard plot to see if there is a trend (e.g. increasing or decreasing), and if there is any indication of dependence within the data; data characteristics beyond the observed period remain unknown and speculative.

In this simulation, the performance of methods in single trial arms was explored. However, in health economics, the incremental health state sojourn time/QALYs are of importance, and therefore, it is equally important to assess the performance associated with different modelling approaches where different assumptions are made about the modelling of the treatment effect. This has the propensity to alter decision-making. This is explored in Chapter 9.

9 CHAPTER IX: EVALUATING THE PERFORMANCE OF METHODS USED IN HEALTH ECONOMICS TO ESTIMATE THE INCREMENTAL HEALTH STATE SOJOURN TIME/QALYS: CASE STUDIES

9.1 Chapter overview

This chapter evaluates the performance of the PSM and STM in estimating the health state sojourn times and QALYs, incrementally, and also separately, for a control arm (part of routine practice) and a new intervention arm for which less information is available. The performance of the methods is judged using real cancer trial datasets, and in addition to the predicted mean health state sojourn times and QALYs, the visual predictions are considered.

Section 9.2 introduces this Chapter. The methods for these case studies is presented in Section 9.3. Results for each case study are described in Section 9.4. In Section 9.5, I provide a conclusion on whether an approach is more accurate compared with others in estimating the incremental health state sojourn times/QALYs in the case study considered, and discuss the potential implications.

9.2 Introduction

Chapter 8 presented a simulation study to evaluate the performance of methods commonly used in health economics, alongside additional methods identified in Chapter 6 to estimate health state sojourn time/QALYs. This simulation study was conducted in single trial arms for pragmatic reasons and ease of interpretation and showed that both the PSM and STM were generally appropriate when estimating health state sojourn time in single trial arms.

In health economics, the incremental health gain (QALYs) associated with the introduction of a new technology is a key outcome of interest; this is driven by health state sojourn time predicted by the model for both arms.

When trying to predict the effect associated with the introduction of the new technology, the following elements need to be considered: (i) what is the relative treatment effect?; (ii) which outcomes/endpoints are affected by the new treatment?; (iii) how long does the treatment effectiveness last? and (iv) what happens when the treatment effect diminishes?

The effect of the new technology on OS is modelled differently between the PSM and the STM. In the PSM, a treatment effect (hazard ratio or time-varying treatment effect) is applied directly to OS estimated in the control arm for some duration (a patient's remaining lifetime or some shorter period). It should be noted that curves could also be fitted independently. In the STM, OS is estimated as a function of any gain in PFS (compared with the control arm), but also any potential gain (if any) in PPS.

Consequently, in addition to differences in the general prediction for health state sojourn time for a given arm (control arm), the different analytic approaches rely on different assumptions on how health state sojourn time would be estimated for the intervention arm.

A key criticism of PSMs is that they do not model the underlying natural history process, and therefore, it could be challenging to predict OS in the long-term for the intervention arm where less information on long-term outcomes is available compared with the control arm (which may have been used in practice for some time). External information, including clinical input, can help in selecting the most appropriate curves for the transitions or survival endpoints of interest, although this is invariably associated with uncertainty. In contrast, the long-term relative treatment effect associated with the introduction of a new technology is typically less clear. Clinical judgement regarding the plausible long-term extrapolation, and therefore the choice of curve, is therefore more challenging. The relative treatment effect on PFS and OS estimated in the trial is typically extrapolated over a lifetime horizon, or is assumed to diminish by some arbitrary time-point without any evidence to substantiate such assumptions.

In contrast, the STM which allows for the explicit modelling of the underlying natural history process may provide a more plausible prediction for the health state sojourn time/QALYs when external evidence or clinical validation are more limited for the intervention arm, by borrowing information from the control arm. An assumption commonly made in health economics (Chapter 5.3.5.5) is that PPS is the same between treatment arms (in the absence of significant differences in the trial), and therefore, it is argued that it is possible to borrow information from the control arm in order to have an estimate of OS for the intervention arm (by pooling PPS across arms, and therefore assuming no treatment effect after progression). Such an approach equates approximately to assuming that any gain in PFS would translate into a commensurate gain in OS (with small variation due to the pre-progression death), which might be considered reasonable in some cancers. A further reason why the STM is sometimes argued to be more appropriate than the PSM in modelling the intervention arm is that it is believed to be easier to use external evidence (e.g. using external evidence in second-line to model the pathway for a first-line treatment). However, such evidence sources can also be used indirectly to inform the most appropriate OS fit for the PSM.

When evaluating the performance of methods in Chapter 8, the key assumption was that the analyst possessed some insight/information through either clinical validation or external evidence on the long-term expected OS and PFS for the treatment arm of interest, which is less likely for a new intervention arm that is not the standard of care. A single trial arms approach was also used. In addition, in Chapter 8, performance was only assessed in terms of predicted mean health state sojourn time/QALYs. Whilst this was a pragmatic decision given the nature of the simulation study, as highlighted in Section 8.3.3, such an approach has some limitations as the mean predicted health state sojourn time/QALYs may appear unbiased, but visual predictions may be poor (e.g. the model could over-predict at the beginning and under-predict at the end of the function, but overall generate the same LYGs). In Chapter 8, hypothetical trial arms were generated and methods were assessed in these hypothetical datasets. Whilst data were generated with the intention of ensuring realism, there are inherent limitations with using simulated datasets, as the data-generating process itself may introduce some biases. Perhaps more importantly, when considering an additional arm (treatment arm), additional assumptions are required to characterise the effect of the intervention on both PFS and OS. Whilst possible, generating realistic treatment effects adds challenges.

Consequently, in this chapter, I explore the performance of the PSM and STM in estimating the incremental health state sojourn time and QALYs associated with the introduction of a new technology in a series of real case studies, where the truth is known because data are complete.

9.3 Description of the method

This study is an extension of the simulation study conducted in Chapter 8, and uses real case-studies. Although it uses real data from trial arms, this study is considered exploratory because of previously discussed challenges associated with interpreting incremental results and using the mean time in health state (e.g. incremental may be unbiased, but predictions for each separate arm may be different to the truth) and the large number of choices available with respect to how the treatment effect for the new intervention could be implemented in both the PSM and the STM. Real case studies are used instead of simulated data to avoid making any assumption about the treatment effect. This also therefore mean that predictions are compared against the truth for the sample (e.g one dataset), rather than unbiased truth (which would account for sampling variation).

9.3.1 Aims of the exploratory analysis

The aims of this exploratory analysis are to examine the performance of the PSM and STM when estimating the incremental health state sojourn times and QALYs associated with the introduction of a new technology in real case studies which include:

- a control arm, which represents the current standard of care, whereby the analyst has some information (e.g. clinical opinion and/or external evidence) regarding the expected long-term PFS and OS (similar to the approach used in Chapter 8),
- and a new intervention for which no external information is available on long-term treatment effects or outcomes on PFS and OS beyond the observed period of the trial.

Put simply, this analysis attempts to address the question of whether directly fitting parametric models to OS data when information about the long-term prognosis of an intervention intervention is unknown could be appropriate, despite not explicitly modelling the underlying natural history process.

9.3.2 Clinical trial datasets (case studies) considered in the exploratory analysis

The PSM and STM were applied to real cancer trial datasets to reflect the treatment effects typically observed in cancer trials and also to limit the number of other unsupported assumptions when modelling the underlying natural history.

It should be noted that whilst using real cancer trial datasets reduces the number of assumptions, the number of scenarios examined will be limited by the data used and may not therefore cover all possible practical solutions.

Trial data usually containss some censoring. However, in order to assess the performance of alternative modelling methods, the truth needs to be known and complete data are required. Consequently, for these case-studies, complete datasets containing full information (no censoring) are required alongside a censored/truncated dataset (reflective of trial data used to estimate health state sojourn time in HTA).

Obtaining data from a cancer clinical trial in which both complete and censored information is available is challenging. Ideally, a clinical trial which reported outcomes at the primary analysis and a final analysis (with events observed for all individuals) would be used in this exploratory analysis. However, in the majority of cancer trials identified (through searches of databases), a large number of patients remained censored in the final analysis (i.e. the trials stopped before everyone in the trial died). Cancer trials with complete information (no censoring) are rare. Furthermore, for a number of trials which report both a primary and final analysis (although still not complete), not all outcomes are reported for both analyses. Indeed, the cut-off time for the final analysis for PFS is often different to the final analysis for OS (for instance, PFS may stop being collected when a certain of number of events occurred [enough to estimate the treatment effect], whilst OS data continues to be collected to ensure sufficient data for estimation of the treatment effect.).

The complete data representing the (sample) truth are described in Section 9.3.2.1. Censoring is subsequently introduced into these datasets in Section 9.3.2.2 to create truncated datasets to reflect information typically available in HTA when implementing the different methods of interest.

9.3.2.1 Complete (or near complete) datasets used in this exploratory analysis

The alternative methods assessed in this case study were applied to a number of cancer trials for advanced/recurrent gastric cancer publicly available in the R package *surrosurv* (gastadv dataset) provided by the GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) group. The selected datasets are part of a broader 20 trials in gastric cancer used in a meta-analysis looking at the surrogacy between PFS and OS.[135-137] These trials were selected as data were publicly available. No similar complete data were identified in other trials in other disease area. Data identified in other disease areas typically contained a large amount of censoring, rendering them unsuitable for the case studies.

As the purpose of the meta-analysis[135-137] was to assess the surrogacy between PFS and OS, included trials examined a number of different treatments options, sometimes in different subgroups. Consequently, an intervention in one trial could be the control arm in another. Similarly, comparators could also be different between trials despite examining the treatment effect for the same intervention. Researchers who compiled the dataset defined the experimental (intervention) treatment as the treatment arm in those trials which contained the largest number of drugs irrespective of the active ingredient or regimen used (for instance triplet therapy compared with doublet therapy) or the treatment arm which included newer agent when the number of regimens given was the same (for instance when both arms were doublet therapy). For these reasons, the data were not pooled across trials.

The number of patients enrolled/randomised and the proportion of recorded PFS and OS events, for the control and intervention arms and whole trial population are presented in Table 15 for each of the 20 individual trials included in the meta-analysis.[135-137]

Table 15 : Characteristics of trials included in the gastric cancer dataset

	Combined arm			Control arm			Intervention arm		
Trial number	n	% PFS events	% OS events	n	% PFS events	% OS events	N	% PFS events	% OS events
1	60	0.97	0.95	30	0.97	0.97	30	0.97	0.93
2	62	0.92	0.9	29	0.97	0.93	33	0.88	0.88
3	119	1	0.99	61	1	0.98	58	1	1
4	256	1	1	72	1	1	184	1	1
5	118	0.95	0.85	60	0.95	0.85	58	0.95	0.84
6	58	0.59	0.52	27	0.63	0.48	31	0.55	0.55
7	156	1	1	53	1	1	103	1	1
8	90	0.99	0.94	45	1	0.98	45	0.98	0.91
9	387	0.95	0.9	132	0.95	0.92	255	0.95	0.89
10	206	0.93	0.87	101	0.94	0.9	105	0.92	0.85
11	135	0.98	0.86	45	1	0.91	90	0.97	0.83
12	279	0.97	0.95	105	0.95	0.93	174	0.99	0.97
13	337	0.92	0.88	165	0.94	0.88	172	0.91	0.87
14	148	0.87	0.8	73	0.95	0.89	75	0.8	0.72
15	457	0.86	0.75	230	0.86	0.77	227	0.85	0.74
16	158	0.83	0.8	79	0.82	0.77	79	0.84	0.82
17	86	0.97	0.94	43	1	1	43	0.93	0.88
18	704	0.98	0.94	234	0.99	0.96	470	0.97	0.93
19	120	0.91	0.89	38	1	0.97	82	0.87	0.85
20	133	1	1	46	1	1	87	1	1

Abbeviations: n=number; OS: overall survival; PFS: progression free survival

As shown in Table 15, a number of the available trials have complete (or nearly complete) information on PFS and OS. Ideally, this exploratory analysis would only consider those trials which had full information with no censoring (e.g. Datasets 4, 7 and 20). However, those trials typically had a small sample size (<75 patients) in one of the treatment arms (typically the control group) which could introduce some uncertainty. Furthermore, as previously highlighted, trials included in the meta-analysis (and therefore in this dataset) examined different treatment regimens in gastric cancer, and therefore the observed treatment effect in terms of gain in PFS and OS varies between the different trials included. Conversely, including trials which contain a large amount of censoring is likely to introduce biases and require a larger number of assumptions.

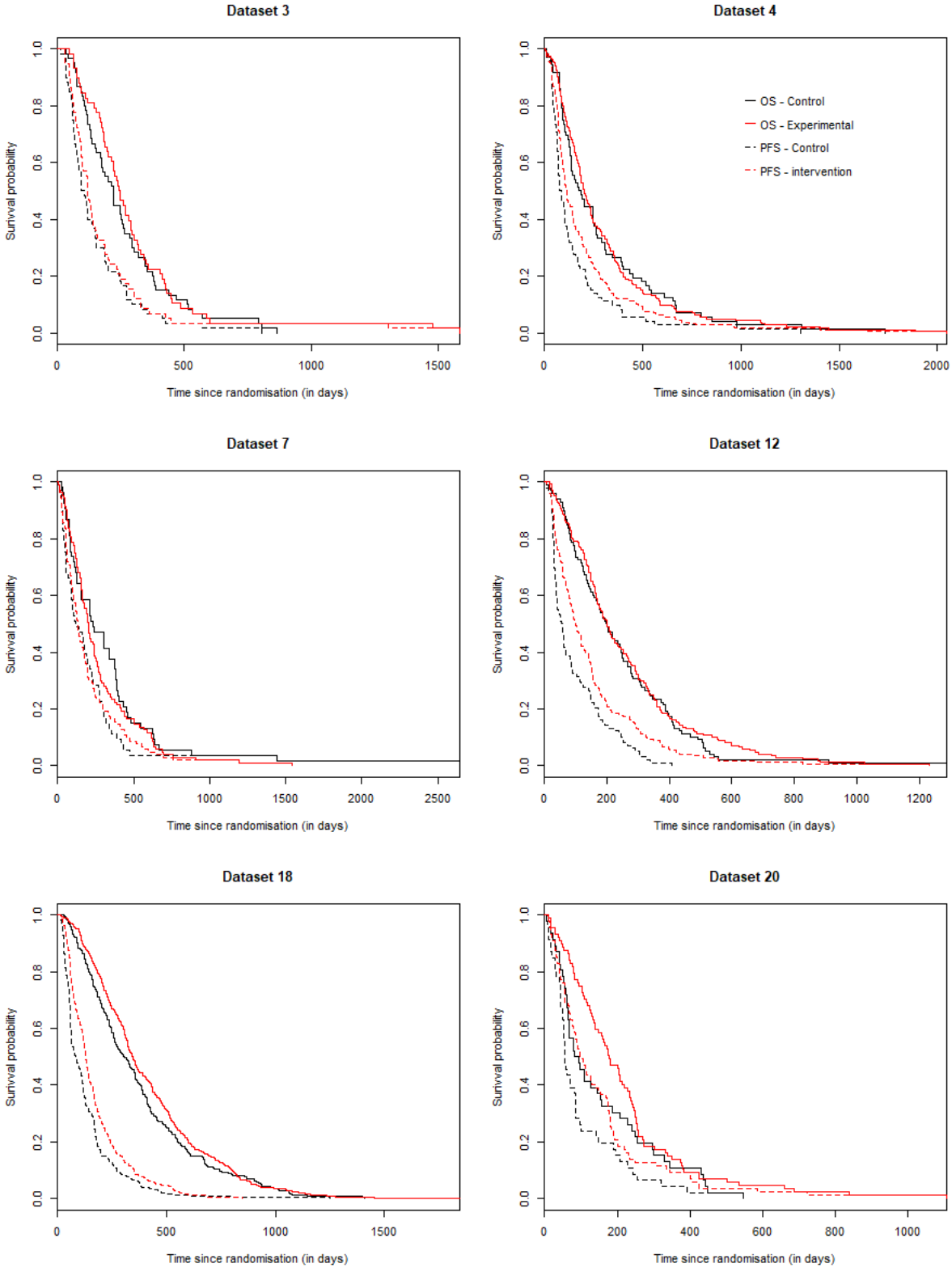
In addition to the three trials which had complete information (e.g. Datasets 4, 7 and 20), the following trials were also included in this exploratory analysis:

- Dataset 3: only one patient was censored for OS for the control arm, and therefore the dataset was almost complete.
- Datasets 12 and 18: the sample size in each arm was >100 patients, >90% of events were observed and these two datasets illustrated different scenarios for the treatment effects on PFS and OS.

Other datasets were not considered further as more than 10% of OS events in at least one arm were censored and the trials did not provide additional scenarios of interest in terms of treatment effects for OS and PFS.

For the three datasets with near complete information (Datasets 3, 12 and 18), assumptions had to be made for the small number of patients that were censored. For simplicity, censored patients were removed from the trial. This is a simplification. It should be noted that an alternative approach could have been to assume the event time to be the same as the censoring time. As the interest is not around quantifying the treatment effect between treatment arms with full accuracy, removing the small number of censored of patients was deemed to be reasonable as this would not affect conclusion from this analysis. Unsurprisingly, the KM estimate following removal of the censored patients was similar to the original KM (not shown), as only a small number of patients were censored. The KM plots for OS and PFS for the included six datasets are summarised in Figure 60.

Figure 60 : PFS and OS in the datasets considered in this exploratory analysis (excludes censored patients where necessary) – Gastric cancer



As shown in Figure 60, different scenarios regarding the treatment effects on OS and PFS for the intervention vs. control arm are covered in these six datasets. These scenarios can be categorised in terms of whether the intervention was associated with a gain in PFS and/or a gain in OS compared with the control arm (Table 16).

Table 16 : Scenario examined in this exploratory analysis

Trial	PFS gain	OS gain
Dataset 3	NO	YES
Dataset 4	YES	NO
Dataset 7	NO	NO
Dataset 12	YES	NO
Dataset 18	YES	YES
Dataset 20	YES	YES

It should be noted that these scenarios represent the truth (once data are complete after the final analysis), rather than the treatment effect observed at the primary analysis when a proportion of patients are censored.

9.3.2.2 Truncation of the datasets

The six gastric cancer trial datasets described in Figure 60 were truncated to reflect information typically available in HTA.

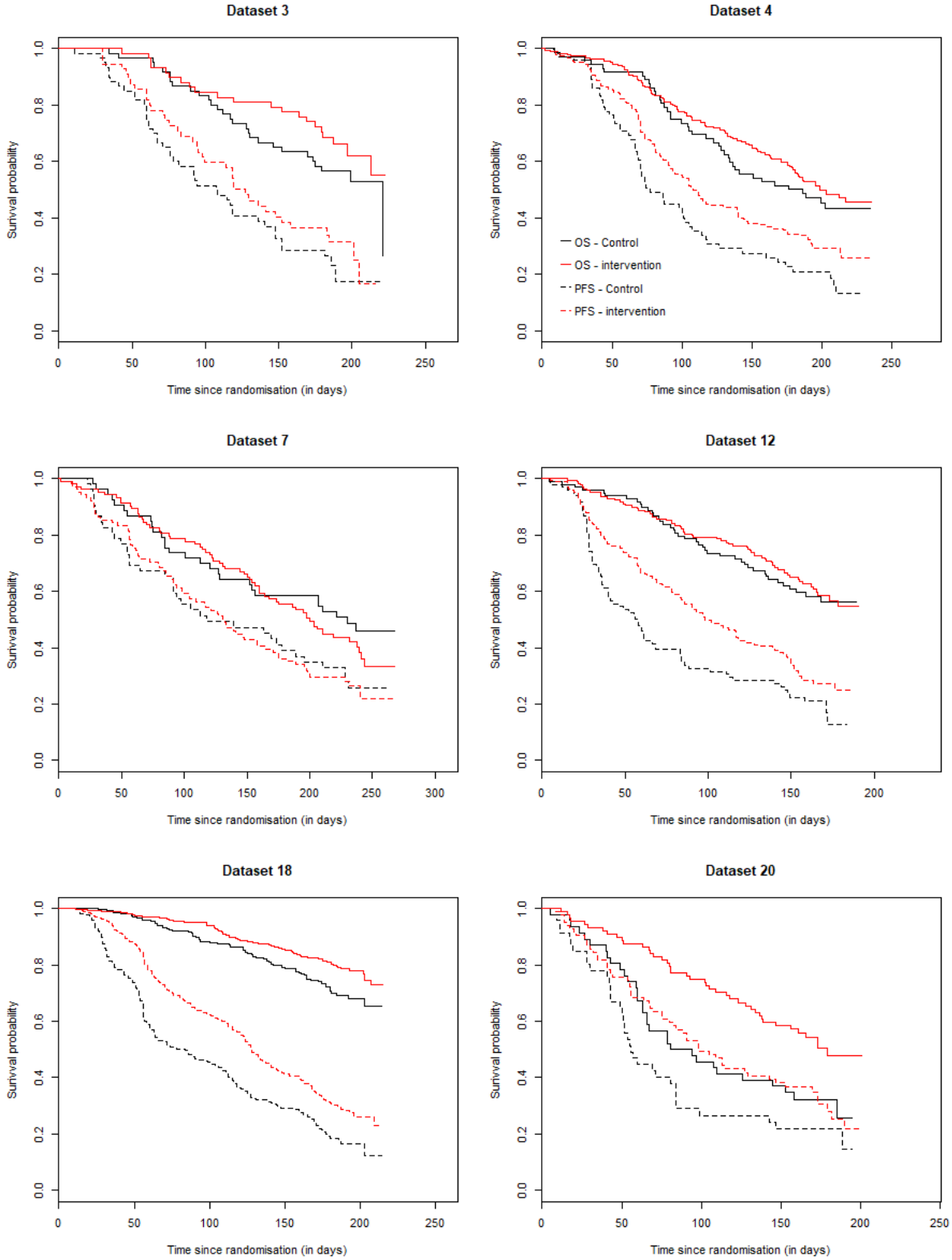
The following steps were followed. The resulting truncated datasets that are used to generate predictions are shown below in Figure 61:

- **Step 1: Assumption that patients do not enter the trial at the same time (random entry).**
  The datasets considered did not contain information on the time to entry into the trial. Only information on PFS and OS was available from the dataset. Whilst this is not a limitation when estimating survival outcomes, ignoring time to entry could affect the censoring mechanism as this would inappropriately assume that all patients enter the trial at the same time.
  To add realism, random entry to the trial was added assuming that patients entered the trial within a 50-day window. For each patient randomised, a time to entry was assigned, randomly sampled from a uniform distribution (between 0 to 50 days).

- **Step 2: Introduction of random censoring associated with loss of follow-up.** Once time to random entry in the trial was assigned for each patient, random censoring was introduced to reflect loss of follow-up for PFS, assuming the time to random censoring (due to loss of follow-up) follows an exponential distribution with an arbitrary rate parameter of 0.000747. As with the simulation study presented in Chapter 8, random censoring was assumed to affect PFS but not OS (as OS is known with certainty). The rate parameter was selected to ensure that approximately 5% to 15% of patients would be lost to follow-up for PFS in the truncated dataset.
- **Introduction of administrative censoring (early termination).** The final step was to introduce the time at which the trial would terminate to report on its primary analysis. In this exploratory analysis, trials were assumed to terminate and report outcomes when 70% of PFS events occurred across both treatment arms. This assumption was generally in line with the design of the actual trials. Consequently, patients with a recorded PFS and OS greater than the administrative censoring time were censored at this time point.

It should be noted, as expected, that the censored datasets were very close to the complete datasets, as censoring is assumed to occur at random, with the exception of administration censoring.

Figure 61: KM for PFS and OS observed in the gastric cancer datasets after truncation – Gastric cancer



9.3.3 Estimands considered

As described in Section 9.3.2, methods (the PSM and STM) were applied to real truncated datasets. Predictions were then compared against the PFS and OS from the complete dataset, representing the truth for the sample, rather than the underlying population truth (which is not affected by sampling variation). This is different to the simulation study described in Chapter 8, whereby the model predictions estimated in a smaller dataset (n=240) were compared against the unbiased truth estimated assuming a large number of patients (240,000 individuals) to reflect variation associated with the smaller sample size of clinical trial. This was necessary given the data available.

It should be noted that the target considered in this exploratory analysis is the truth from the sample, and therefore is prone to sample variation with a wide CI around the estimated treatment effect due to the small sample size".

Furthermore, in this exploratory analysis the sample truth was calculated by fitting a parametric model rather taking the observed average survival time (or restricted mean from the KM). This was done given the small sample size and also because of the step change in the KM which dropped to 0 quickly in some datasets. A flexible model (spline model with 3 knots) was used to estimate the sample truth.

A time horizon of 10 years was examined given the poor prognosis in gastric cancer.

The PSM and the STM methods were applied to the truncated dataset to estimate the following outcomes for each treatment arm individually and incrementally:

- Mean time spent in the PF health state
- Mean time spent in the PD health state
- Total LYGs
- Total QALYs.

Utility values of 0.80 and 0.50 were arbitrarily assumed for individuals in the PF and PD states, respectively, in order to estimate total QALYs. It should be noted that it is difficult to interpret the incremental QALY gain in isolation as it is a function of: (a) the time spent in the health states in each individual arm, and (b) the assumed utility values. Consequently, for each of the six case-studies, the predicted time in each health state, for each individual arms, is described alongside the incremental LYGs and QALYs.

9.3.4 Implementation of methods

Similar to Chapter 8, the process for curve selection was automated in order to: (i) have a structured transparent and reproducible process for curve selection and (ii) minimise any influence from

unconscious biases. However, predictions associated with each methods were double-checked manually to ensure face validity (which was not possible in Chapter 8).

A daily cycle length was used to increase accuracy in the predictions. For the sake of simplicity, general population mortality was not used to cap modelled OS rates in this exploratory analysis. This is because the age of patients included in the trial was not available; hence, additional assumptions would have been required. This was also considered a pragmatic decision given the poor prognosis of patients with advanced/gastric cancer in the trial considered with the large majority of patients recruited dying within 3 years, with no patients alive after 7 years (Figure 60). Discounting was also not considered. This simplification was made to make the interpretation of results easier. These simplifications are unlikely to affect the findings of the analysis.

9.3.5 Methods considered in this exploratory analysis

Given the nature of this analysis and for ease of interpretation, only the PSM and STM (unadjusted and adjusted in the log scale) were considered in this exploratory analysis. The more formal MSM was not explored given its similarities with the STM (see Chapter 5 and Chapter 8). Furthermore, the Li model[91] was not explored further given that the model is constrained by the need to use Weibull distributions, which is very restrictive when extrapolating beyond the observed period of the trials. Finally, the Fu model (joint modelling of PFS and OS under semi-competing risks) was not considered here for pragmatic reasons, given the aim of this analysis and its exploratory nature. Whilst it would have been possible to include this model, significant alteration would have been necessary as the model is fitted individually to data from single trial arms. As described later, there are already a number of challenges when looking at the performance of the PSM (independent model) when estimating health state sojourn time/QALYs between two treatment arms given the different possible assumption regarding the nature of the treatment effects. It was considered that including the Fu model[89] would make the interpretation of the exploratory analysis more challenging.

A key challenge in this exploratory analysis is that for the PSM and the STM, the estimation of the health state sojourn times and QALYs for the intervention arm could adopt a variety of alternative assumptions regarding how the treatment effect is modelled. Whilst this is also the case for the control arm, where different assumptions could be made, these assumptions and decisions are perhaps more limited and less subject to debate.

For PFS, for both the PSM and STM, a survival curve can be fitted directly to the KM for each arm separately or using a relative treatment effect parameter (in the form of an HR for PH models or a

constant acceleration factor for AFT models). It should be noted that the estimation of PFS would only affect the estimate of OS for the STM.

For OS, within the PSM, whilst a survival distribution is typically fitted directly to the survival data for OS, the effect of the intervention arm could be modelled by;

1. Fitting a separate curve to the intervention arm (either assuming the same distribution as the control arm or using different parametric model forms)
2. Applying a treatment effect (an HR for PH or a constant acceleration factor for AFT models) to the baseline OS and PFS curves from the control arm. This treatment effect may be applied indefinitely or may be assumed to wane at the end of the observed follow-up period or at some other timepoint (e.g. based on clinical opinion).

Within the STM, the effect of the intervention on PPS can be estimated by;

1. Fitting PPS separately to each arm, and therefore assuming that the time to death following progression differs between arms.
2. Applying a treatment effect (an HR for PH or a constant acceleration factor for AFT models) to the baseline PPS from the control arm. Treatment effects may be applied over the patient's remaining lifetime, up to the end of the observed follow-up period of the trial or some other defined time-point.
3. Pooling PPS data from both the control and intervention arm, and therefore assuming that the time to death following progression is the same between arms. A shift in PFS is therefore assumed to lead to a commensurate shift in OS (with some variation attributable to the proportion of patients who may die prior progression).

Further variations are also possible, for example using the observed KM up to a certain timepoint followed by parametric extrapolation.

It is not practical to examine all possible implementations of the STM and PSM and the choice of implementation is often guided by the analyst's judgement. Consequently, for this exploratory analysis only a limited number of alternative implementations were considered. These were considered by myself to be the most common and appropriate assumptions made when reviewing previous TAs. Key assumptions considered in this exploratory analysis on how PFS and OS is modelled for the intervention arm for the PSM and STM are summarised in Table 17. Lifetime treatment effects are often deemed clinically implausible, and assuming the relative treatment effects apply over a shorter duration (such as 3 to 5 years depending on the condition and length of follow-up) is often considered to be more

plausible. To reflect this, in addition to applying a treatment effect over the lifetime, analyses are presented where the treatment effect is limited to the trial duration given the poor prognosis of patients with gastric cancer (most patients die and progress within 3 years).

Analyses are separated into primary analyses and scenario analyses depending on the most likely implementations in HTA (according to my judgement). In this exploratory analysis, irrespective of the analysis, PFS was always modelled for the intervention arm assuming a treatment effect (an HR for PH or a constant acceleration factor for AFT models). This was a pragmatic decision. This was found to be most commonly assumed approach in health economics for PFS.[9] However, it is recognised that different analysts may have different views and could be reluctant to use a treatment effect for any outcomes, and therefore prefer to model PFS using separate parametric distributions.

Table 17 : Method used to model outcomes for the treatment arm.

		PFS	OS
PSM	Primary analysis	HR applied to control	HR applied to control (whole model duration)
	Scenario analysis 1	HR applied to control	HR applied to control (trial duration only). The same hazard used after
	Scenario analysis 2	HR applied to control	Separate fit (same distributional form as control)
STM (unadjusted)	Primary analysis	HR applied to control	Pooled PPS if $p<0.05$ Separate PPS if $p>0.05$
	Scenario analysis 1	HR applied to control	Pooled PPS if $p>0.05$ Separate PPS if $p<0.05$
	Scenario analysis 2	HR applied to control	Separate PPS (same distributional form as control)
STM (adjusted using	Primary analysis	HR applied to control	Pooled PPS if $p<0.05$ Separate PPS if $p>0.05$
	Scenario analysis 1	HR applied to control	Pooled PPS if $p>0.05$ Separate PPS if $p<0.05$
	Scenario analysis 2	HR applied to control	Separate PPS (same distributional form as control)

Abbreviations : HR: hazard ratio; OS: overall-survival; PFS: progression-free survival; PPS: post-progression survival; STM: state-transition model; TTP: time to progression

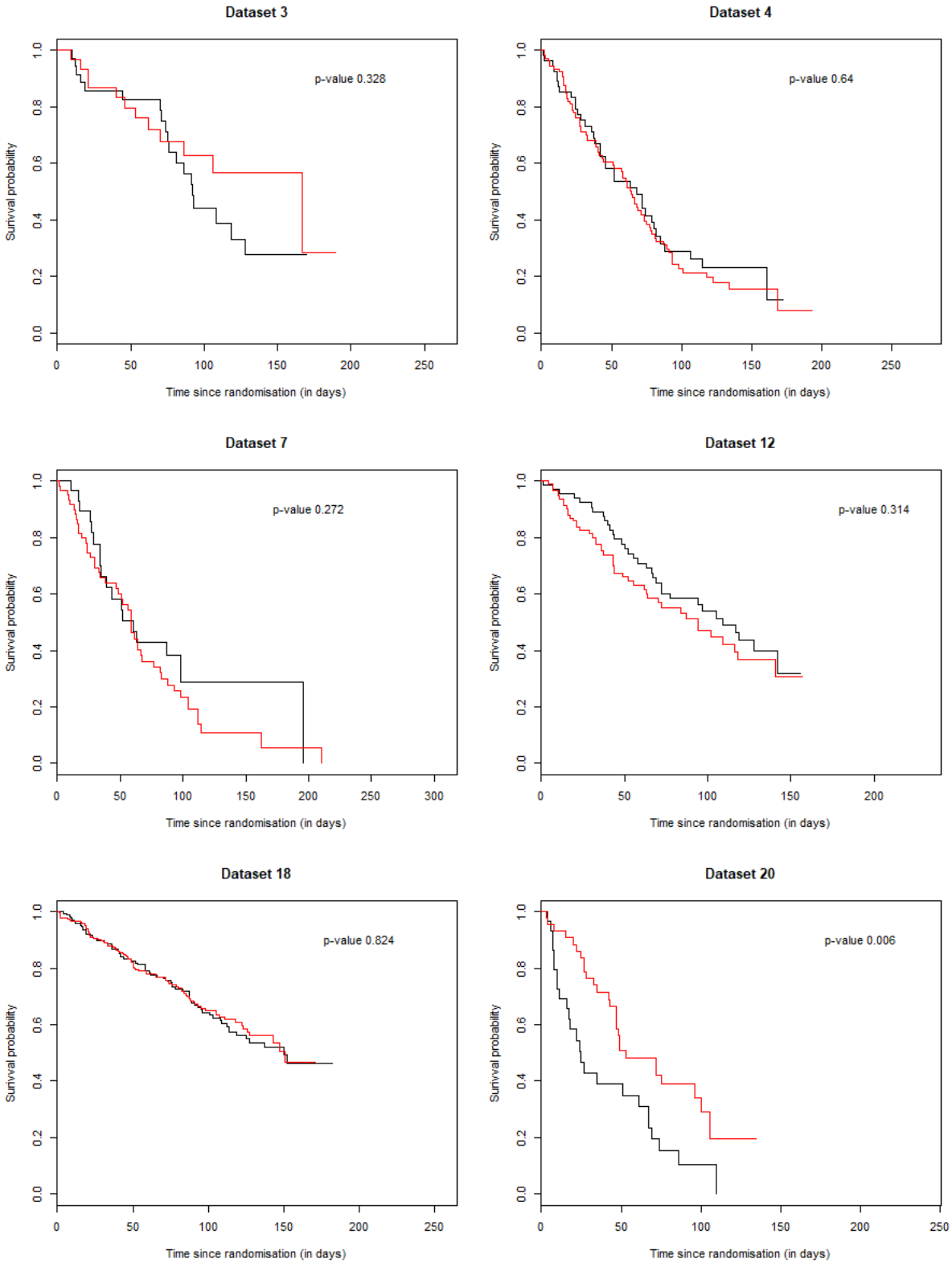
For the primary analysis, for the STM, PPS was pooled if there was no statistically significant difference for the HR for PPS between treatments arms, as this assumption is commonly made in health economics to justify pooling PPS (See Chapter 5.3.5.5). PPS for each treatment arm, and associated p-value (HR for PPS between treatment arms) is shown in Figure 62. It can be seen that PPS was visually similar and there was no statistically significant difference in the HR for PPS in Datasets 3, 4, 7 or 18. There were some visual differences in PPS between arms in Dataset 12, but the *p*-value for the HR was >0.05, and therefore this could be attributable to the sample size. PPS was visually and statistically different in Dataset 20; thus different PPS was used between arms in the primary analysis. It should be noted that using a p value to decide on whether or not to pool PPS has limitation, as absence of evidence does not mean evidence of absence. Furthermore, the trial isn't powered to find a significant difference in PPS.

9.3.6 Curve selection process

As highlighted in Chapter 8.3.4.1, the choice of parametric model for extrapolation is a key issue when judging the performance of a method. Consequently, a similar process for curve selection to that described in Chapter 8.3.4.1 was used. Analysts are assumed to possess some (imperfect) information on the long-term PFS and OS for the control arm only (but not for the intervention arm). This is because the control arm is typically part of clinical practice already, and therefore, external evidence on its long term outcomes is typically available and clinicians are able to have an estimate of its long-term effect.

It should be noted that it is possible that no information is available for both the control and intervention if neither are part of standard of care and are new interventions. However, this situation is not considered here.

Figure 62 : KM for PPS in the truncated datasets and associated p-value – Gastric cancer



In Chapter 8, analysts were assumed to be given information on when 90% of events in the trial arm were expected to have occurred (in the primary analysis).

As highlighted in Chapter 8, results are likely to be impacted if different criteria for curve selection are used. Whilst this is true for both approaches, in the STM, OS is a function of both the predictions of PFS and PPS, and thus, any misspecification of PFS will also impact on OS, as shown in Chapter 8.

For ease of interpretation and transparency, two analyses are conducted assuming:

1. A stringent criterion for curve selection (i.e. more certain) whereby the analyst is assumed to have elicited information regarding the time at which 95% of PFS and OS events in the control arm would occur, with a margin of error of 2.5%. This is considered stringent as this is allow to better represent a possible tail in the data.
2. A less stringent criterion for curve selection (less certain, similar to that used in Chapter 8 in the primary analysis), whereby the analyst is assumed to have elicited information regarding the time at which 90% of PFS and OS events in the control arm would occur, with a margin of error of 5%.

For each of the six trials, the complete datasets are used to inform the clinical plausibility for the fit to the control arm only, in order to select the transitions/extrapolation that are likely to be plausible. Analysts are assumed to possess no information on the OS or PFS for the intervention arm, and therefore, must rely on either extrapolating OS directly from the truncated trial data in the PSM (using a treatment effect or using the same distributional form), or modelling the underlying process in the STM using information available for the control arm only.

9.3.7 Performance measure

Bias was the only performance measure considered in this exploratory analysis. As previously stated, only bias against the sample truth (point estimate) was considered given the wide 95% CI due to small sample size.

As highlighted in Chapter 8, using the mean time in health state has limitations given that it is possible for the mean time to be unbiased, even if the model does not provide a good representation of the data. As this exploratory analysis is limited to six case studies, it was possible to visually examine predictions for each dataset; this was not possible in the simulation study conducted in Chapter 8. Consequently, a description of the visual predictions on OS and PFS generated by each method is presented alongside the mean predicted health state sojourn time/QALYs (for each arm individually and incremental results between treatment arms).

9.4 Results

Given the wealth of information, only key findings are summarised and the description of results in this section focuses on the primary analysis using the stringent criterion for curve selection (e.g whereby the analyst is assumed to have elicited information regarding the time at which 95% of PFS and OS events in the control arm would occur, with a margin of error of 2.5%), unless stated. Results from scenarios assuming different assumptions regarding the modelling of the treatment effects (waning treatment effect, no use of HR, separate PPS, etc) and different curve selection criteria (less stringent definition) are presented in Appendix 14; and referenced in this thesis when necessary. The timepoint used for the curve selection process, and *p*-values for PPS are also presented separately for each dataset alongside the selected distributions (based on visual fit, AIC and long-term plausibility) for OS, PFS, PrePS and PPS for the two scenarios examined (stringent and less stringent curve selection criteria) in Appendix 15.

In Section 9.4.1, I report how methods performed in each dataset in estimating the mean incremental LYGs and QALYs and provide a brief description (key messages). Due to challenges in interpreting the mean time in health states, as well as incremental values (as the incremental may be unbiased, but the estimate for each individual arm biased), the fit (visual) is discussed in Section 9.4.2 for relevant datasets to provide further details and interpretation only when methods predicted appropriate incremental LY/QALYs.

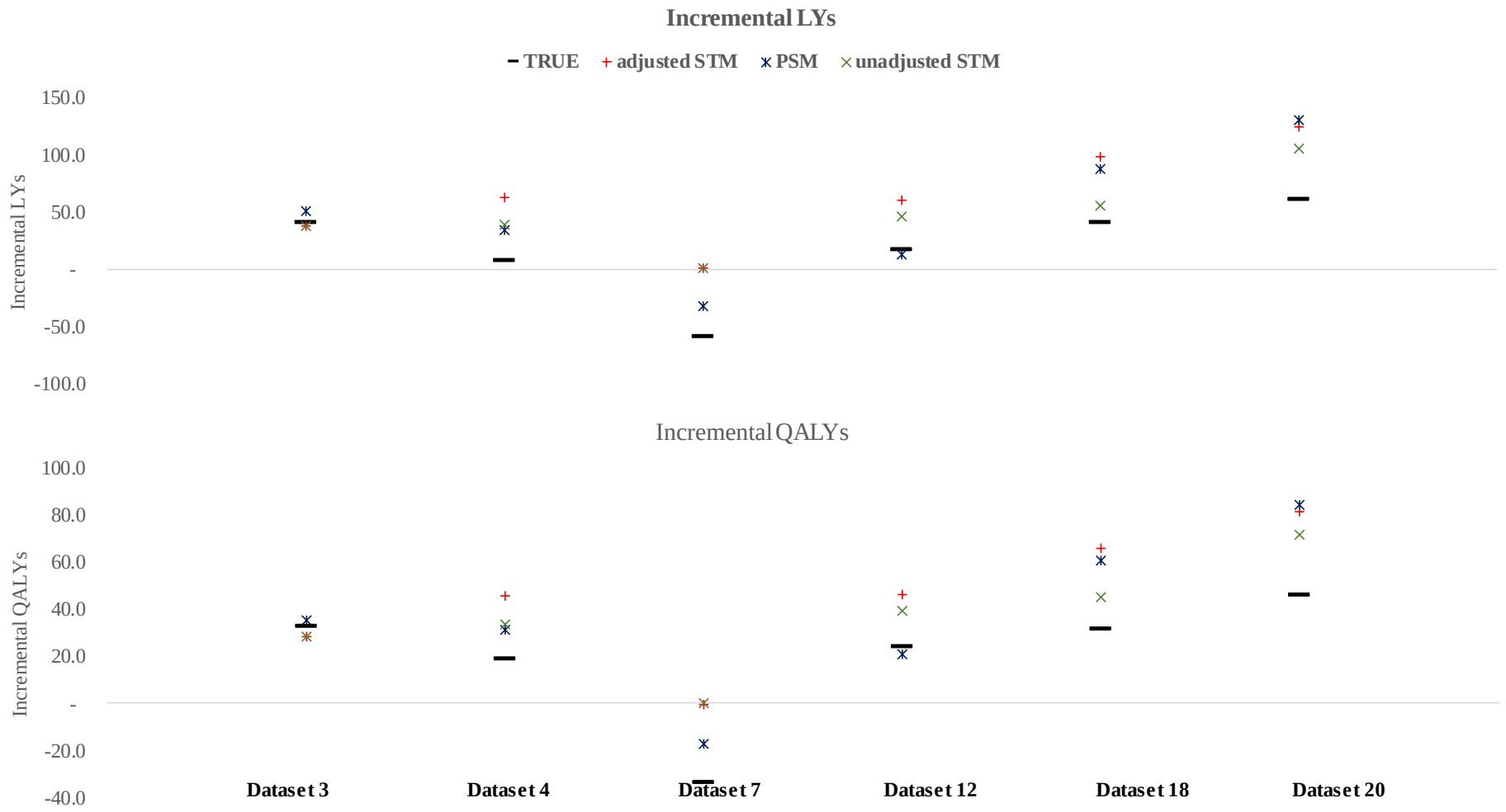
9.4.1 Performance of the PSM and STM in estimating the incremental mean LY and QALYs

The predicted incremental mean LYGs and QALYs generated using the PSM and STM (under the base-case assumption about the modelling of the treatment effect and curve selection criteria) are displayed in Figure 63.

In summary, no approach was consistently better at predicting the incremental LYGs and QALYs. In fact, approaches were prone to biases and generated varying incremental mean LYGs and QALYs. These findings were generally not affected when considering different assumptions to model the treatment effect; unless stated:

- both the PSM and STM generated reasonable incremental LY/QALYs in Dataset 3,
- no approaches generated reasonable incremental LY/QALYs in Dataset 4 and 20,
- the PSM generated more plausible estimates for the incremental LY/QALYs in Dataset 7, but remained inaccurate (Table 18).

Figure 63: Comparison of the predicted mean incremental LYs and QALYs gained (base-case) against the truth from the sample dataset



- the PSM generated more plausible estimates for the incremental LY/QALYs in Dataset 12, unlike both versions of the STM,
- the unadjusted STM generated more plausible estimate for the incremental LY/QALYs in Dataset 18, but remained inaccurate (Table 18).
- No approach were appropriate in Dataset 20 under the base-case selection criteria. However, when pooling PPS (despite significant difference; p value <0.05), the unadjusted STM generated reasonable estimates for the incremental LY/QALY in Dataset 20.
- As expected findings were similar between LYs and QALYs.
- Approaches remained inaccurate when a less stringent curve selection criteria was used for the base-case assumption for the modelling of the treatment effect (Table 19).

Table 18 : Predicted incremental LY and percent bias (using stringent curve selection criteria – base-case assumption regarding the modelling of the treatment effect)

	Truth for the Sample	Predicted Inc LY				Percent Bias		
		PSM	Unadjusted STM	Adjusted STM		PSM	Unadjusted STM	Adjusted STM
Dataset 3	**40.3**	51.5	37.5	38.0		27.68%	-6.93%	-5.77%
Dataset 4	**7.5**	34.5	39.2	62.5		362.30%	425.42%	737.59%
Dataset 7	**-59.8**	-32.6	1.5	1.4		-45.52%	-102.44%	-102.28%
Dataset 12	**16.1**	13.3	46.7	60.5		-17.52%	189.32%	275.10%
Dataset 18	**39.9**	87.4	56.0	98.2		119.03%	40.27%	146.02%
Dataset 20	**60.3**	130.5	105.4	124.8		116.48%	74.85%	107.08%

Table 19 : Predicted incremental LY and percent bias (using less stringent curve selection criteria – base-case assumption regarding the modelling of the treatment effect)

	Truth for the Sample	Predicted Inc LY				Percent Bias		
		PSM	Unadjusted STM	Adjusted STM		PSM	Unadjusted STM	Adjusted STM
Dataset 3	**40.3**	66.0	37.5	38.0		63.83%	-6.93%	-5.77%
Dataset 4	**7.5**	34.5	39.2	62.5		362.30%	425.42%	737.59%
Dataset 7	**-59.8**	-32.6	-1.5	1.4		-45.52%	-102.44%	-102.28%
Dataset 12	**16.1**	15.4	46.7	60.5		-4.77%	189.32%	275.10%
Dataset 18	**39.9**	87.4	56.0	112.1		119.03%	40.35%	180.98%
Dataset 20	**60.3**	148.8	105.4	139.2		146.87%	74.85%	130.91%

9.4.2 Visual predictions for each individual dataset, in each treatment arms

The description of results in this Section will focus on those scenarios where at least one approach was considered to be reasonable (<20% variation from the truth) at estimating the incremental mean LYs/QALYs (e.g Dataset 3 and, Dataset 12) under the base-case selection criteria. While the incremental LYs/QALYs predicted was considered reasonable, the aim in this section is to assess whether the estimate and visual predictions for each individual arm were accurate.

No detailed description is provided for Dataset 4, 7, 18 and 20 as none of the approaches generated reasonable incremental LYs/QALYs under the base-case criteria; suggesting that no approach accurately predicted OS for either one or both the control and intervention arm. It should be noted that the unadjusted STM generated reasonable mean incremental LY in Dataset 20 when PPS was pooled (despite p-value <0.05).

9.4.2.1 Comparison of predictions using the PSM and STM in Dataset 3

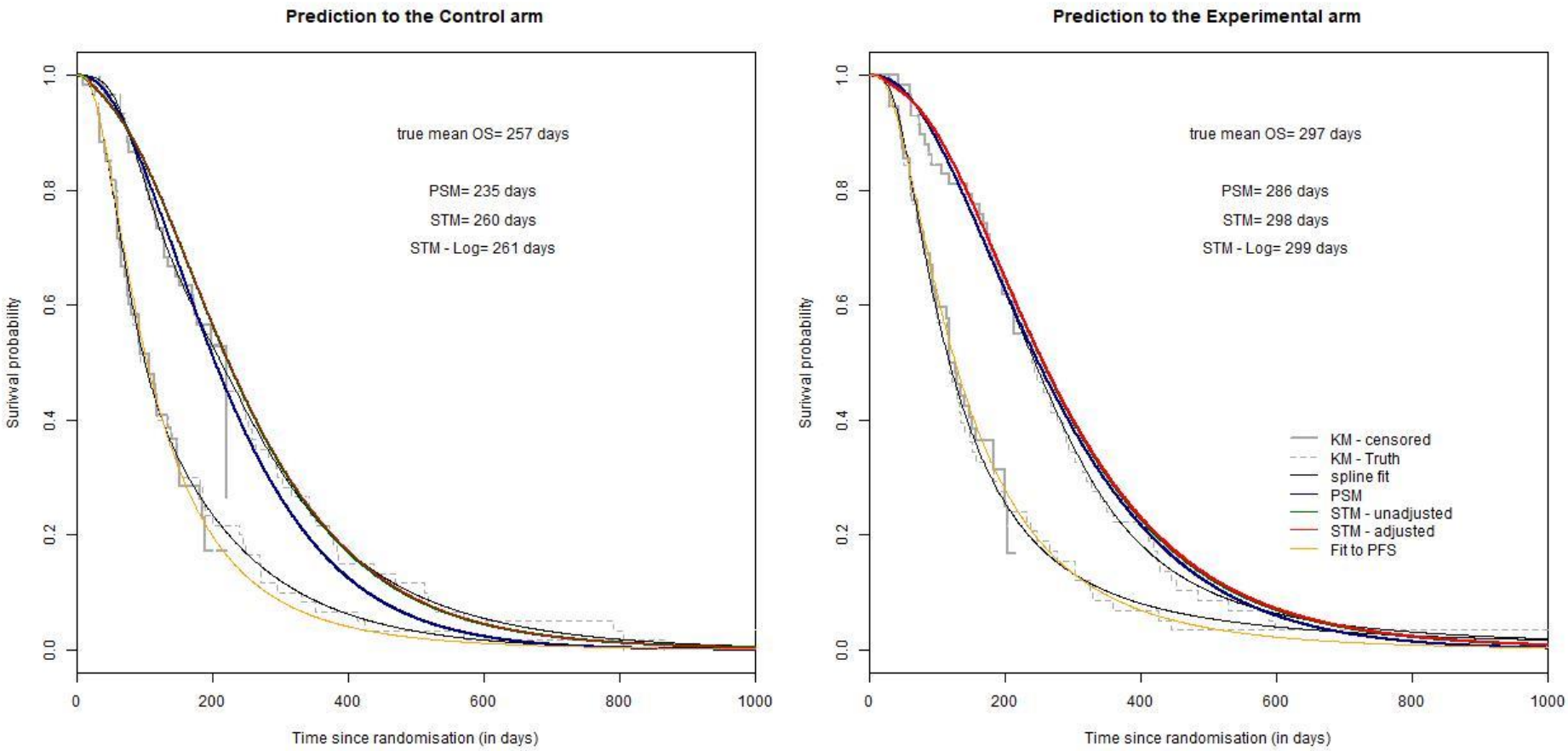
In this dataset, patients enrolled in the intervention arm experienced a small improvement in PFS compared with the control arm. OS was also improved in the intervention arm, compared with the control arm. Using a less stringent curve selection criterion led to the selection of a different OS curve for the PSM compared with the more stringent curve selection criteria.

The predicted PFS assuming a lifetime treatment effect provided a good visual fit to the observed data, and a reasonable visual prediction compared with the unobserved period of the full dataset for both the control and intervention arms. Despite the good visual fit, the mean predicted time in the PF health state (Appendix 14) was slightly under-estimated for both treatment arms, as the selected PFS distribution had a smaller tail compared with that observed in the complete dataset.

When considering predictions for OS, using the stringent curve selection criterion, the PSM fitted the data well during the observed period for the control arm, but the visual fit beyond the observed period of the trial was under-estimated compared with the complete dataset (Figure 64). In contrast, the STMs (when PPS was unadjusted or adjusted using TTP on the log scale) did not provide a good visual fit during the observed period of the trial, but the prediction beyond the trial period was closer to what was observed in the complete dataset (Figure 64). The predicted OS for the intervention arm was broadly similar between approaches and was in line with the OS seen in the complete dataset (Figure 64).

The STMs (adjusted or not) predicted mean LYGs and QALYs which were close to those estimated in the complete dataset, when considering each treatment arm individually.The PSM slightly underestimated the mean LYGs and QALYs for both treatment arms; as a result, the incremental mean LYGs/QALYs predicted by the STMs and PSM were close to those estimated in the complete dataset, with predictions from the STMs being much closer to the truth in that sample.

Figure 64 : Predicted PFS and OS using the PSM and STM in Dataset 3 using a stringent curve selection criterion



Using a less stringent curve selection criterion, the curve selection process did not change compared with Scenario 1 for the STM, as PFS and PPS remained unchanged. However, a different OS curve (to the one used using the stringent definition) was selected which affected the PSM predictions. Predictions were worse, with the predicted OS being over-estimated compared with the OS seen in the complete dataset. Alternative assumptions on the modelling for OS for the PSM (i.e. waning treatment effects and fitting separate models to each arm) led to a slight improvement in predictions, but the PSM remained more biased compared with the STM.

9.4.2.2 Predictions using the PSM and STM in Dataset 12

In this case study, patients enrolled in the intervention group had a significantly better PFS compared with the control group, but no OS gain was observed. Using a less stringent curve selection criterion, led to a different OS curve selection for the PSM. The curves selected for PFS and PPS were the same in both scenarios.

The predicted PFS assuming a constant treatment effect was reasonable (Figure 65) for both the control and intervention arm. As a result, the predicted mean time in the progression-free health state (Appendix 14) was similar, but slightly over-estimated, compared with the complete dataset.

Using the stringent curve selection criterion (Figure 65), the visual fit for OS to the observed data was broadly similar between approaches. Beyond the observed period, no methods generated visual predictions that were close to that observed in the complete dataset for the control arm. The PSM and unadjusted STM under-estimated OS whilst the adjusted STM over-estimated OS. When looking at the intervention arm, the unadjusted STM generated predictions for OS which were close to that observed in the complete dataset, both visually and in terms of mean predicted LYGs (Appendix 14). The PSM provided a good visual fit during the observed period, but under-estimated both the mean LYGs and QALYs. In contrast, the adjusted STM over-estimated both the mean LYGs and QALYs.

Figure 65 : Predicted PFS and OS using the PSM and STM in Dataset 12 using a stringent curve selection criterion

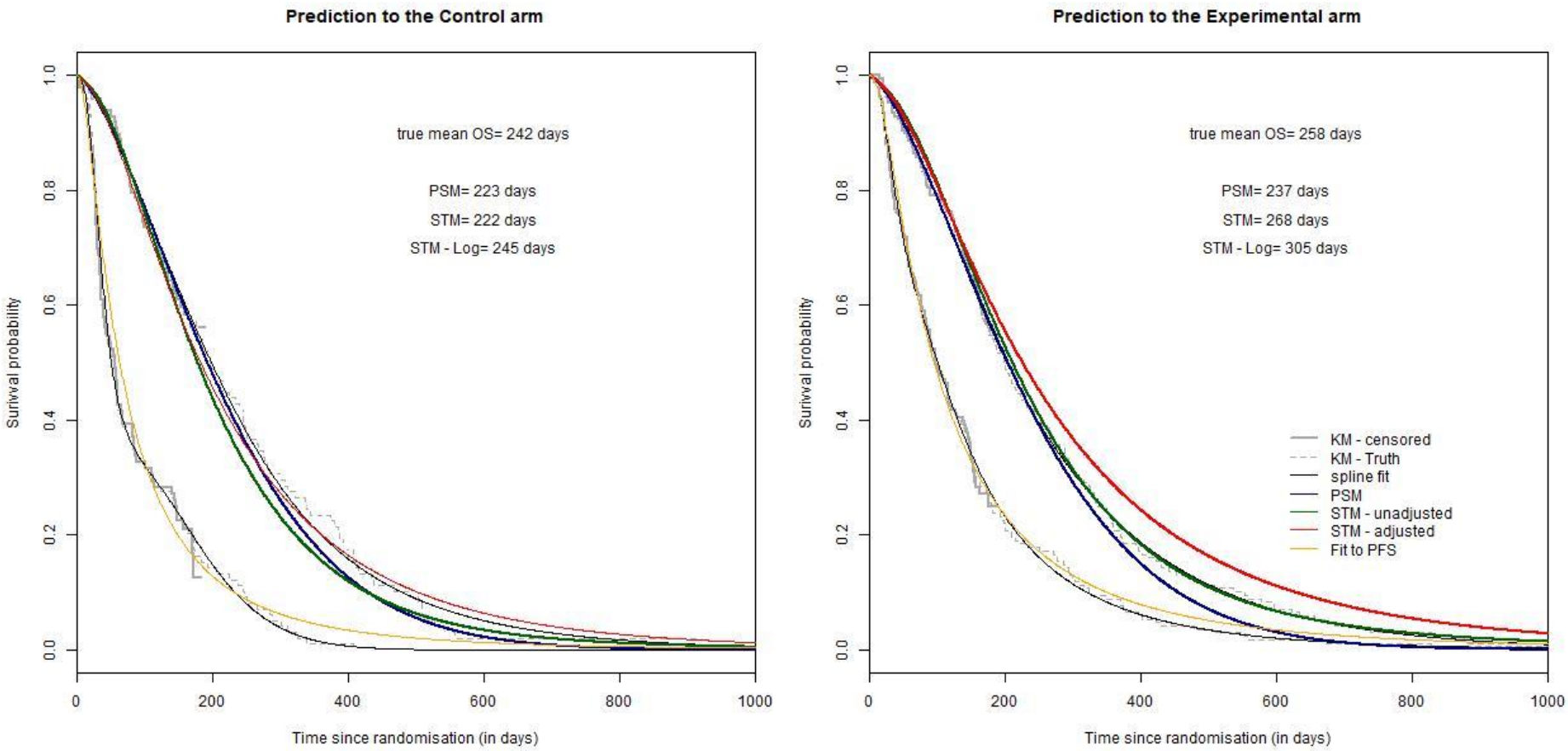
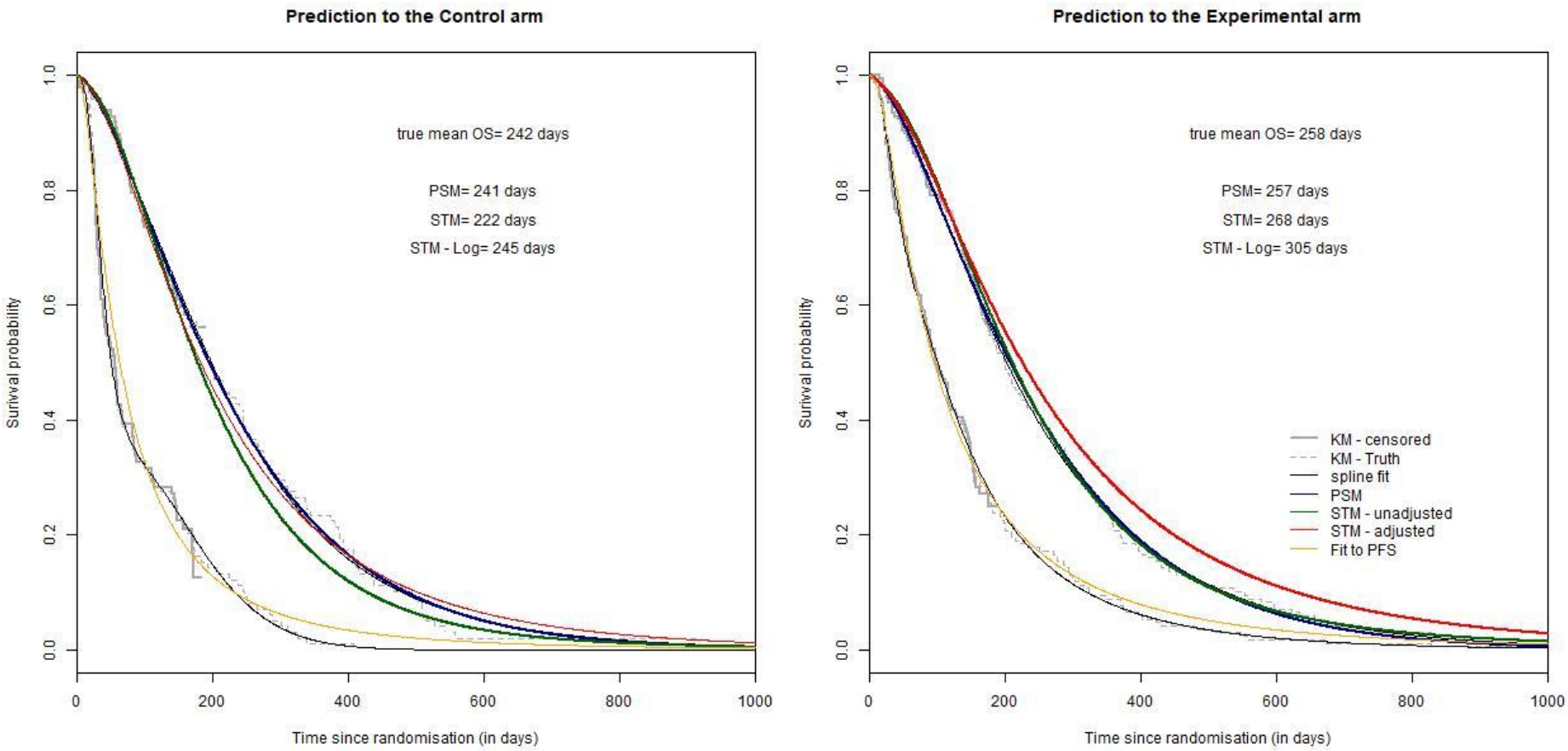


Figure 66 : Predicted PFS and OS using the PSM and STM in Dataset 12 using a less stringent curve selection criterion



As shown in Figure 63, the PSM predicted incremental LYGs/QALYs in the same order to that estimated in the complete dataset. However, when looking at predictions for each individual arm, predictions were not as accurate. The incremental LYGs/QALYs were over-estimated using the STM approach.

In summary, in this particular case study (Dataset 12), despite the PSM generating accurate incremental LYs/QALYs, no method was accurate at predicting OS for both the control and intervention arm separately. Using the stringent curve selection criteria (base-case) methods generated accurate predictions for one arm only..

Using a less stringent curve selection criterion (Figure 66), the curve selection process did not change for transitions included in the STM, and therefore predictions remained the same. A different curve was selected for OS for the PSM which improved prediction in each individual arms, and lead to accurate prediction of the incremental LY (Appendix 14).

9.5 Discussion and conclusions

This chapter examined the comparative performance of PSM and STM in estimating health state sojourn time/QALYs when information about the long-term prognosis of an intervention is unknown. Methods were compared in six case-studies. This showed that despite not modelling the underlying natural history process explicitly, in the case-studies considered the PSM was not less accurate compared with the STM in predicting OS and QALYs for the intervention arm, even under the simplified assumption of the treatment effect applied throughout the model duration.

In fact, predictions were often inaccurate. This study highlights that neither the PSM or STMs are bias-free and that the choice between the two is heavily reliant on information available about the prognosis for both the control and intervention arms. Indeed, whilst both approaches are generally appropriate, any misspecification in the curve selection affects the accuracy of predictions of health state sojourn time/QALYs.

When modelling the intervention arm, there are limitations with both approaches. With the PSM, assumptions have to be made on how long the treatment effect will persist. Assuming the treatment effect persists over the patient's remaining lifetime, as assumed within the primary analysis in this study, is likely to be optimistic. Assumptions could be made that the treatment effect wanes after a certain time point. Whilst this may be a more realistic assumption, it is often arbitrary. For STMs, PPS is often pooled when no statistical difference is observed between arms, and therefore the time to death following progression between arms is assumed to be same (as was done in the primary analysis in this study). Therefore, the benefits of the intervention are driven primarily from the gain in PFS. Such an approach is attractive as it allows borrowing of information from the control arm which is often more complete. However, assuming the same PPS between treatment arms may be an oversimplification. Indeed, despite no significant differences observed in the censored dataset, this may not necessarily mean that there is no difference in PPS had the data been complete. Furthermore, as shown in Chapter 7, using PPS is subject to potential limitations relating to selection bias and informative censoring. Finally, estimated OS with the STM depends on both the extrapolation of PFS and PPS, which could introduce inaccuracy.

A key strength of this study is that real trial cancer datasets were used, and therefore they reflect the prognoses and treatment effects observed in actual oncology clinical trials. Furthermore, datasets with different treatment effects (on PFS, OS or both) were examined allowing exploration of the performance of these alternative modelling methods in a number of scenarios. Different implementations for the PSM and STMs were also examined to estimate outcomes for the treatment arm to cover the most

common scenarios considered in applied health economic evaluations. The process for curve selection was also automated to reduce any unconscious biases when selecting curves for the transitions/endpoints of interest, and two selection criteria were explored.

This study is subject to several limitations.

- the target considered in this exploratory analysis is the truth from the sample, and therefore is prone to sample variation with a wide CI around the estimated treatment effect due to the small sample size.
- in this exploratory analysis the sample truth was calculated by fitting a parametric model (spline model with 3 knots). This was done given the small sample size and also because of the step change in the KM which dropped to 0 quickly in some datasets.
- furthermore, in this exploratory analysis the sample truth was calculated by fitting a parametric model rather estimating the mean time from the KM (restricted mean). This was done given the small size, but also because of the step change in the KM which dropped to 0 quickly in some datasets.
- a limited number of trials were examined. Only six trials were considered, all of which were conducted in gastric cancer since data was available. Nevertheless, the treatment effect on PFS and OS, as well as the definition of the control and intervention arms in the trials included differed between the trials. Therefore, the trials included cover different scenarios.
- the majority of trials had a small sample size which could increase the uncertainty. Three trials had less than 103 patients in each arm. The sample size was also considerably lower for the control arm (half of the size of the intervention in five out of six trials). Despite their small sample size, conclusion from this study are generalisable to larger datasets in that predictions are often inaccurate and that neither the PSM or STMs are bias-free.
- methods had to be applied to datasets containing censoring to reflect information typically available in HTA, but the respective complete datasets had to be available in order to assess their performance. In this study, I used complete datasets, and introduced censoring (random and administration) using assumptions. Whilst results are influenced to some degree by these censoring assumptions, the key findings from this study are not.
- only the most common implementations of the PSM and STM were considered in this study when estimating outcomes for the intervention arms. This is because the implementation could be different between analysts, but also the decision problem at hand. Therefore, what was considered as primary analysis in this study, may not be considered the approach taken by other analysts. For the PSM, the primary analysis considered a treatment effect applied for the

lifetime in order to model the intervention arm. This is a simplification and possibly an optimistic assumption. In practice, analysts may restrict the treatment effect to a particular duration (waning treatment effect). For the STM, I assumed that PPS is pooled if the HR for PPS between arms was not statistically significant, despite occasional visual differences. This approach was employed for the primary analysis as it is often taken in HTA. However, the absence of statistical difference does not preclude any difference, inparticular given the small number of patients considered in some of the arms in this study. However, alongside the primary analysis, whilst described minimally in the main body of this thesis, a number of alternative scenario analyses were presented for transparency using different structural assumptions, such as waning treatment effects or use of separate curves between treatment arms for PPS. These alternative implementations did not affect the key findings.

- a key limitation concerns the choice of model for extrapolation. As previously described, the performance of methods is heavily dependent on the curve selection for the transitions/outcomes of interest. In order to limit unconscious biases, the process used in the simulation study was also used here, whereby curves are selected using an automated process. A key criterion for curve selection is about the long-term plausibility for PFS and OS, and therefore the performance of each method varies depending on the level of certainty of the long-term prediction. Two scenarios are presented in this study, assuming a stringent and a less stringent curve selection criterion. It can be seen that results were sometimes different between analyses, highlighting the importance for clinical validation on the long-term outcomes.
- it should be acknowledged that only the PSM and STM (adjusted and unadjusted) was examined in this study. Alternative methods considered in the simulation study could have been examined. However, they were not considered in this study for pragmatic reasons.
- Information about the control arm was assumed to be available. This is because the control arm is typically part of clinical practice already, and therefore, external evidence on its long-term impact is often available and clinicians may be able to estimate this. However, it is possible that no information is available for either the control or intervention if neither are part of standard of care.
- As highlighted in this study, interpreting QALYs or incremental LYGs and QALYs is very challenging as these are the results of a number of estimates. Indeed, in this study, despite poor visual prediction compared with the complete datasets, approaches sometimes generated similar mean health state sojourn time/QALYs. Similarly, they were a number of cases where predictions for both arms were sub-optimal when taken on their own, but the predicted incremental LYGs or QALYs were close to what observed in the complete dataset. Whilst the

mean incremental LYGs/QALYs are the outcome of interest in HE, the mean predicted health state sojourn time/QALYs and visual predictions for each arm need to be considered.

- the treatment effects in the case studies were very low, so assuming the effect is retained for a lifetime might not make much difference. If there was a big treatment effect that was lost over time, assuming it applies indefinitely might have more of an impact. However, the conclusion would remain the same that neither approach is bias-free.

In conclusion, this exploratory analysis shows that, despite not modelling explicitly the underlying natural history process, the PSM does not perform less well compared with the STM when estimating health state sojourn time/QALYs for the intervention arm. In fact, all these approaches, whilst reasonable, were often inaccurate to varying degree when compared with the predictions in the complete dataset. Unsurprisingly, this study further highlights the importance of clinical validation in order to select the most appropriate curve for the transitions in the STM or direct fit for the PSM.

PART V: IMPLICATIONS FOR DECISION-MAKERS AND ANALYSTS

10 CHAPTER X: KEY FINDINGS / STRENGTHS/LIMITATIONS / AREAS OF FURTHER RESEARCH

10.1 Chapter overview

In this chapter, I summarise the findings of this thesis, discuss its contribution to health economics, and discuss areas of future research. Findings from previous chapters are drawn together in order to formulate a series of recommendations for both analysts and decision-makers to help guide the choice of analytical models when assessing the cost-effectiveness of oncology treatments.

10.2 Overview of the thesis

In Chapter 1, I discussed the motivations for this thesis that alternative models are currently used to address the same decision problem but have the propensity to change the conclusions of an economic analysis; therefore the decisions made on the basis of these analyses. The aim of this thesis was to guide the choice of modelling approach for estimating health state sojourn time for anticancer therapies conditional on the nature of data available. In particular, this study attempted to address the following question – *"Is it possible to identify when particular analytical approaches may perform better than others, subject to the nature of the data available (e.g. under different levels of censoring, dependence and follow-up)?"*

In Chapter 2, I provided a description of the different approaches that are currently used in health economics when modelling oncology treatment. I summarised the underlying assumptions associated with approach, and described their strengths and limitations. I then discussed briefly previous comparisons available; and highlighted that approaches had not been compared systematically and therefore it is difficult to determine whether one approach is consistently superior to another. I therefore proposed in this chapter that a systematic comparison of the different approaches is required to identify whether one approach is superior to another.

In Chapter 3, I provided some theoretical background and a brief description of some of the key concepts used in survival analysis that are used throughout this thesis. These concepts are important as they underpin the alternative modelling approaches evaluated within the thesis.

In Chapter 4, I described the MSM – a type of STM which combines transitions under a competing risk framework. The MSM is rarely used in health economics despite the availability of several software packages and clear tutorials by Williams et al.[50] I focused on the implementation of the MSM using two packages available in R, the `msm` package and the `mstate` package. I showed that the MSM is easily implemented in R using these two packages and explained the key differences and how transitions are estimated and combined within these two packages. As highlighted by other researchers,[24, 50] I showed that using the MSM is relatively straightforward and therefore barriers for its use in health economics are low. While the MSM using the `mssample` has been previously described by Williams *et al*, this chapter is original and significant as it provides a comparison of the `msm` and `mssample` functions in R, but also provides an in-depth description on how transitions are combined within the function.

In Chapter 5, I described the simplifications that are typically made in health economic when modelling oncology treatments using the STM approach and highlighted how this "simple" approach compares with the competing risk approach (the MSM). I conducted a review of previous NICE appraisals in order to identify the key assumptions and understand how the STM is currently implemented in health economics. I showed that, whilst it is difficult to directly compare the MSM and the Simplified STM due to the differences in inputs, any differences are more likely to be the result of the choice of parametric functions and extrapolation of the transitions, rather than the approach itself. This chapter is original and significant as it clearly sets out the different implementations of the STM in health economics (to model oncology treatments), and provide a more direct and fairer comparison between the STM and MSM; showing little differences.

In Chapter 6, I presented the methods and results of a review of methods for the joint modelling of progression and survival. I described challenges associated with searching the methodological literature, and described how these were addressed through the use of iterative searching and use of expert opinion. Despite some overlap, identified methods could be categorised according to two groups: (1) methods that include the dependence between transitions in an illness-death model (joint/conditional modelling of transitions) and (2) methods that include the dependence between PFS and OS under a semi-competing risk framework. All of the approaches identified within this review had limitations. Example of implementation were also often not available in an appropriate software package. This chapter is original and significant as it summarises approaches that are currently available that could potentially be used in health economics to jointly model progression and survival.

In Chapter 7, I described the potential biases associated with the use of PPS estimated only in a subset of patients who progressed when generalised to the overall randomised population in the same randomised control trial. I illustrated these using hypothetical simulated datasets as well as real datasets.

I then explored whether simple approaches that are typically considered reduce biases and showed that adjustments that are currently proposed (suggested by analyts when I described my topic to external audience) may often be inaccurate. Consequently, I suggested that analysts should consider reporting both the unadjusted PPS and adjusted PPS in order to reflect the uncertainty so that it can be considered in the decision-making process. This chapter is original and significant as this demonstrated clearly potential biases associated with PPS when there is dependence in the data. It also highlighted that current suggested adjustments for PPS may not always have the desired effect of reducing bias.

In Chapter 8, I designed and implemented a simulation study specifically to address the research questions posed in the first chapter of this thesis. A simulation study was chosen because this allowed the bias associated with the methods to be compared given a known truth, during both the observed and unobserved period, which would have otherwise not been possible using published trial data. The simulation study followed the ADEMP framework set out by Morris *et al*.[139] When applied to single arm studies, I showed that whilst all approaches were prone to biases, the PSM and STM remained generally reasonable. I showed that it not possible to identify whether the PSM or STM fare "better" that the other based on observed data characteristics alone and that their performance differed in different scenarios. The PSM was affected more by the difficulty for simple parametric extrapolation to reflex complex hazard (changes in hazards during and beyond the observed period), whereas the STM also has the problem with the generalisability of PPS when there is dependence. Other methods (properly modelling competing risks, trying to adjust for prognosis in PPS, modelling a relationship between PFS and OS) generally made very little difference and didn't represent a clear improvement to the standard PSM and STM approaches. This chapter is original and significant as it provides a systematic comparison and showed that is not possible to select approaches based on the observed data characteristics alone and that the performance of methods was principally explained by unobserved data characteristics like how the hazard look beyond the observed period, and how well curves are selected.

Finally, in Chapter 9, I assessed the performance of methods in estimating incremental outcomes using six case-studies in gastric cancer. I showed that the PSM was not less accurate compared with the STM in predicting OS and QALYs for the intervention arm, even under the simplified assumption of the treatment effect applied throughout the model duration. In fact, all these approaches, whilst reasonable, were often inaccurate to varying degrees when compared with the predictions in the complete datasets. This chapter also re-iterated the importance of clinical validation when selecting extrapolation method. This chapter is original and significant as it provides a direct comparison of the performance between approaches using complete data and shows that all approaches are prone to biases, and that it is possible for none of the approach typically considered to generate accurate predictions for the incremental LYs/QALYs.

10.3 Discussion

This thesis demonstrated that approaches currently used in health economics models to estimate health state sojourn time are generally reasonable, but, could be inaccurate and that it is not possible to determine with certainty *a priori,* based only on the observed characteristics of the available data, whether one approach is likely to fare better than the other. While it is possible to learn something from looking at the hazard plot (to identify trend) or assement of relationship in the data (between PFS and PPS for instance), it remain unknow how the hazard would look like beyond the observed period.

In this thesis, I attempted to highlight the different strengths but also limitations, as well as the key underlying assumptions (when modelling the introduction of a new technology) associated with each approach. These needs to be recognised when selecting a particular approach, as without a proper understanding and knowledge of these, it is not possible to justify robustly why an approach should be used/prefered compared with another.

The aim of this thesis was to attempt to identify particular data characteristics that could help analysts and decision-makers select the most appropriate analytical method to estimate health state sojourn time for both the control and intervention arm. At present, there is no framework for making this choice, and modelling approaches are used inconsistently between appraisals, often with no or inadequate justification. My experience has been that analysts often have a preference for a particular approach. Some analysts also inappropriately assume naively that all approaches are the same, inaccurately referencing outputs from previous research that showed in some limited case study that both the PSM and STM generated similar predictions.[20] Identifying cases where a particular approach is perhaps more suitable compared with another is also valuable, given the ongoing debates on which model structure to use to answer a particular problem. For instance, within the NICE STA (or MTA) process, where the company submits a model, there are often disagreements between the structure chosen by the company and the one preferred by the ERG (or AG).

I showed through the use of a simulation study in single-trial arms, that is not possible to select approaches based on the observed data characteristics alone, and instead, whilst some trends were seen in terms of more observable data characteristics (such as when dependence is present in the data; although this is only partially observable), the performance of methods was primarily driven by unobserved data characteristics (in particular the shape of the hazard beyond the observed period; e.g turning points) but also how well curves are selected. As expected, when there are dependence in the data, modelling the underlying process was less adequate due to limitations associated with the generalisability of using PPS estimated only in the subset of patients who progressed. However, the

STM/MSM remain reasonable approaches, and the level of dependence examined in the simulation study was perhaps extreme compared with that observed in real trials. I do not suggest that the STM/MSM cannot or should not be used in case of dependence; more that that caution should be exercised and analysts need to have a full understanding of the potential limitations. Even so, there is no guarantee that an alternative approach would necessarily generate more accurate predictions. It is also important to recognised that when adjusting PPS by including TTP as a covariate in the statistical model (to account for the dependence) for example, the effect of this adjustment is unclear, and therefore I suggest that both the unadjusted and adjusted be presented for transparency.

In health economics, we are often more interested in incremental outcomes between arms. There is also an ongoing debate regarding whether the PSM (direct fit to OS) is appropriate, notably when modelling the treatment effect for the intervention. I showed through the use of six case studies that the PSM was no less accurate than the STM in predicting OS and QALYs for the intervention arm, even under the simplified assumption of the treatment effect applied throughout the model duration. In fact, whilst both approaches were generally reasonable, they were often inaccurate; with none approaches generating accurate predictions in some cases.

The MSM is rarely used in health ecoconomics. My experience has been that analysts often have a limited understanding of the differences between the MSM (formal modelling of the competing transitions under a competing risk framework) and the STM as implemented in health economics (which uses PFS directly). I showed in this thesis that both approaches generate similar results when curves are selected appropriately and that differences highlighted in previous research are perhaps more attributable to assumptions about the choice of parametric extrapolations, rather that the approach itself. A number of approaches are also available in the broader literature (outside health economics) to jointly model progression and survival. However, clear examples are not available for many of those. This is an important barrier given the need for transparency and technical skills of analysts typically in charge of building or reviewing cost-effectiveness models. Methods without a proper tutorial are unlikely to be adopted. The method developed by Fu model was identified to jointly model PFS and OS under a semi-competing risk using a copula; with a clear example available in a suitable statistical package. I showed that it is easy to implement and avoids arbitrary assumptions currently made in health economics for the structural relationship between survival endpoints where a constraint is added to ensure that PFS is consistent with OS. However, I also showed in the simulation study, that when the hazard of death and progression are complex, such approach may be less desirable.

When modelling the treatment arm, approaches make different assumption which need to be recognised and undertood. With the PSM, assumptions have to be made on how long the treatment effect will

persist (when using a HR). It is often unclear whether the treatment effect will persist over the patient's remaining lifetime or wanes after a certain time point. With the STM, typically, PPS is pooled between treatment arms when no statistical difference is observed between arms, and therefore the benefits of the intervention are driven primarily from the gain in PFS. However, assuming the same PPS between treatment arms may be an oversimplification. Indeed, despite no significant differences observed in the censored dataset, this may not necessarily mean that there would be no difference in PPS had the data been complete. Furthermore, as showed when dependence is present in the data, PPS estimated in the subset of patients who progressed is biased when generalised to the overall randomised population.

My experience has been that analysts have typically strong views and preferences for a particular approach; based on their perception of the strengths and limitations. Indeed, when describing my topic to external, but also internal audiences, strong views has been expressed either side in favour of an approach, which has been challenging to reconcile. In particular, depending on their model choice preferences, concerns were expressed that results may be biased against their preferred approach. In this thesis, because of these strong views, I tried to be as objective as possible, and not be influenced by any *a priori* preferences. As an example, when I started my thesis, my pre-conception was that when data are immature, using a STM (which allow an explicit modelling of the process) would always fare better compared with the direct fit to OS. This thesis showed that this is not necessarily the case, and that in fact, all approaches are biased, but subject to different limitations. Consequently, what is more relevant is to recognise, but also be able to communicate effectively the underlying assumptions and implications when selecting a particular approach.

While this thesis focused on the choice of analytical approach, extrapolation is a key component when determining the performance of a method. All these approaches use different inputs (transitions or survival endpoints). It is crucial that a robust extrapolation method is used as otherwise, the wrong conclusion will be made.

In summary, I showed in this thesis that relying on a single analytical approach (e.g economic model structure) is not advisable as it is not possible to know for certain if an approach fare better than the other (due to the need to extrapolate beyong the trial). This is an important finding because at present, decisions are often made based on the choices made *a priori* by the analyst responsible of building models, but often, without any knowledge on whether using a different approach would provide a different economic conclusion. The decision-making process should therefore capture both the uncertainty around the analytical approach and extrapolation method (clinical expectation).

10.4 Recommendations

Based upon findings from this thesis, I formulated the following recommendations for the choice of analytical approach when assessing the cost-effectiveness of oncology treatments for both analysts and decision-makers. It should be noted that these recommendations only apply to cases where there is a choice between the PSM and STM. It is possible that in some cases, the choice of analytical approach is driven by the decision problem instead, and therefore analysts have little choice to use a particular approach. Recommendations also consider the most common 3-state oncology model and it is recognised that some of the recommendations formulated below may be more difficult to implement when considering more health states; although not impossible.

10.4.1 Recommendations for analysts responsible for building models

As it is not possible to identify when one approach fare better than the other based on observable data characteristics alone, I formulated the following recommendations for analysts responsible for building oncology cost-effectiveness models; which I hope would increase both transparency, but also consistency between appraisals:

- to include automatically, as standard, the functionality within the cost-effectiveness model to assess the different analytical approaches; notably the PSM and STM (adjusted or not). My experience (during and prior to my PhD) has been that exploring alternative model choices does not require a significant time if the model is designed from the start to be flexible. Indeed, considering only the methods that are commonly used in health economics (the PSM and STM), including the functionality for the PSM when the model is initially built as a STM is relatively straightforward as the trace for OS and PFS can typically simply be over-written (unless tunnel state are required for other reasons such as costing). In contrast, if the model is initially built as a PSM (which is the case for many oncology appraisals), the addition of the functionality to assess the STM become more challenging and time consuming. Consequently, it is reasonable to recommend as a starting point that oncology models should follow as standard a STM structure, with the PSM approach applied subsequently. Decision-makers are more likely to be more confident in their decision-making if analysts can show that predictions between approaches are the same; although this can still be biased. If there are differences, these could then considered in the decision-making process. It is also valuable for analysts who built a STM to compare the modelled OS against the direct fit for the sake of validation.

- analysts need to be able to communicate transparently and effectively why one approach is preferred, and to explain the different assumptions made and their implications. This needs to be communicated effectively to decision-makers so that they are able to account for the uncertainty appropriately, and reject scenarios that are perhaps less appropriate. Ultimately, decision-makers should be responsible for selecting the approach they feel to be the most appropriate to form a basis for decision-making, rather than the other way around.
- analysts should be transparent about the different key assumptions when modelling the treatment effect and ensure that all options are included within the model as standard (even if deemed not relevant by the analyst). This is important as decision-makers or other analysts may not share the view expressed by the analyst responsible for building models and may want to consider the uncertainty around different assumptions. This can also act as a matter of validation.
- analysts need to be able to communicate transparently and effectively how clinical validation is undertaken. The extrapolation method is crucial. Clinical opinion is often used inconsistently between appraisals. This process needs to be transparent and standardised.
- when constructing a STM, results using both the adjusted and unadjusted STM should be reported as standard in order for decision-makers to capture any uncertainty (as it cannot be known for certainty which one is more appropriate). Indeed, if predictions using the adjusted and unadjusted PPS vary widely, this uncertainty should be reflected in the decision-making process, but also explained by analysts. Decision-makers are more likely to be confident in their decision-making if predictions using the adjusted and unadjusted PPS are similar. But, there remains a risk that neither are correct. This is unknown, and decision-makers will need to consider the possibility that all these estimates are uncertain and may be biased.
- when using a PSM and selected curves for PFS and OS appear to cross, analyts should consider reporting results by jointly modelling PFS and OS (in addition to independent) using the Fu's model.

10.4.2 Recommendations for decision makers and model reviewers

In addition to recommendations for analysts in charge of building oncology cost-effectiveness models, I formulated the following recommendation for for decision-makers and model reviewers; which if adopted I hope would ensure consistent decision-making:

- there is a need to develop a framework to standardise how, in what format, and level of information (i.e key assumptions, predictions for each approaches) that should presented to decision-makers. Decision-makers often have a large amount of information to process, even when considering a single modelling approach. If multiple scenarios are presented, a clear framework needs to be developed for decision-makers to avoid information overload.
- training is required to ensure that decision-makers fully understand the underlying assumptions for each modelling approach.
- decision-makers should request that different structures are explored to ensure that the decision is not influenced by the choice of approach. If alternative structures are not presented, decision-makers may want to consider this uncertainty within their decision-making process.
- a framework for handling structural uncertainty within the decision-making process would be helpful to promote consistency between appraisals.
- consider the uncertainty associated with expert elicitation for survival model selection

10.5 Area of future research

This thesis demonstrates that analysts and decision-makers cannot rely solely on predictions from a single approach, but should instead consider the structural uncertainty associated with the modelling approach. However, there remain a number of methodological issues which require further research:

(i) scenarios covered in this simulation study were limited to non-cure processes. There is an increasing number of novel treatments undergoing assessments which have the potential to offer a cure. It may therefore be informative to explore whether the same findings would be seen if cure processes were considered.

(ii) additional research on how to account for the structural uncertainty in the decision-making process is required. Should recommendations from this thesis be taken forward by analysts and decision-makers, a transparent and structured process needs to be defined to understand how structural uncertainty could be accounted for within the decision-making process. For instance, if two approaches provide very different answers, how should decision-makers synthesise and

use this information? While there is already some research on how to account for structural uncertainty within models, this needs to be extended to consider decision-makers preferences. Other methods could be considered such as model averaging for instance or multi-criteria decision analysis (MCDA) type process.

(iii) there is also a need for further research to assess how best to elicit expert information to inform survival model selection and how uncertainty associated with this approach is accounted for. Whilst there is general guidance on how to select extrapolation methods, these guidelines do not provide clear guidance on how clinical opinion should be used when selecting the most appropriate extrapolation method. My experience has been that clinical opinion is used inconsistently between appraisals. Standardisation is important as inaccurate selection for the extrapolation method, lead to incurrate method's prediction. Research is also needed on how to capture the uncertainty; for instance bayesian model average accounting for both the statistical fit and clinical expectation,

(iv) a number of approaches to jointly model progression and survival were identified in the review of methods. Some of these could not be assessed due to the absence of a clear tutorial/example. This is a key barrier to their implementation. Further research is therefore required on how to implement these methods, should they be taken forward in health economics,

(v) it is important to understand from both decision-makers but also analysts, the potential barriers they foresee for the recommendations formulated in this thesis. This is important as decision-making is already challenging. My recommendation adds significant additional complexity which needs to be recognised. For analysts responsible for building or reviewing cost-effectiveness models, this will add work. For decision-makers this adds an additional level of complexity and more information to process. Additional research should be conducted to understand what information decision-makers require and to understand whether the additional analysis demands can be met by analysts or is useful for decision makers

(vi) Additional research is required to develop alternative methods that could be used to adjust PPS. In Chapter 7, I explored simple adjustments to PPS. For instance, IPCW has been suggested as an alternative approach. It is unclear if such approach or an alternative would necessarily be more accurate,

(vii) Incorporation of bayesian method within the development of the PSM and STM being expanded.

10.6 Conclusion

I aimed to guide the choice of modelling approach for cost-effectiveness models for oncology treatment; in particular identify if and when one approach may fare better than the other based on observed data characteristics.

I demonstrated in both my simulation study, but also case studies, that this is as not simple, and that approaches cannot be selected based on observed data characteristics alone as that their performance/appropriateness is primarily explained by unobserved data characteristics such as the complexity of the underlying hazards (for both OS and PFS), and the parametric model selection process. This is an important finding.

Analysts and decision-makers should therefore be careful when relying on predictions from a single approach. It is unknown whether ICERs generated using a single analytic approach are adequate, or whether, in some cases, decision-making should consider ICERs from a range of alternative approaches to account for these possible structural uncertainty.

Based upon findings from this thesis, I have formulated series of recommendations for both analysts and decision-makers, which I believe if they are taken forward would improve the transparency of health economic analyses and increase decision-makers' confidence in the use of those models.

REFERENCE

1. Excellence National Institute for Health and Care. Guide to the methods of technology appraisal. 2013. Available from URL: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 (Accessed 20 sept 2020).
2. Excellence National Institute for Health and Care. Judging whether public health interventions offer value for money. 2013. Available from URL: https://www.nice.org.uk/advice/lgb10 (Accessed 20 Sept 2020).
3. Drummond MF Sculpher M, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes (3rd edition). ; 2005.
4. Excellence National Institute for Health and Care. NICE highly specialised technologies guidance. 2020. Available from URL: https://www.nice.org.uk/About/What-we-do/Our-Programmes/NICE-guidance/NICE-highly-specialised-technologies-guidance (Accessed 20 Sept 2020).
5. Administration Food and Drug. Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. U.S. Department of Health and Human Services,; 2007. Available from URL: https://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf (Accessed 20 Sept 2020).
6. Agency European Medicines. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man. 2008. Available from URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137126.pdf (Accessed 20 Sept 2020).
7. Ciani O Davis S, Tappenden P, Garside R, Stein K, Cantrell A, Saad ED, Buyse M, Taylor RS. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care* 2014;30:312-24.
8. Mills EJ Bucher HC, D'Oca K, Rafia R. Progression free survival versus overall survival as important clinical endpoints in cancer clinical trials: can improved validation of surrogate endpoints improve the utility of trial evidence? . 2014. Available from URL: https://www.ispor.org/home (Accessed 20 Sept 2020).
9. Woods B Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. 2017. Available from URL: http://www.nicedsu.org.uk (Accessed 20 Sept 2020).
10. Tappenden P Brennan A, Squires H, Stevenson M. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. *Value Health* 2012;15:1127-36.
11. N Latimer. The role of treatment crossover adjustment methods in the context of economic evaluation. 2012. Available from URL: http://etheses.whiterose.ac.uk/3720/1/Thesis_Final_with_corrections.pdf (Accessed 20 Sept 2020).
12. N Latimer. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2013. Available from URL: http://www.nicedsu.org.uk (Accessed 20 Sept 2020).
13. N Latimer. Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. Inconsistencies, limitations, and a practical guide. *Medical Decision Making* 2013;33 743-54.
14. N Latimer. Response to "Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach" by Bagust and Beal. . *Medical Decision Making* 2013;34:279-82.
15. I Cromwell. Development and application of a whole disease model of oral cancer to inform health technology management. University of British Columbia; 2019. Available from URL: https://open.library.ubc.ca/clRcle/collections/ubctheses/24/items/1.0379922 (Accessed 20 Sept 2020).
16. Brennan A Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006;15:1295-310.
17. Barton P Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy* 2004;9:110-8.

18. Bullement A Cranmer HL, Shields GE,. A Review of Recent Decision-Analytic Models Used to Evaluate the Economic Value of Cancer Treatments. *Applied Health Economics and Health Policy* 2019;17:771-80.

19. Excellence National Institute for Health and Care. TA257. Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses HER2. 2012. Available from URL: http://publications.nice.org.uk/lapatinib-or-trastuzumab-in-combination-withanaromataseinhibitor-for-the-first-line-treatment-of-ta257 (Accessed 20 Sept 2020).

20. Excellence National Institute for Health and Care. TA472. Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab. 2017. Available from URL: https://www.nice.org.uk/guidance/ta472 (Accessed 20 Sept 2020).

21. Excellence National Institute for Health and Care. TA381: Olaparib for maintenance treatment of relapsed, platinum-sensitive, BRCA mutation-positive ovarian, fallopian tube and peritoneal cancer after response to second-line or subsequent platinum-based chemotherapy. 2016. Available from URL: https://www.nice.org.uk/guidance/ta381 (Accessed 20 Sept 2020).

22. Damien P Muller P. A Bayesian bivariate failure time regression model. *Computational Statistics & Data Analysis* 1998;28:77-85.

23. Briggs A Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics* 1998;13:397-409.

24. Williams C Lewsey JD, Mackay DF, Briggs AH. Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. *Med Decis Making* 2017;37:427-39.

25. Briggs A Baker TM, Gilloteau I, Orsini L, Wagner S, Paly V. Partitioned survival versus state transition modelling in oncology: a case study with nivolumab in advanced melanoma. *Value in Health* 2015;18:A338.

26. Smare C Lakhdari K, Doan J, Posnett J, Johal S. Evaluating Partitioned Survival and Markov Decision-Analytic Modeling Approaches for Use in Cost-Effectiveness Analysis: Estimating and Comparing Survival Outcomes. *PharmacoEconomics* 2020;38:97-108.

27. Batteson R Hart R, Hemstock M, Gooden K, Kotapati S, Roze S, Lee D,Amadi A. Modelling Survival of Patients Treated with Adjuvant Nivolumab Who Have Melanoma with Lymph Node Involvement or Metastatic Disease After Complete Resection. *PharmacoEconomics* 2020;4:343-51.

28. Cranmer H Shields GE, Bullement A. A comparison of partitioned survival analysis and state transition multi-state modelling approaches using a case study in oncology. *J Med Econ* 2020;23:1176-85.

29. D Collett. Modelling Survival Data in Medical Research, Third Edition; 2014.

30. Sphere Project Data. CALGB 40502: A Randomized Phase III Trial of Weekly Paclitaxel Compared to Weekly Nanoparticle Albumin Bound Nab-paclitaxel or Ixabepilone With or Without Bevacizumab as First-Line Therapy for Locally Recurrent or Metastatic Breast 2020. Available from URL: https://www.projectdatasphere.org/projectdatasphere/html/home (Accessed 20 Sept 2020).

31. Rugo HS Barry WT, Moreno-Aspitia A, Lyss AP, Cirrincione C, Leung E, Mayer EL, Naughton M, Toppmeyer D, Carey LA, Perez EA, Hudis C, Winer EP. Randomized Phase III Trial of Paclitaxel Once Per Week Compared With Nanoparticle Albumin-Bound Nab-Paclitaxel Once Per Week or Ixabepilone With Bevacizumab As First-Line Chemotherapy for Locally Recurrent or Metastatic Breast Cancer: CALGB 40502/NCCTG N063H (Alliance). *J Clin Oncol* 2015;33:2361-9.

32. Fizazi K Higano CS, Nelson JB, Gleave M, Miller K, Morris T, Nathan FE, McIntosh S, Pemberton K, Moul JW. Phase III, randomized, placebo-controlled study of docetaxel in combination with zibotentan in patients with metastatic castration-resistant prostate cancer. . *J Clin Oncol* 2013;31:1740-7.

33. Pirker R Ramlau RA, Schuette W, Zatloukal P, Ferreira I, Lillie T, Vansteenkiste JF. Safety and efficacy of darbepoetin alpha in previously untreated extensive-stage small-cell lung cancer treated with platinum plus etoposide. *J Clin Oncol* 2008;26:2342-9.

34. Leung KM Elashoff RM, Abdelmonem AA. Censoring issues in survival analysis. *Annu Rev Public Health* 1997;18:83-104.

35. Wikipedia. Kaplan–Meier estimator. Available from URL: https://en.wikipedia.org/wiki/Kaplan%E2%80%93Meier_estimator (Accessed 20 Sept 2020).

36. Wikipedia. Nelson–Aalen estimator. Available from URL: https://en.wikipedia.org/wiki/Nelson%E2%80%93Aalen_estimator (Accessed 20 Sept 2020).

37. DR Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)* 1972;34:187-220.

38. PC Lambert. Welcome and Introduction to Flexible Parametric Survival Models. *Workshop on Applications and Developments of Flexible Parametric Survival Models (Stockholm 10/11/2011)* 2011.

39. P Royston. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002;21:2175-97.

40. Zare A Hosseini M, Mahmoodi M, Mohammad K, Zeraati H,Holakouie Naieni K. A Comparison between Accelerated Failure-time and Cox Proportional Hazard Models in Analyzing the Survival of Gastric Cancer Patients. *Iran J Public Health* 2015;44:1095-102.

41. Population Health Methods: Competing Risk Analysis. 2020. Available from URL: https://www.publichealth.columbia.edu/research/population-health-methods/competing-risk-analysis (Accessed

42. Fine JP Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 1999;94:496-509.

43. Putter H Fiocco M, Gesku RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007;26:2389-430.

44. Andersen PK Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res* 2002;11:91-115.

45. C Jackson. Multi-state modelling with R: the msm package. 2016. Available from URL: https://cran.r-project.org/web/packages/msm/vignettes/msm-manual.pdf (Accessed 20 Sept 2020).

46. Excellence National Institute for Health and Care. TA586 :Multiple myeloma - lenalidomide (post bortezomib) (part rev TA171). 2019. Available from URL: https://www.nice.org.uk/guidance/TA586 (Accessed 20 Sept 2020).

47. Excellence National Institute for Health and Care. TA587: Multiple myeloma (newly diagnosed) - lenalidomide. 2019. Available from URL: https://www.nice.org.uk/guidance/TA587 (Accessed 20 Sept 2020).

48. Excellence National Institute for Health and Care. Abiraterone for treating newly diagnosed high risk metastatic hormone-naive prostate cancer [ID945]. 2019. Available from URL: https://www.nice.org.uk/guidance/indevelopment/gid-ta10122 (Accessed 20 Sept 2020).

49. Crowther MJ Lambert PC. Multistate: User-friendly Stata software. 2020. Available from URL: https://www.mjcrowther.co.uk/software/multistate/ (Accessed

50. Williams C Lewsey JD, Mackay DF, Briggs AH. Cost-effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. *Med Decis Making* 2017;37:340-52.

51. Excellence National Institute for Health and Care. TA263: Bevacizumab in combination with capecitabine for the first-line treatment of metastatic breast cancer. 2012. Available from URL: https://www.nice.org.uk/guidance/TA263 (Accessed 20 Sept 2020).

52. Excellence National Institute for Health and Care. TA214: Bevacizumab in combination with a taxane for the first-line treatment of metastatic breast cancer. 2011. Available from URL: https://www.nice.org.uk/guidance/ta214 (Accessed 20 Sept 2020).

53. Excellence National Institute for Health and Care. TA563: Abemaciclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. 2019. Available from URL: https://www.nice.org.uk/guidance/ta563 (Accessed 20 Sept 2020).

54. Excellence National Institute for Health and Care. TA496: Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. 2017. Available from URL: https://www.nice.org.uk/guidance/ta496 (Accessed 20 Sept 2020).

55. Excellence National Institute for Health and Care. TA593: Ribociclib with fulvestrant for treating hormone receptor-positive, HER2-negative, advanced breast cancer. 2020. Available from URL: https://www.nice.org.uk/guidance/indevelopment/gid-ta10448/documents (Accessed 20 Sept 2020).

56. Excellence National Institute for Health and Care. TA604: Idelalisib for treating follicular lymphoma refractory to 2 treatments. 2019. Available from URL: https://www.nice.org.uk/guidance/ta604 (Accessed 20 Sept 2020).

57. Excellence National Institute for Health and Care. TA243: Rituximab for the first-line treatment of stage III-IV follicular lymphoma. 2012. Available from URL: https://www.nice.org.uk/guidance/ta243 (Accessed 20 Sept 2020).

58. Excellence National Institute for Health and Care. TA226: Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma. 2011. Available from URL: https://www.nice.org.uk/guidance/ta226 (Accessed 20 Sept 2020).

59. Excellence National Institute for Health and Care. TA513: Obinutuzumab for untreated advanced follicular lymphoma. 2018. Available from URL: https://www.nice.org.uk/guidance/ta513 (Accessed 20 Sept 2020).

60. Excellence National Institute for Health and Care. TA380: Panobinostat for treating multiple myeloma after at least 2 previous treatments. 2016. Available from URL: https://www.nice.org.uk/guidance/ta380 (Accessed 20 Sept 2020).

61. Excellence National Institute for Health and Care. TA343: Obinutuzumab in combination with chlorambucil for untreated chronic lymphocytic leukaemia. 2015. Available from URL: https://www.nice.org.uk/Guidance/TA343 (Accessed 20 Sept 2020).

62. Excellence National Institute for Health and Care. TA193: Rituximab for the treatment of relapsed or refractory chronic lymphocytic leukaemia. 2010. Available from URL: https://www.nice.org.uk/Guidance/ta193 (Accessed 20 Sept 2020).

63. Excellence National Institute for Health and Care. TA174: Rituximab for the first-line treatment of chronic lymphocytic leukaemia. 2009. Available from URL: https://www.nice.org.uk/guidance/ta174 (Accessed 20 Sept 2020).

64. Excellence National Institute for Health and Care. TA400: Nivolumab in combination with ipilimumab for treating advanced melanoma. 2016. Available from URL: https://www.nice.org.uk/Guidance/TA400 (Accessed 20 Sept 2020).

65. Excellence National Institute for Health and Care. TA 384: Nivolumab for treating advanced (unresectable or metastatic) melanoma. 2016. Available from URL: https://www.nice.org.uk/guidance/ta384 (Accessed 20 Sept 2020).

66. Excellence National Institute for Health and Care. TA370: Bortezomib for previously untreated mantle cell lymphoma. 2015. Available from URL: https://www.nice.org.uk/guidance/ta370 (Accessed 20 Sept 2020).

67. Excellence National Institute for Health and Care. TA502:Ibrutinib for treating relapsed or refractory mantle cell lymphoma. 2018. Available from URL: https://www.nice.org.uk/guidance/ta502 (Accessed 20 Sept 2020).

68. Excellence National Institute for Health and Care. TA578: Durvalumab for treating locally advanced unresectable non-small-cell lung cancer after platinum-based chemoradiation. 2019. Available from URL: https://www.nice.org.uk/guidance/ta578 (Accessed 20 Sept 2020).

69. Excellence National Institute for Health and Care. TA258: Erlotinib for the first-line treatment of locally advanced or metastatic EGFR-TK mutation-positive non-small-cell lung cancer. 2012. Available from URL: https://www.nice.org.uk/guidance/ta258 (Accessed 20 Sept 2020).

70. Excellence National Institute for Health and Care. TA387: Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated. 2016. Available from URL: https://www.nice.org.uk/guidance/ta387 (Accessed 20 Sept 2020).

71. Excellence National Institute for Health and Care. TA386.Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. 2016. Available from: https://www.nice.org.uk/Guidance/TA386 (Accessed 20 Sept 2020).

72. Excellence National Institute for Health and Care. TA439: Cetuximab and panitumumab for previously untreated metastatic colorectal cancer. 2017. Available from URL: https://www.nice.org.uk/guidance/ta439 (Accessed 20 Sept 2020).

73. Excellence National Institute for Health and Care. TA491: Ibrutinib for treating Waldenstrom's macroglobulinaemia. 2019. Available from URL: https://www.nice.org.uk/guidance/ta491 (Accessed

74. Project Data Sphere. 2020. Available from URL: https://www.projectdatasphere.org (Accessed 20 Sept 2020).

75. Dissemination Centre for Reviews and. Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care, University of York. 2009. Available from URL: www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf (Accessed 20 Sept 2020).

76. Higgins JPT Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011. Available from URL: www.cochrane-handbook.org (Accessed 20 Sept 2020).

77. Schlosser RW Wendt O, Bhavnani S, Nail-Chiwetalu B. Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review. *Int J Lang Commun Disord* 2006;41:567-82.

78. A Booth. Unpacking your literature search toolbox: on search styles and tactics. *Health Information and Libraries Journal* 2008;25:313-7.

79. Hutton JL Ashcroft R. What does 'systematic' mean for reviews of methods? Black N, Brazier J, Fitzpatrick R, Reeves B, editors. . In: Black N, editor. *Health services research methods: a guide to best practice*: London, BMJ Books; 1998.

80. Edwards SJL Lilford RJ, Kiauka S,. Different types of systematic review in health services research In: Black N BJ, Fitzpatrick R, Reeves B, editors., editor. *Health services research methods: a guide to best practice*: London: BMJ Books; 1998.

81. S Paisley. Identifying evidence to inform decision-analytic models of cost-effectiveness : a qualitative study of information seeking processes and behaviour. . 2012. Available from URL: https://eprints.whiterose.ac.uk/98498/1/Paisley_Pharmacoeconomics_2016_accepted_version.pdf (Accessed 20 Sept 2020).

82. Mihaylova B Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ* 2011;20:897-916.

83. Tosh TJ. Simulation optimisation to inform economic evaluations of sequential therapies for chronic conditions: a case study in Rheumatoid Arthritis. *PHD Thesis, University of Sheffield* 2015.

84. Glasziou PP Cole BF, Gelber RD, Hilden J, Simes RJ. Quality adjusted survival analysis with repeated quality of life measures. *Stat Med* 1998;17:1215-29.

85. C Jackson. Multi-State Models for Panel Data: The msm Package for R. *JournalofStatisticalSoftware* 2011;38.

86. Belkacemi MC Castelli C, Remita MR, Fournel P, Daures JP. Modelling of overall survival by an association between progression-free and post-progression survival using a conditional distribution. *Statistical Modelling* 2014;14:77-98.

87. Dejardin D Lesaffre E, Verbeke G. Joint modeling of progression-free survival and death in advanced cancer clinical trials. *Statistics in Medicine* 2010;29:1724-34.

88. Fleischer F Gaschler-Markefski B, Bluhmki E. A statistical model for the dependence between progression-free survival and overall survival. *Statistics in Medicine* 2009;28:2669-86.

89. Fu H Wang Y, Liu J, Kulkarni PM, Melemed AS. Joint modeling of progression-free survival and overall survival by a Bayesian normal induced copula estimation model. *Stat Med* 2013;32:240-54.

90. Król A Mauguen A, Mazroui Y, Laurent A, Michiels S, Rondeau V. Tutorial in Joint Modeling and Prediction: a Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event. *Journal of Statistical Software* 2017;81.

91. Lia YM Zhang Q. A Weibull multi-state model for the dependence of progression-free survival and overall survival. *Statistics in Medicine* 2015;34:2497-513.

92. Mazroui Y Mathoulin-Pelissier S, Soubeyran P, Rondeau V. General joint frailty model for recurrent event data with a dependent terminal event: Application to follicular lymphoma data. *Stat Med* 2012;31:1162-76.

93. Meller M Beyersmann J, Rufibach K,. Joint modelling of progression-free and overall survival and computation of correlation measures. *Statistics in Medicine* 2019;38:4270-89.

94. D Oakes. A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B-Methodological* 1982;44:414-22.

95. Rondeau V Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* 2007;8:708-21.

96. Rondeau V Mazroui Y, Gonzalez JR,. An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation. *Journal of Statistical Software* 2012;47.

97. Weber EM Titman AC. Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's tau. *Statistics in Medicine* 2019;38:703-19.

98. Sildnes B Lindqvist BH. Modeling of semi-competing risks by means of first passage times of a stochastic process. *Lifetime Data Analysis* 2017;24:153-75.

99. Le-Rademacher JG Peterson RA, Therneau TM, Sanford BL, Stone RM, Mandrekar SJ. Application of multi-state models in cancer clinical trials. *Clin Trials* 2018;15:489-98.

100. Crowther MJ Lambert PC. Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. *Stat Med* 2017;36:4179-742.

101. Wilkerson J Abdallah K, Hugh-Jones C, Curt G, Rothenberg M, Simantov R, Murphy M, Morrell J, Beetsch J, Sargent DJ, Scher HI, Lebowitz P, Simon R, Stein WD, Bates SE, Fojo T. Estimation of tumour regression and growth rates during treatment in patients with advanced prostate cancer: a retrospective analysis. *Lancet Oncol* 2017;18:143-54.

102. Madan J Chen YF, Aveyard P, Wang D, Yahaya I, Munafo M, Bauld L, Welton N. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *Journal of the royal Statistical Society Series A* 2014;177:295-314.

103. Riley RD Jackson D, Salanti G, Burke DL, Price M, Kirkham J, White IR. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples *BMJ* 2017;358.

104. Riley RD Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008;9:172-86.

105. Amdahl J Chen L, Delea TE. Network Meta-analysis of Progression-Free Survival and Overall Survival in First-Line Treatment of BRAF Mutation-Positive Metastatic Melanoma. *Oncol Ther* 2016;4:239-56.

106. Excellence National Institute for Health and Care. TA561: Venetoclax with rituximab for previously treated chronic lymphocytic leukaemia. 2019. Available from URL: https://www.nice.org.uk/guidance/ta561 (Accessed 20 Sept 2020).

107. Excellence National Institute for Health and Care. Lung cancer: diagnosis and management NICE guideline [NG122]. 2019. Available from URL: https://www.nice.org.uk/guidance/ng122 (Accessed 20 Sept 2020).

108. T Snowsill. A New Method for Model-Based Health Economic Evaluation Utilizing and Extending Moment-Generating Functions. *Med Decis Making* 2019;39:523-39.

109. Oakes D. A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B-Methodological* 1982;44:414-22.

110. Fournel P Robinet G, Thomas P, Souquet PJ, Lêna H, Vergnenégre A, Delhoume JY, Le Treut J, Silvani JA, Dansin E. Randomized phase III trial of sequential chemoradiotherapy compared with concurrent chemoradiotherapy in locally advanced non-small-cell lung cancer: Groupe Lyon-Saint-Etienne d'Oncologie Thoracique – Groupe Francais de Pneumo-Cancérologie NPC 95-01 study. *Journal of Clinical Oncology* 2005;23:5910-7.

111. P Hougaard. Modeling multivariate survival. *Scandinavian Journal of Statistics* 1987;14:291-304.

112. Liu D Kalbfleisch JD, Schaubel DE. A Positive Stable Frailty Model for Clustered Failure Time Data with Covariate-Dependent Frailty. *Biometrics* 2011;67:8-17.

113. D Dejardin. Statistical models for the analysis of oncology endpoints. 2013. Available from URL: https://ibiostat.be/publications/phd/daviddejardin (Accessed 20 Sept 2020).

114. Y Wang. Estimation of accelerated failure time models with random effects. 2007. Available from URL: https://lib.dr.iastate.edu/rtd/3062/ (Accessed 20 Sept 2020).

115. Rice JD Tsodikov A. Semiparametric Time-to-Event Modeling in the Presence of a LatentProgression Event. *Biometrics* 2017;73:463-72.

116. Xu J Kalbfleisch JD, Tai B. Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics* 2010;66:716-25.

117. Liu L Huang X, Yaroshinsky A, Cormier JN. Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. *Biometrics* 2016;72:204-14.

118. Goethals K Janssen P, Duchateau L. Frailty models and copulas: similarities and differences. *Journal of Applied Statistics* 2008;35:1071-9.

119. Nelsen RB. An introduction to copulas. : Springer-Verlag New York; 2006.

120. Klara Goethals Paul Janssen & Luc Duchateau. Frailty models and copulas: similarities and differences. *Journal of Applied Statistics* 2008:1071-9.

121. NPTEL. M5L11: Introduction to Copulas1. Available from URL: https://nptel.ac.in/courses/105/105/105105138/ (Accessed 20 Sept 2020).

122. Dorey M Joubert P. Modelling Copulas: An Overview. Online do: The staple Inn Actuarial Society; 2005. Available from URL: http://not-normal-consulting.co.uk/web_documents/modellingcopulas.pdf (Accessed 20 Sept 2020).

123. Wikipedia. Copula (probability theory). Available from URL: https://en.wikipedia.org/wiki/Copula_(probability_theory) (Accessed 20 Sept 2020).

124. Shih JH Louis TA. Inferences on association parameter in copula models for bivariate survival data. *Biometrics* 1995;51:1384-99.

125. J Segers. Copulas: An Introduction I - Fundamentals. 2013. Available from URL: http://www.columbia.edu/~rf2283/Conference/1Fundamentals%20(1)Seagers.pdf (Accessed 20 Sept 2020).

126. Emura T Nakatochi M, Murotani K, Rondeau V. A joint frailty-copula model between tumour progression and death for meta-analysis. *Stat Methods Med Res* 2015;26:2649-66.

127. Foster NR Qi Y, Shi Q, Krook JE, Kugler JW, Jett JR, Molina JR, Schild SE, Adjei AA, Mandrekar SJ. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer* 2011;117:1262-71.

128. Felizzi Bennett I, Pletscher M, Thuresson P, Paracha N, Ray J. PRM 155: Joint modeling of overall survival and progression-free survival. *Value in Health* 2018;21:S382-S3.

129. Rotolo F Legrand C, Van Keilegom I. A simulation procedure based on copulas to generate clustered multi-state survival data. *Comput Methods Programs Biomed* 2013;109:305-12.

130. Furman E Kuznetsov A, Su J, Zitikis R. Tail dependence of the Gaussian copula revisited. *Insurance: Mathematics and Economics* 2016;69:97-103.

131. ÇT Gülöksüz. Comparison of Some Selection Criteria for Selecting Bivariate Archimedean Copulas. *AKÜ FEMÜBİD* 2016;16:250-5.

132. Xue X Brookmeyer R. Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Anal* 1996;2:277-89.

133. Shemyakin AE Youn H. Copula models of joint last survivor analysis. *Bayesian Models in Business and Industry* 2006;22:211-24.

134. SM Karadimitriou. Correlation in R. Available from URL: https://www.sheffield.ac.uk/polopoly_fs/1.536458!/file/MASH_Correlation_R.pdf (Accessed 20 Sept 2020).

135. group The GASTRIC. Benefit of adjuvant chemotherapy for resectable gastric cancer: a metaanalysis. *JAMA* 2010;303:1729-37.

136. Oba K Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, Catalano P, Lordick F, Michiels S, Morita S, Ohashi Y, Pignon JP, Rougier P, Sasako M, Sakamoto J, Sargent D, Shitara K, Cutsem EV, Buyse M, Burzykowski T; GASTRIC group. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. *J Natl Cancer Inst* 2013;105:1600-7.

137. Buyse M Molenberghs G, Paoletti X, Oba K, Alonso A, Van der Elst W, Burzykowski T. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biom J* 2016;58:104-32.

138. Davies CC Briggs A, Lorgelly P, Garellick G. The ‘‘Hazards’’ of Extrapolating Survival Curves. *medical decision making* 2013;33:369-80.

139. Morris TP White I, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med* 2019;38:2074-102.

140. Whyte S Walsh C, Chilcott J. Bayesian Calibration of a Natural History Model with Application to a Population Model for Colorectal Cancer *Med Decis Making* 2011;31:625-41.

141. optim {stats} - General-purpose Optimization. Available from URL: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/optim.html (Accessed 20 Sept 2020).

142. Finn RS Martin M, Rugo HS, Jones S, Im SA, Gelmon K, Harbeck N et al. Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* 2016;375:1925-36.

143. Hortobagyi GN Stemmer SM, Burris HA, Yap YS, Sonke GS, Paluch-Shimon S et al. Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N Engl J Med* 2016;375:1738-48.

144. (ONS) Office for National Statistics. Interim life tables from the Office of National Statistics (based on 2011-2013 UK population data) 2014. Available from URL: http://www.statistics.gov.uk/ (Accessed 20 Sept 2020).

145. Karadimitriou SM. Correlation in R. *University of Sheffield*.

146. Zaorsky NG Churilla TM, Egleston BL, Fisher SG, Ridge JA, Horwitz EM, Meyer JE. Causes of death among cancer patients. *Ann Oncol* 2017;28:400–7.

147. GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group Oba K, Paoletti X, Bang YJ, Bleiberg H, Burzykowski T, Fuse N, Michiels S, Morita S, Ohashi Y, Pignon JP, Rougier P, Sakamoto J, Sargent D, Sasako M, Shitara K, Tsuburaya A, Van Cutsem E, Buyse M. Role of chemotherapy for advanced/recurrent gastric cancer: an individual-patient-data meta-analysis. *Eur J Cancer* 2013;49:1565-77.

148. Li Y Sun JG, Song SG. Statistical analysis of bivariate failure time data with Marshall-Olkin Weibull models. *Computational Statistics & Data Analysis* 2012;56:2041-50.

149. Gasparini A White A. rsimsum: Analysis of Simulation Studies Including Monte Carlo Error. 2020. Available from URL: https://CRAN.R-project.org/package=rsimsum (Accessed 20 Sept 2020).

150. Rücker G Schwarzer G. Presenting simulation results in a nested loop plot. *BMC Med Res Methodol* 2014;14.

APPENDICES

Appendix 1 : Parametric fit to the BC dataset

Figure 67 : Extrapolation for PFS in the BC dataset

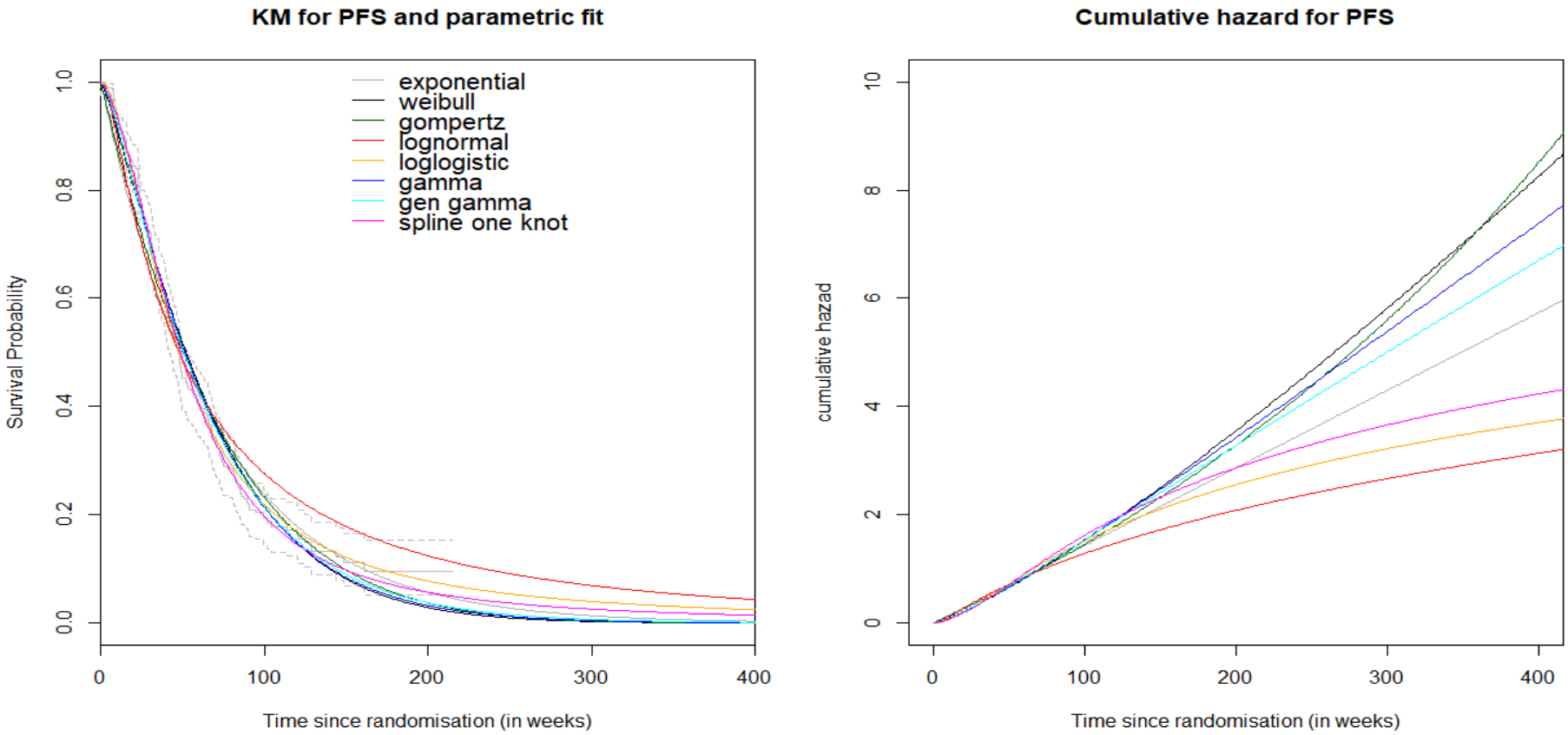


Figure 68 : Extrapolation for TTP in the BC dataset

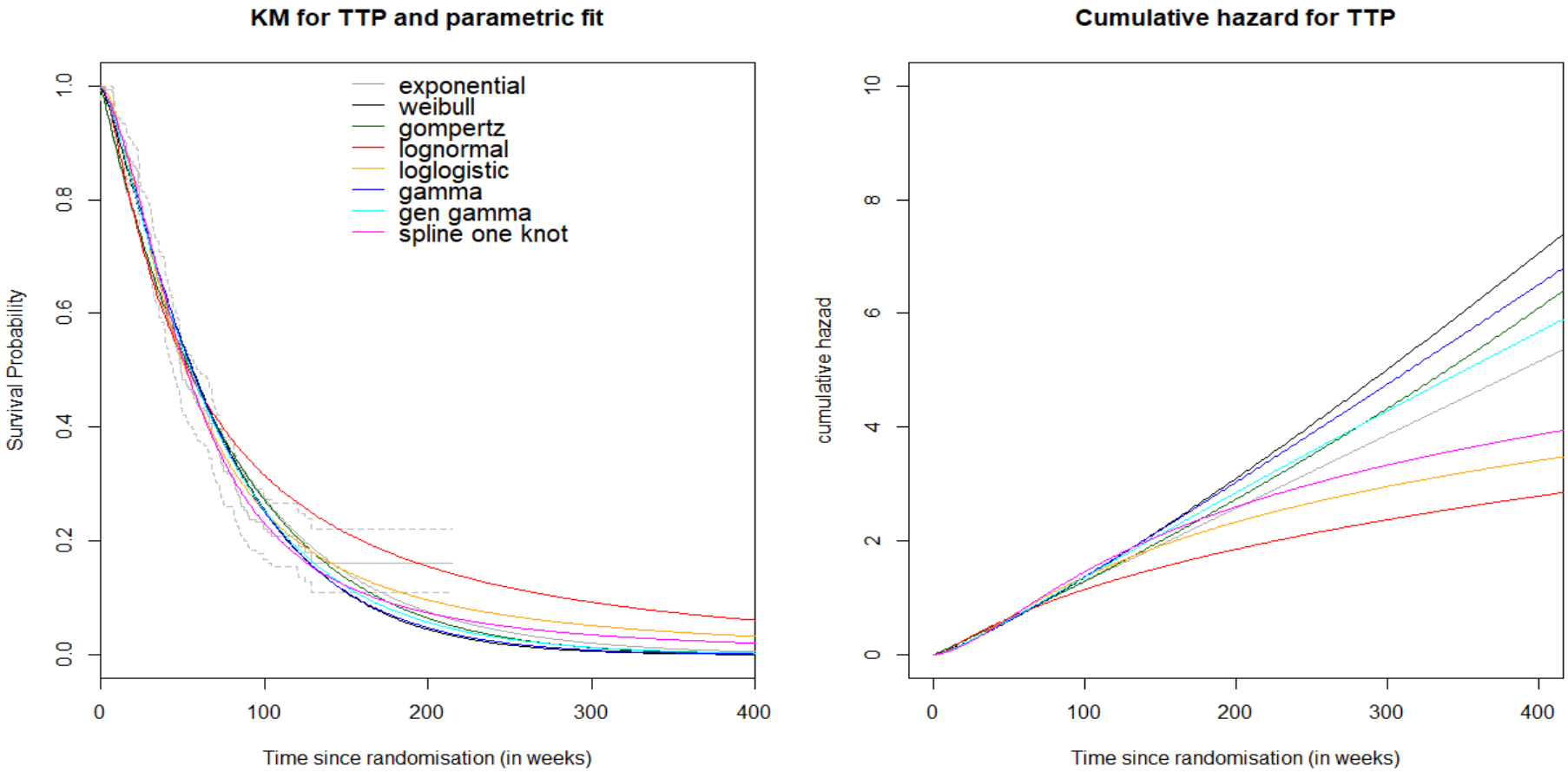


Figure 69 : Extrapolation for prePS in the BC dataset

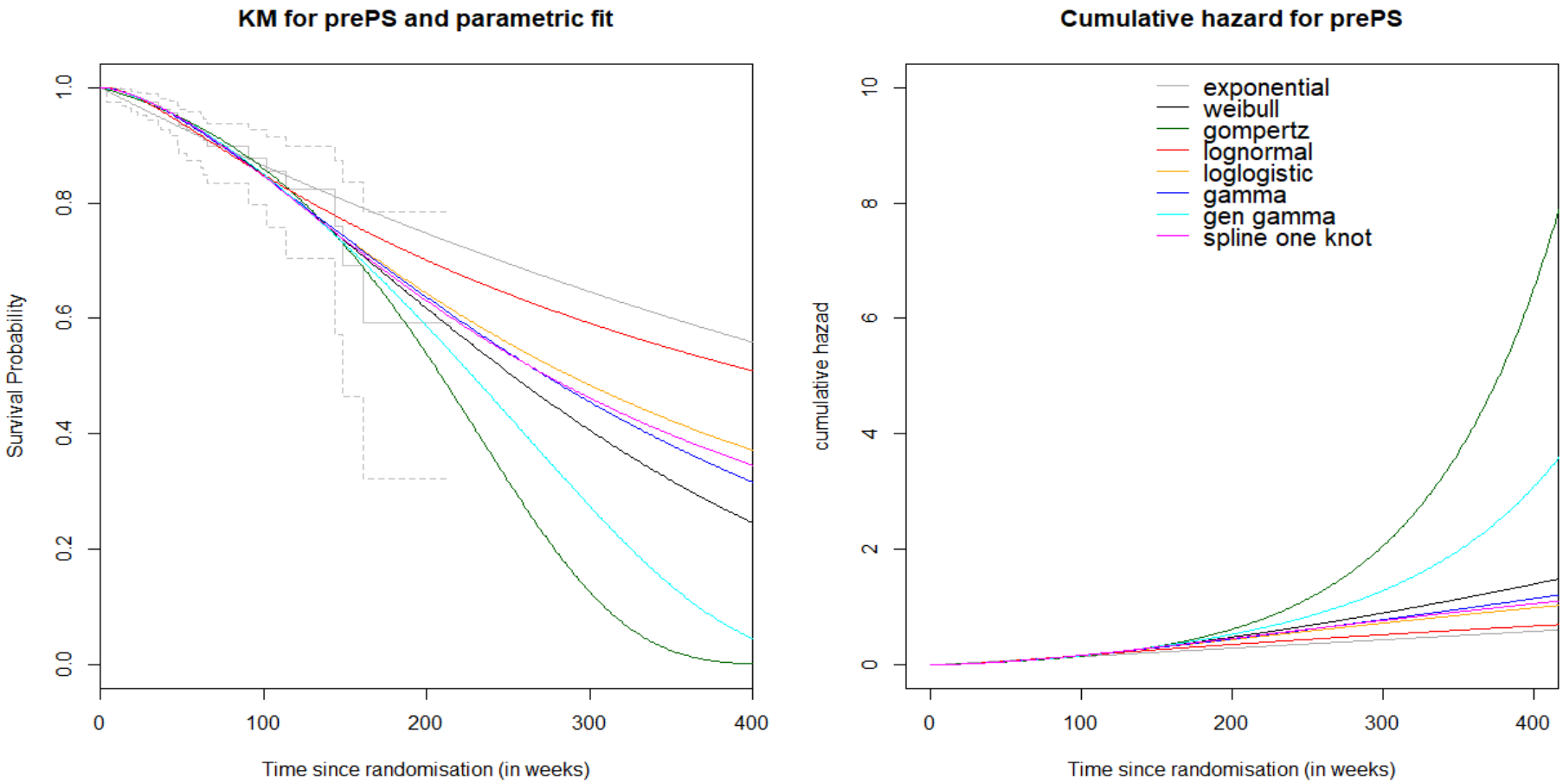
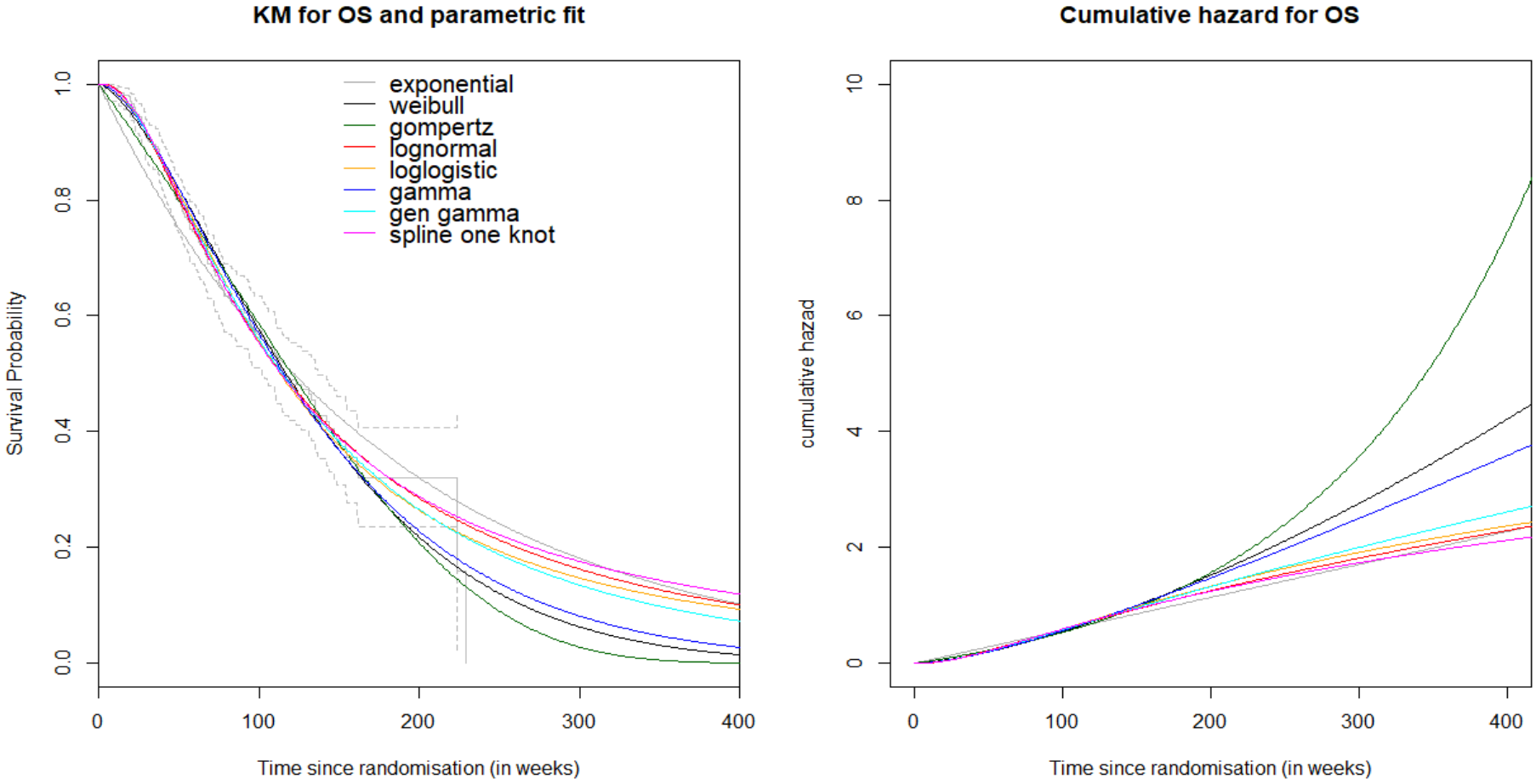


Figure 70 : Extrapolation for OS in the BC dataset



Appendix 2 : Key characteristics of appraisals included in the review

Key characteristics	TAs	
Review of NICE cancer TAs		
STAs	ID945; TA604; TA593; TA586; TA587 ; TA578 ; TA563 ; TA513 ; TA502 ; TA496 ; TA491 ; TA472 ; TA387 ; TA400 ; TA386 ; TA384 ; TA380 ; TA381 ; TA370 ;TA343 ; TA263 ; TA258 ; TA226 ; TA214 ; TA193 ; TA174	
MTAs	TA439; TA257; TA243	
Conditions	Follicular lymphoma (FL)	TA604; TA513; TA472; TA243; TA226
	Breast cancer (BC)	TA593; TA563; TA496; TA263 ; TA257 ; TA214
	Chronic lymphocytic lymphoma (CLL)	TA343 ; TA193 ; TA174
	Mantle cell lymphoma	TA502; TA370
	Myelofibrosis	TA386
	Multiple myeloma (MM)	TA586; TA587; TA380
	ovarian cancer (OC)	TA381 ;
	melanoma	TA400 ; TA384
	NSCLC	TA578; TA258
	Prostate Cancer (PC)	ID945; TA387
	Colorectal Cancer (CC)	TA439;
	Waldenstrom's Macroglobulinaemia (WM)	TA491
Setting	(Locally) advanced /metastatic	ID945; TA593; TA563; TA513; TA496 ; TA439 ; TA387 ; TA400 ; TA384 ; TA263 ; TA257 ; TA258 ; TA243 ; TA226 ; TA214
	Locally advanced only	TA578;
	Relapsed/refractory	TA604; TA586; TA502; TA491; TA472 ; TA380 ; TA381 ; TA193
	Mix pop/unclear	TA587; TA386 ; TA370; TA343 ; TA174
Justifications provided for choice of structure	Clear reference to OS immaturity	ID945; TA593; TA578; TA563; TA502 ; TA496 ; TA491 ; TA472 ; TA400 ; TA384 ; TA381 ; TA370 ; TA343 ; TA257
	Modelling of the pathway / natural history / use external evidence	ID945; TA593; TA563; TA496; TA491 ; TA439 ; TA472 (indolent) ; TA387 ; TA400 ; TA386 ; TA384 ; TA380 ; TA381 ; TA370 ; TA343 ; TA243 ; TA226
	Structural relationship PFS/OS – avoid crossing	TA586; TA587; TA578; TA563
	Structural relationship (patient experience in 2 phase for hazard of death)	TA257

	No clear justification for STM and/or reference to previous appraisals	TA604; TA214; TA193 ; TA174 ; TA513; TA263 ; TA258
Composite endpoint different to PFS	TTD	TA387; TA386
	FST	TA381 ;
Simulation approach (other than MSM?)		TA593/TA496; TA387; TA386 ; TA243
MSM approach		ID945; TA586; TA587
Analysis primarly based on trial for both arms)		TA604; TA593; TA586; TA587; TA578 ; TA513 (supplemented by additional trial for late PD) ; TA472; TA387; TA386; TA380 ; TA381 ; TA370 ; TA343 ; TA263 ; TA257 ; TA214 ; TA193 ; TA174
PFS + PPS from same trial (but comparator different to key trial)		TA502; TA258
PFS and PPS taken from different trials		ID945; TA563; TA496; TA491; TA439 ; TA400 ; TA384 ; TA243 ; TA226
Modelling of competing transitions		
PFS modelled as composite endpoint	Proportion of PFS events assumed to be death (explicit)	TA593; TA496; TA439 (implicit); TA386 ; TA380 (logistic) ; TA381 ; TA257 (implicit); TA243
	Death rate based on number of death and total PFS time	TA513; TA502; TA491; TA472; TA343 (implicit) ; TA263 ; TA258 ; TA226 ; TA214 ; TA193 ; TA174
	Death rate based on prePS	TA370
	Probability of progression based on TTP	TA578 (same distribution as for PFS)
Min time to death and time to progression (sampling time)		TA387
Two competing transitions modelled separately	TTP + parametric function for prePS	TA604; TA400 ; TA384
	TTP+negative binomial for prePS	TA563
Modelling of PPS		
PPS same double arm trial	Pooled (no sig diff)	TA593; TA578; TA513 (ERG comment); TA472 ; TA370 ; TA257 ; TA214 ; TA193 ; TA174 ; TA380 (pano - pathway)
	Different PPS between arms	TA586 (MSM); TA587(MSM); TA381 ; TA263 (KM+exp) ; TA257 ; TA386 (stopping rule)
PPS taken from one arm (single arm trial)		TA604; TA502; TA258 (different comparator to trial)
PPS taken from external source	PPS curve	ID945; TA496; TA491; TA439; TA400 ; TA384 ; TA343 ; TA243 (pathway) ; TA226
	Pay-off applied	TA563
PPS not really used (simulation)		TA387

Distribution used for PPS		
PPS – exponential only?	No clear justification	TA604; TA513; TA491 (unclear if other examined); TA214; TA193; TA174
	Avoid tunnel state/overcomplication	TA502; TA380 ; TA343 (implicit mention) ; TA258 ; TA226
	MSM	ID945; TA586; TA587
PPS – exponential in base case but alternative used		TA593; TA578; TA370
PPS - non exponential used in base-case		TA563(pay off) ; TA496 (external trial); TA439 (external trial); TA472; TA400; TA386 ; TA384 ; TA381 ; TA263(KM+exp) ; TA257 (exp in one model, non exp in another); TA243(pathway)
Modelling of late and early progressor separately		TA513; TA380
PPS (using age as a covariate)		TA343 (exponential adjusted by age)
Alternative structure explored		
	STM as base-case (but PSM explored in SA)	TA578
	STM used for one arm only	TA400 ; TA384
Other notes		TA604 (Several models using different data); TA593 – First line model same as TA496 TA439 – pathway modelled TA386- Stopping rule applied ; pathway modelled TA380 – pathway modelled TA381 – pathway modelled TA257 – 2 models built by AG

Appendix 3 : Parametric fit to the Prostate and Lung cancer datasets

Figure 71 : Parametric fit to the prostate cancer dataset

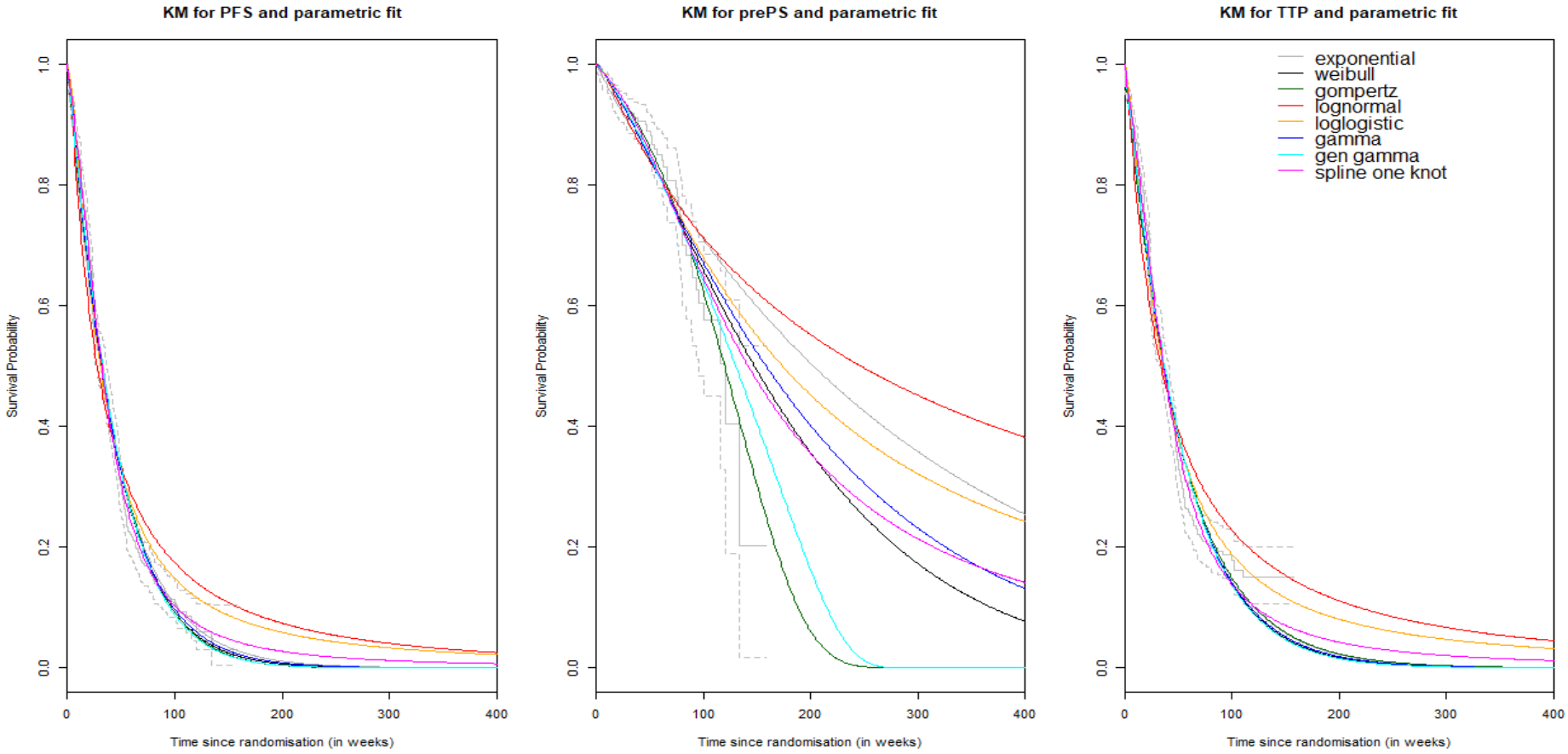
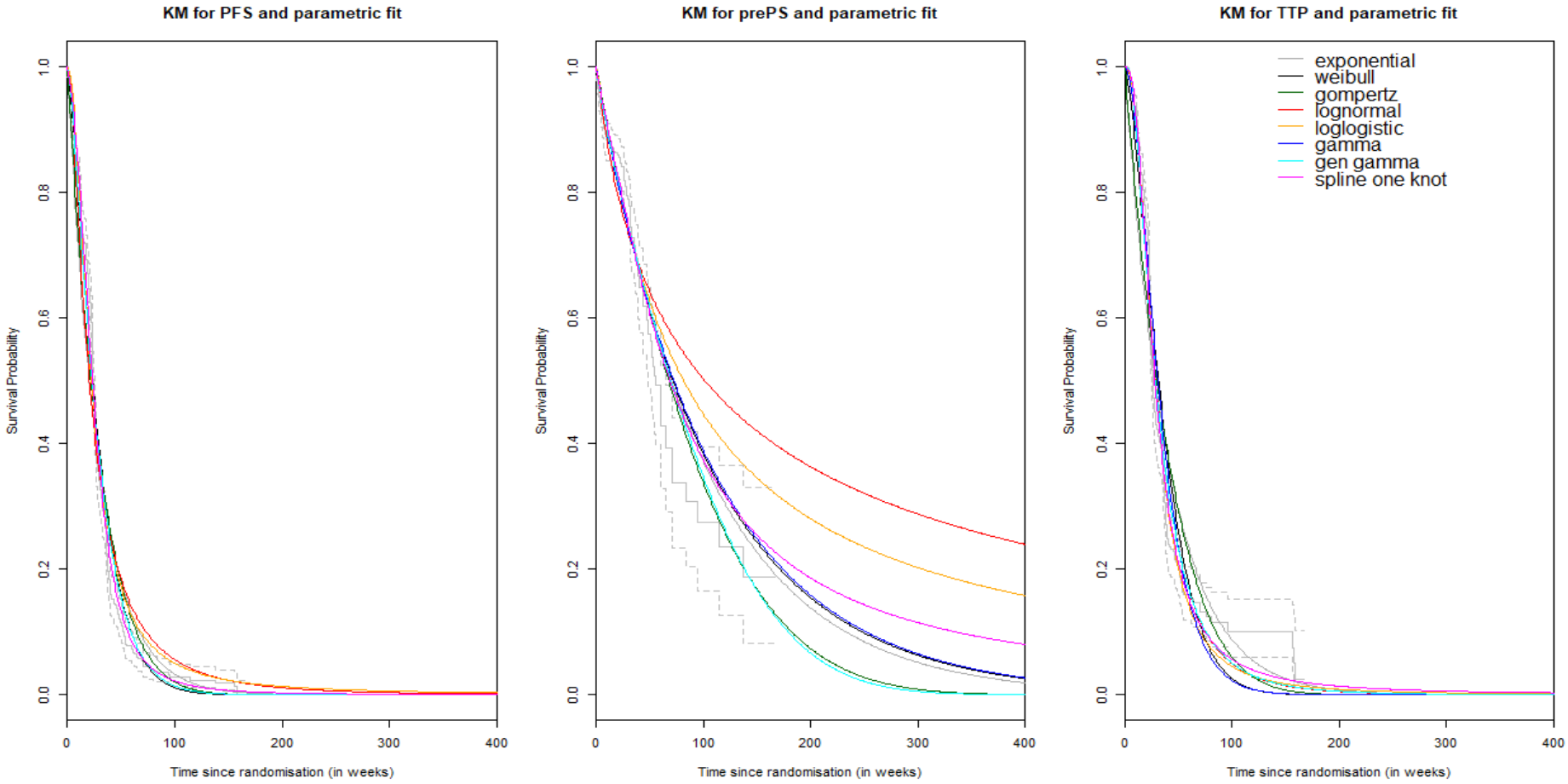


Figure 72 : Parametric fit to the Lung cancer dataset



Appendix 4 : Comparison of prediction between simplified STM vs. MSM

Table 20 : Comparison of prediction between simplified STM vs. MSM – Breast cancer dataset

PFS/ TTP	prePS	Simplified STM			MSM			Relative difference (%)		
		PF	PD	LY	PF	PD	LY	PF	PD	LY
exp	exp	70.2	68.7	138.9	71.2	68.5	139.6	1.41%	-0.41%	0.51%
exp	weib	70.2	66.6	136.7	68.4	66.7	135.1	-2.54%	0.29%	-1.16%
exp	gomp	70.2	63.5	133.7	67.6	67.1	134.7	-3.65%	5.64%	0.76%
exp	lnorm	70.2	68.3	138.4	69.5	68.2	137.7	-0.93%	-0.07%	-0.51%
exp	llogis	70.2	67.3	137.5	69.0	68.1	137.1	-1.68%	1.25%	-0.25%
exp	gam	70.2	67.0	137.2	67.8	66.9	134.6	-3.43%	-0.25%	-1.88%
exp	ggam	70.2	65.0	135.2	69.5	67.0	136.5	-1.04%	3.08%	0.94%
exp	spl	70.2	67.1	137.3	69.2	66.6	135.9	-1.34%	-0.69%	-1.02%
weib	exp	66.6	69.2	135.7	66.6	69.9	136.4	0.03%	1.01%	0.53%
weib	weib	66.6	68.0	134.6	65.7	68.6	134.4	-1.26%	0.88%	-0.18%
weib	gomp	66.6	66.9	133.4	65.6	67.8	133.3	-1.46%	1.35%	-0.05%
weib	lnorm	66.6	68.8	135.4	67.1	66.7	133.9	0.87%	-3.07%	-1.13%
weib	llogis	66.6	68.3	134.9	66.7	67.7	134.4	0.24%	-0.94%	-0.36%
weib	gam	66.6	68.2	134.8	66.7	69.4	136.1	0.20%	1.74%	0.98%
weib	ggam	66.6	67.6	134.1	66.1	68.9	134.9	-0.75%	1.97%	0.62%
weib	spl	66.6	68.2	134.8	66.9	68.4	135.2	0.50%	0.19%	0.34%
gomp	exp	67.2	69.1	136.3	68.2	69.0	137.2	1.43%	-0.06%	0.68%
gomp	weib	67.2	67.6	134.9	67.7	68.3	136.0	0.71%	0.95%	0.83%
gomp	gomp	67.2	66.1	133.3	66.7	67.4	134.1	-0.80%	2.07%	0.62%
gomp	lnorm	67.2	68.7	135.9	68.9	69.2	138.1	2.54%	0.70%	1.61%
gomp	llogis	67.2	68.0	135.2	66.8	68.6	135.4	-0.58%	0.78%	0.10%
gomp	gam	67.2	67.9	135.1	68.3	68.2	136.5	1.67%	0.40%	1.03%
gomp	ggam	67.2	67.0	134.2	66.8	67.3	134.1	-0.60%	0.40%	-0.10%
gomp	spl	67.2	67.9	135.1	67.0	66.5	133.5	-0.27%	-2.05%	-1.16%
lnorm	exp	101.4	65.1	166.5	93.6	65.5	159.1	-7.70%	0.58%	-4.46%
lnorm	weib	101.4	58.0	159.5	81.8	63.4	145.2	-19.31%	9.20%	-8.93%
lnorm	gomp	101.4	57.0	158.4	75.1	61.7	136.8	-25.98%	8.31%	-13.65%
lnorm	lnorm	101.4	64.5	166.0	92.9	65.9	158.8	-8.43%	2.14%	-4.32%
lnorm	llogis	101.4	61.4	162.9	87.6	64.6	152.1	-13.67%	5.10%	-6.59%
lnorm	gam	101.4	59.4	160.8	83.4	62.2	145.5	-17.82%	4.61%	-9.53%
lnorm	ggam	101.4	57.4	158.8	78.4	62.0	140.5	-22.68%	8.05%	-11.57%
lnorm	spl	101.4	60.7	162.1	84.2	64.1	148.3	-16.99%	5.57%	-8.54%
llogis	exp	84.8	67.0	151.7	78.6	67.7	146.2	-7.34%	1.08%	-3.62%

llogis	weib	84.8	63.0	147.8	72.4	65.3	137.6	-14.66%	3.57%	-6.89%
llogis	gomp	84.8	61.8	146.6	69.2	65.0	134.2	-18.38%	5.09%	-8.48%
llogis	lnorm	84.8	66.7	151.5	78.0	66.6	144.6	-8.03%	-0.16%	-4.57%
llogis	llogis	84.8	64.8	149.6	73.9	67.7	141.5	-12.89%	4.41%	-5.39%
llogis	gam	84.8	63.8	148.6	74.2	66.9	141.2	-12.44%	4.89%	-5.00%
llogis	ggam	84.8	62.4	147.2	70.7	65.4	136.1	-16.61%	4.76%	-7.55%
llogis	spl	84.8	64.4	149.2	73.8	66.1	140.0	-12.92%	2.69%	-6.18%
gam	exp	67.0	69.1	136.1	67.2	68.7	135.9	0.25%	-0.54%	-0.15%
gam	weib	67.0	67.9	134.9	65.8	69.5	135.3	-1.79%	2.34%	0.29%
gam	gomp	67.0	66.4	133.4	65.7	68.0	133.6	-1.98%	2.33%	0.16%
gam	lnorm	67.0	68.8	135.8	67.6	69.2	136.8	0.79%	0.67%	0.73%
gam	llogis	67.0	68.2	135.3	66.4	69.3	135.7	-0.87%	1.51%	0.33%
gam	gam	67.0	68.1	135.1	66.1	68.5	134.6	-1.39%	0.62%	-0.38%
gam	ggam	67.0	67.2	134.3	66.8	68.3	135.1	-0.37%	1.58%	0.60%
gam	spl	67.0	68.1	135.1	65.9	68.6	134.5	-1.70%	0.73%	-0.47%
ggam	exp	67.8	69.0	136.8	68.1	67.8	135.9	0.50%	-1.78%	-0.65%
ggam	weib	67.8	67.6	135.4	67.4	68.0	135.4	-0.52%	0.52%	0.00%
ggam	gomp	67.8	65.8	133.6	66.7	67.2	133.9	-1.59%	2.19%	0.27%
ggam	lnorm	67.8	68.7	136.5	68.2	68.0	136.2	0.62%	-0.95%	-0.17%
ggam	llogis	67.8	68.1	135.8	68.3	68.6	136.9	0.83%	0.76%	0.80%
ggam	gam	67.8	67.9	135.7	68.1	67.1	135.2	0.46%	-1.13%	-0.34%
ggam	ggam	67.8	66.8	134.5	67.2	68.5	135.7	-0.86%	2.56%	0.84%
ggam	spl	67.8	67.9	135.7	67.8	68.5	136.3	0.04%	0.91%	0.48%
spl	exp	75.2	68.1	143.3	71.7	67.9	139.6	-4.63%	-0.26%	-2.55%
spl	weib	75.2	65.3	140.5	67.8	66.4	134.2	-9.76%	1.61%	-4.48%
spl	gomp	75.2	64.0	139.1	66.4	66.9	133.3	-11.69%	4.66%	-4.18%
spl	lnorm	75.2	67.9	143.1	71.7	67.5	139.2	-4.59%	-0.56%	-2.68%
spl	llogis	75.2	66.7	141.8	71.2	68.6	139.8	-5.32%	2.91%	-1.45%
spl	gam	75.2	66.0	141.1	69.8	67.7	137.5	-7.21%	2.65%	-2.60%
spl	ggam	75.2	64.6	139.7	66.8	66.0	132.8	-11.12%	2.18%	-4.97%
spl	spl	75.2	66.4	141.6	69.9	67.6	137.4	-7.06%	1.73%	-2.94%

Abbreviations : exp : exponential ; weib : weibull ; gomp :gompertz ; lnorm : lognormal ; llogis : loglogistic ; gam : gamma ; ggam : generalised gamma ; MSM : multi-state model ; PFS : progression-free survival ; prePS : pre-progression mortality survival ; spl : spline hazard with one knot ; STM : state-transition model ; TTP : time to progression

Table 21 : Comparison of prediction between simplified STM vs. MSM – Gastric cancer dataset

PFS/ TTP	prePS	Simplified STM			MSM			Relative difference (%)		
		PF	PD	LY	PF	PD	LY	PF	PD	LY
exp	exp	44.9	60.0	104.9	45.0	60.7	105.7	0.15%	1.23%	0.76%
exp	weib	44.9	58.8	103.8	44.2	59.6	103.8	-1.74%	1.35%	0.01%
exp	gomp	44.9	55.4	100.3	43.1	58.6	101.8	-4.05%	5.91%	1.45%
exp	lnorm	44.9	60.5	105.5	45.4	60.2	105.7	1.11%	-0.46%	0.21%
exp	llogis	44.9	59.8	104.8	45.3	60.1	105.4	0.79%	0.50%	0.62%
exp	gam	44.9	59.3	104.2	44.5	59.9	104.4	-1.00%	1.09%	0.19%
exp	ggam	44.9	56.6	101.6	43.9	57.9	101.7	-2.42%	2.21%	0.16%
exp	spl	44.9	58.9	103.8	44.4	60.4	104.8	-1.21%	2.50%	0.89%
weib	exp	44.1	60.2	104.3	43.9	60.8	104.8	-0.38%	1.09%	0.47%
weib	weib	44.1	59.3	103.4	43.8	60.4	104.2	-0.66%	1.80%	0.75%
weib	gomp	44.1	56.7	100.8	43.4	57.8	101.2	-1.50%	1.83%	0.37%
weib	lnorm	44.1	60.6	104.7	44.9	60.7	105.6	1.83%	0.17%	0.87%
weib	llogis	44.1	60.1	104.2	44.4	59.8	104.2	0.61%	-0.48%	-0.02%
weib	gam	44.1	59.7	103.7	44.2	59.6	103.8	0.30%	-0.08%	0.08%
weib	ggam	44.1	57.7	101.8	43.7	58.6	102.3	-0.93%	1.58%	0.50%
weib	spl	44.1	59.4	103.5	43.5	59.2	102.7	-1.25%	-0.40%	-0.76%
gomp	exp	43.7	60.3	104.0	45.4	59.8	105.1	3.79%	-0.83%	1.11%
gomp	weib	43.7	59.5	103.2	44.2	59.5	103.7	1.13%	-0.05%	0.45%
gomp	gomp	43.7	57.2	100.9	43.2	57.6	100.8	-1.05%	0.73%	-0.04%
gomp	lnorm	43.7	60.7	104.4	45.3	60.3	105.6	3.57%	-0.59%	1.15%
gomp	llogis	43.7	60.2	103.9	45.0	59.6	104.5	2.94%	-1.05%	0.63%
gomp	gam	43.7	59.8	103.5	44.6	58.9	103.5	2.09%	-1.53%	0.00%
gomp	ggam	43.7	58.1	101.8	43.4	59.2	102.7	-0.60%	1.91%	0.83%
gomp	spl	43.7	59.5	103.2	44.4	59.4	103.8	1.64%	-0.24%	0.55%
lnorm	exp	67.9	54.3	122.2	56.3	58.2	114.4	-17.16%	7.17%	-6.35%
lnorm	weib	67.9	50.2	118.1	52.4	55.5	107.9	-22.85%	10.47%	-8.69%
lnorm	gomp	67.9	49.4	117.3	47.6	52.8	100.4	-29.95%	6.91%	-14.43%
lnorm	lnorm	67.9	57.0	124.9	61.9	56.5	118.3	-8.89%	-0.91%	-5.25%
lnorm	llogis	67.9	54.2	122.2	57.3	57.0	114.4	-15.58%	5.14%	-6.38%
lnorm	gam	67.9	51.4	119.3	55.6	55.6	111.2	-18.15%	8.20%	-6.80%
lnorm	ggam	67.9	50.2	118.1	48.5	54.9	103.4	-28.65%	9.30%	-12.52%
lnorm	spl	67.9	50.9	118.8	53.9	54.4	108.3	-20.65%	6.95%	-8.83%
llogis	exp	64.9	55.2	120.1	53.1	57.4	110.5	-18.23%	4.04%	-8.00%
llogis	weib	64.9	52.2	117.1	50.7	56.0	106.7	-21.93%	7.26%	-8.92%
llogis	gomp	64.9	51.2	116.1	45.2	56.1	101.4	-30.35%	9.73%	-12.68%

llogis	lnorm	64.9	57.3	122.2	57.3	59.5	116.8	-11.72%	3.85%	-4.41%
llogis	llogis	64.9	55.1	120.0	53.4	57.1	110.5	-17.72%	3.64%	-7.91%
llogis	gam	64.9	53.1	118.0	50.7	57.4	108.2	-21.81%	8.13%	-8.34%
llogis	ggam	64.9	51.9	116.8	47.1	55.1	102.1	-27.49%	6.13%	-12.55%
llogis	spl	64.9	52.5	117.4	50.8	55.9	106.7	-21.71%	6.54%	-9.09%
gam	exp	44.5	60.1	104.6	44.9	59.3	104.2	0.90%	-1.33%	-0.38%
gam	weib	44.5	59.1	103.6	44.5	58.9	103.4	0.05%	-0.43%	-0.22%
gam	gomp	44.5	56.2	100.7	43.4	58.3	101.7	-2.46%	3.70%	0.98%
gam	lnorm	44.5	60.6	105.0	45.6	60.8	106.4	2.60%	0.39%	1.32%
gam	llogis	44.5	60.0	104.5	44.5	59.1	103.5	0.00%	-1.54%	-0.88%
gam	gam	44.5	59.5	104.0	44.4	59.6	103.9	-0.23%	0.10%	-0.04%
gam	ggam	44.5	57.3	101.7	43.4	58.3	101.7	-2.30%	1.73%	-0.03%
gam	spl	44.5	59.2	103.7	44.3	59.9	104.2	-0.31%	1.14%	0.52%
ggam	exp	43.6	60.3	103.9	44.4	61.1	105.5	1.84%	1.28%	1.51%
ggam	weib	43.6	59.6	103.2	44.1	60.3	104.4	1.04%	1.30%	1.19%
ggam	gomp	43.6	57.4	101.0	43.2	58.1	101.3	-1.01%	1.21%	0.25%
ggam	lnorm	43.6	60.7	104.3	44.4	59.9	104.3	1.78%	-1.34%	-0.04%
ggam	llogis	43.6	60.2	103.9	43.9	59.5	103.4	0.58%	-1.20%	-0.45%
ggam	gam	43.6	59.8	103.5	43.9	60.2	104.1	0.55%	0.64%	0.60%
ggam	ggam	43.6	58.2	101.9	43.6	60.3	103.9	-0.07%	3.55%	2.00%
ggam	spl	43.6	59.6	103.2	44.2	59.4	103.6	1.24%	-0.36%	0.32%
spl	exp	50.3	58.7	108.9	45.9	59.6	105.6	-8.64%	1.69%	-3.08%
spl	weib	50.3	56.8	107.1	45.1	59.2	104.3	-10.36%	4.21%	-2.63%
spl	gomp	50.3	55.0	105.2	43.1	57.5	100.6	-14.21%	4.53%	-4.42%
spl	lnorm	50.3	59.7	110.0	47.9	60.3	108.2	-4.64%	1.02%	-1.57%
spl	llogis	50.3	58.7	109.0	46.5	59.7	106.2	-7.53%	1.81%	-2.50%
spl	gam	50.3	57.5	107.8	45.4	59.5	104.9	-9.68%	3.51%	-2.64%
spl	ggam	50.3	55.6	105.9	43.7	57.7	101.3	-13.14%	3.63%	-4.33%
spl	spl	50.3	57.4	107.7	45.4	58.6	103.9	-9.80%	2.09%	-3.46%

Abbreviations : exp : exponential ; weib : weibull ; gomp :gompertz ; lnorm : lognormal ; llogis : loglogistic ; gam : gamma ; ggam : generalised gamma ; MSM : multi-state model ; PFS : progression-free survival ; prePS : pre-progression mortality survival ; spl : spline hazard with one knot ; STM : state-transition model ; TTP : time to progression

Table 22 : Comparison of prediction between simplified STM vs. MSM – Lung cancer dataset

PFS/ TTP	prePS	Simplified STM			MSM			Relative difference (%)		
		PF	PD	LY	PF	PD	LY	PF	PD	LY
exp	exp	29.7	19.8	49.5	29.7	20.0	49.7	-0.06%	0.89%	0.32%
exp	weib	29.7	19.9	49.6	29.8	19.8	49.6	0.34%	-0.35%	0.06%
exp	gomp	29.7	19.6	49.3	29.6	19.5	49.1	-0.47%	-0.47%	-0.47%
exp	lnorm	29.7	20.6	50.3	30.9	20.4	51.3	3.99%	-0.90%	1.99%
exp	llogis	29.7	20.4	50.1	30.6	20.4	51.0	3.04%	-0.07%	1.77%
exp	gam	29.7	19.9	49.6	29.8	20.0	49.9	0.51%	0.69%	0.58%
exp	ggam	29.7	19.6	49.3	29.7	19.6	49.3	0.14%	-0.26%	-0.02%
exp	spl	29.7	20.0	49.6	29.9	20.0	49.9	0.83%	0.02%	0.50%
weib	exp	29.2	20.0	49.2	29.4	20.4	49.7	0.62%	1.91%	1.15%
weib	weib	29.2	20.0	49.2	29.5	20.0	49.5	0.94%	0.21%	0.64%
weib	gomp	29.2	20.0	49.2	29.5	20.4	49.9	1.14%	2.13%	1.54%
weib	lnorm	29.2	20.3	49.5	29.4	20.3	49.7	0.61%	-0.23%	0.26%
weib	llogis	29.2	20.3	49.5	29.5	20.3	49.8	1.11%	-0.19%	0.57%
weib	gam	29.2	20.0	49.2	29.0	20.0	49.0	-0.70%	0.12%	-0.37%
weib	ggam	29.2	20.0	49.2	29.3	20.2	49.5	0.20%	1.14%	0.58%
weib	spl	29.2	20.1	49.3	29.6	19.9	49.5	1.35%	-0.92%	0.43%
gomp	exp	29.3	19.9	49.2	29.2	20.2	49.5	-0.18%	1.52%	0.51%
gomp	weib	29.3	20.0	49.3	29.2	20.3	49.5	-0.29%	1.51%	0.44%
gomp	gomp	29.3	19.9	49.1	29.4	20.2	49.6	0.39%	1.65%	0.90%
gomp	lnorm	29.3	20.5	49.8	29.9	20.4	50.3	2.18%	-0.60%	1.03%
gomp	llogis	29.3	20.4	49.7	29.5	20.3	49.7	0.60%	-0.63%	0.09%
gomp	gam	29.3	20.0	49.3	29.1	20.1	49.2	-0.53%	0.32%	-0.19%
gomp	ggam	29.3	19.9	49.1	29.4	20.0	49.4	0.52%	0.55%	0.54%
gomp	spl	29.3	20.0	49.3	29.1	20.5	49.7	-0.56%	2.56%	0.71%
lnorm	exp	34.1	18.9	53.0	29.4	19.8	49.3	-13.55%	4.74%	-7.02%
lnorm	weib	34.1	19.1	53.1	29.4	19.8	49.2	-13.82%	3.85%	-7.47%
lnorm	gomp	34.1	18.4	52.5	29.3	19.8	49.2	-13.94%	7.69%	-6.34%
lnorm	lnorm	34.1	20.3	54.4	29.6	20.6	50.2	-13.13%	1.44%	-7.69%
lnorm	llogis	34.1	20.0	54.1	29.8	20.4	50.1	-12.65%	1.92%	-7.26%
lnorm	gam	34.1	19.2	53.2	29.1	19.8	48.9	-14.56%	3.59%	-8.03%
lnorm	ggam	34.1	18.6	52.6	29.5	20.0	49.5	-13.25%	7.38%	-5.97%
lnorm	spl	34.1	19.5	53.5	29.4	20.5	49.9	-13.83%	5.34%	-6.86%
llogis	exp	35.7	18.7	54.4	28.5	20.2	48.6	-20.19%	7.78%	-10.56%
llogis	weib	35.7	18.9	54.5	29.0	20.4	49.4	-18.73%	8.13%	-9.43%
llogis	gomp	35.7	18.4	54.1	28.8	20.7	49.5	-19.30%	12.33%	-8.53%

llogis	lnorm	35.7	20.1	55.8	29.7	20.7	50.4	-16.69%	2.91%	-9.62%
llogis	llogis	35.7	19.8	55.4	29.1	20.4	49.5	-18.50%	3.31%	-10.72%
llogis	gam	35.7	18.9	54.6	28.9	20.0	48.8	-19.05%	5.46%	-10.55%
llogis	ggam	35.7	18.6	54.3	28.4	20.3	48.7	-20.44%	9.23%	-10.27%
llogis	spl	35.7	19.3	54.9	28.6	20.4	49.0	-19.69%	5.70%	-10.78%
gam	exp	29.2	20.0	49.2	29.1	20.4	49.5	-0.27%	1.63%	0.50%
gam	weib	29.2	20.1	49.3	29.1	20.3	49.4	-0.27%	0.89%	0.20%
gam	gomp	29.2	20.0	49.2	28.7	20.2	48.9	-1.71%	0.99%	-0.61%
gam	lnorm	29.2	20.4	49.6	28.9	20.6	49.5	-0.95%	1.13%	-0.09%
gam	llogis	29.2	20.4	49.6	29.4	20.0	49.4	0.69%	-1.69%	-0.29%
gam	gam	29.2	20.1	49.3	29.0	20.2	49.2	-0.57%	0.32%	-0.21%
gam	ggam	29.2	20.1	49.3	29.0	20.3	49.2	-0.86%	0.84%	-0.17%
gam	spl	29.2	20.3	49.5	29.3	20.3	49.7	0.36%	0.38%	0.37%
ggam	exp	29.2	20.0	49.2	29.3	19.9	49.2	0.27%	-0.73%	-0.14%
ggam	weib	29.2	20.1	49.3	29.5	20.3	49.8	0.93%	0.94%	0.94%
ggam	gomp	29.2	20.0	49.2	29.4	20.4	49.8	0.73%	1.90%	1.20%
ggam	lnorm	29.2	20.4	49.6	30.0	20.5	50.5	2.61%	0.89%	1.91%
ggam	llogis	29.2	20.4	49.6	30.0	20.3	50.3	2.73%	-0.44%	1.42%
ggam	gam	29.2	20.1	49.3	29.7	20.2	49.9	1.77%	0.32%	1.18%
ggam	ggam	29.2	20.1	49.3	29.0	20.3	49.4	-0.60%	1.04%	0.07%
ggam	spl	29.2	20.3	49.5	29.2	19.9	49.1	-0.17%	-1.83%	-0.85%
spl	exp	30.1	19.7	49.8	29.4	19.9	49.4	-2.07%	0.93%	-0.88%
spl	weib	30.1	19.8	49.8	29.6	19.9	49.5	-1.68%	0.72%	-0.73%
spl	gomp	30.1	19.7	49.7	29.0	20.1	49.0	-3.64%	2.14%	-1.35%
spl	lnorm	30.1	20.3	50.4	30.3	20.6	50.9	0.79%	1.42%	1.05%
spl	llogis	30.1	20.2	50.3	29.9	20.2	50.1	-0.38%	0.02%	-0.22%
spl	gam	30.1	19.8	49.9	28.9	20.3	49.3	-3.80%	2.74%	-1.20%
spl	ggam	30.1	19.7	49.7	29.1	20.0	49.1	-3.18%	1.46%	-1.34%
spl	spl	30.1	20.0	50.1	30.2	19.9	50.1	0.51%	-0.44%	0.13%

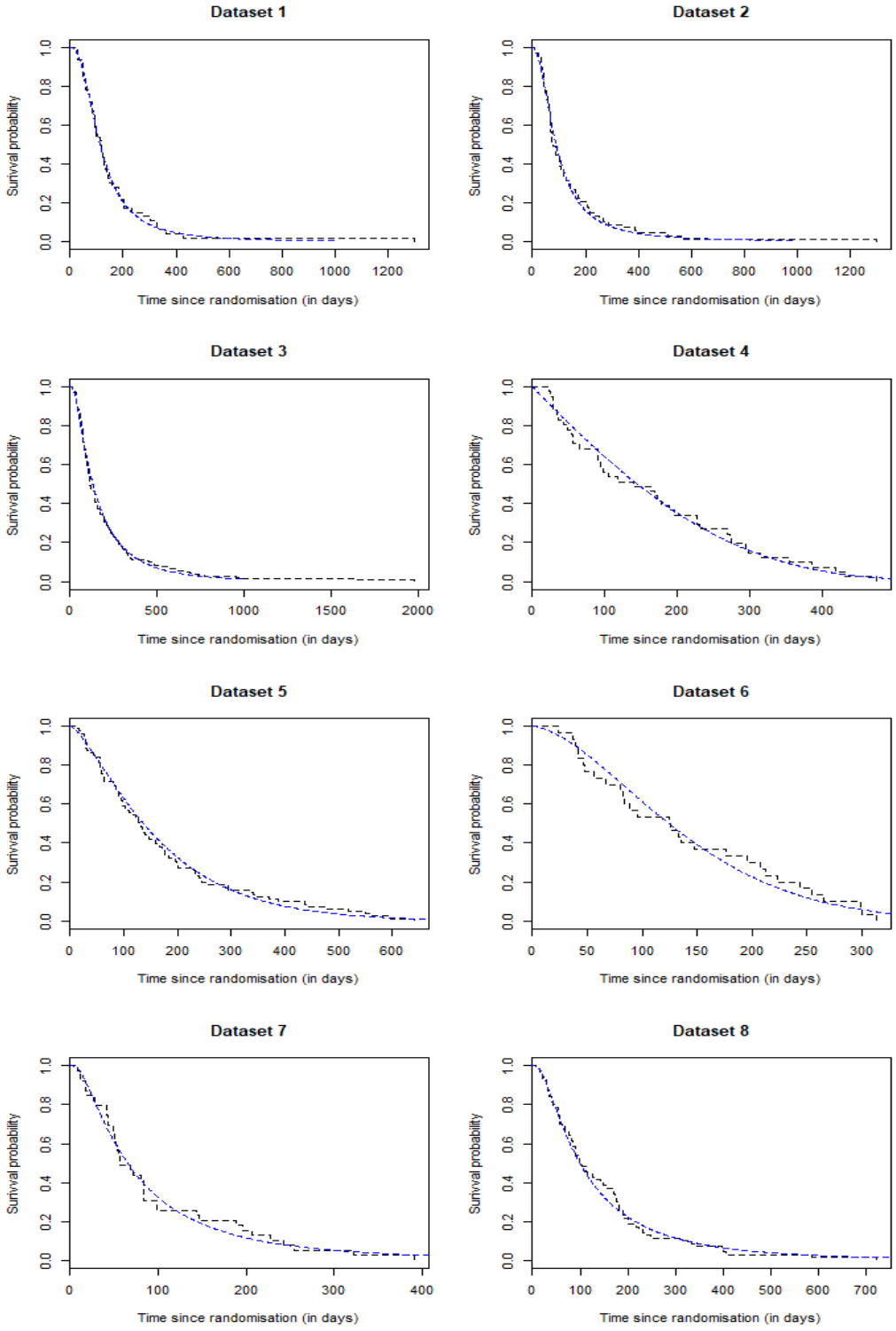
Abbreviations : exp : exponential ; weib : weibull ; gomp :gompertz ; lnorm : lognormal ; llogis : loglogistic ; gam : gamma ; ggam : generalised gamma ; MSM : multi-state model ; PFS : progression-free survival ; prePS : pre-progression mortality survival ; spl : spline hazard with one knot ; STM : state-transition model ; TTP : time to progression

Appendix 5 : Papers included in the systematic review

1. Belkacemi, M. C., et al. (2014). "Modelling of overall survival by an association between progression-free and post-progression survival using a conditional distribution." Statistical Modelling 14(1): 77-98.
2. Dejardin, D., et al. (2010). "Joint modeling of progression-free survival and death in advanced cancer clinical trials." Statistics in Medicine 29(16): 1724-1734.
3. Fleischer, F., et al. (2009). "A statistical model for the dependence between progression-free survival and overall survival." Statistics in Medicine 28(21): 2669-2686.
4. Fu, H. D., et al. (2013). "Joint modeling of progression-free survival and overall survival by a Bayesian normal induced copula estimation model." Statistics in Medicine 32(2): 240-254.
5. Krol, A., et al. (2017). "Tutorial in Joint Modeling and Prediction: A Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event." Journal of Statistical Software 81(3): 1-52.
6. Lia, Y. M. and Q. Zhang (2015). "A Weibull multi-state model for the dependence of progression-free survival and overall survival." Statistics in Medicine 34(17): 2497-2513.
7. Mazroui, Y., et al. (2012). "General joint frailty model for recurrent event data with a dependent terminal event: Application to follicular lymphoma data." Statistics in Medicine 31(11-12): 1162-1176.
8. Meller, M., et al. (2019). "Joint modeling of progression-free and overall survival and computation of correlation measures." Statistics in Medicine 38(22): 4270-4289.
9. Oakes, D. (1982). "A MODEL FOR ASSOCIATION IN BIVARIATE SURVIVAL-DATA." Journal of the Royal Statistical Society Series B-Methodological 44(3): 414-422.
10. Rondeau, V., et al. (2007). "Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events." Biostatistics 8(4): 708-721.
11. Rondeau, V., et al. (2012). "frailtypack: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation." Journal of Statistical Software 47(4): 1-28.
12. Sildnes, B. and B. H. Lindqvist (2018). "Modeling of semi-competing risks by means of first passage times of a stochastic process." Lifetime Data Analysis 24(1): 153-175.
13. Weber, E. M. and A. C. Titman (2019). "Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's tau." Statistics in Medicine 38(5): 703-719.

Appendix 6 : PPS Chapter

Figure 73 : Selected fit to PFS for each dataset



Appendix 7 : Review for key trials characteristics

Key characteristics of trial data that are typically encountered within the HTA context

Prior to setting out the design of the simulation study, it is important to understand the characteristics of trial data that are typically encountered within the HTA context in order to define plausible and realistic scenarios when evaluating the performance of the methods under investigation. This section presents the findings from a rapid review of NICE TAs of therapies for advanced/metastatic cancer.

Objective of the review

The primary aim of this rapid review is to identify the key characteristics of trial data which are observable and which may be deemed to influence the performance of methods that are typically encountered within the HTA context, in terms of: (i) the number of patients enrolled/randomised; (ii) the proportion of observed PFS events amongst enrolled/randomised patients which are either progression or death, and; (iii) the proportion of observed deaths amongst enrolled/randomised patients. The aim of this rapid review is to obtain a broad picture of the key characteristics of trials that are typically encountered within the HTA context in terms of the number of observed PFS and OS events to inform the data-generating mechanism.

- ***Search and selection strategy***

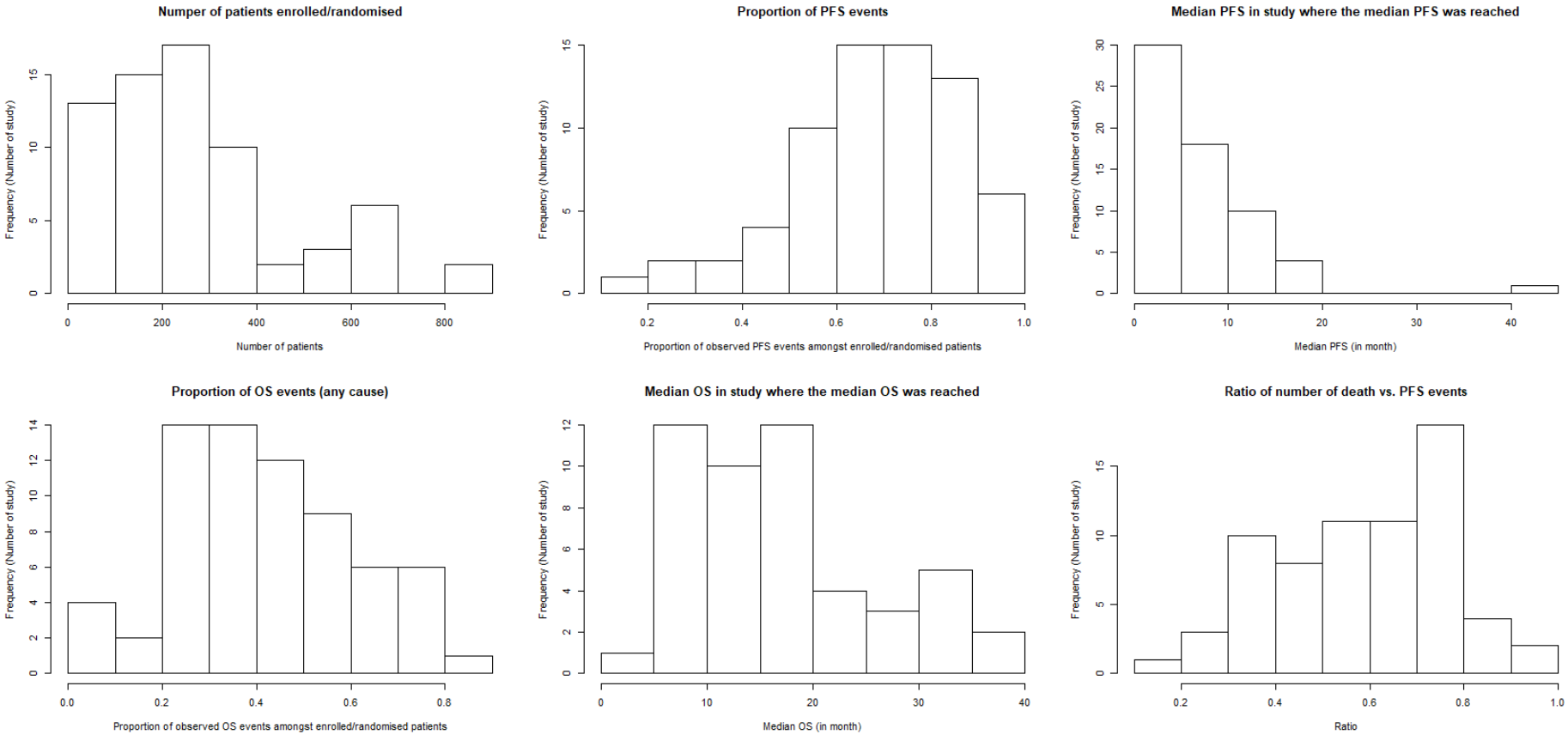
The review focused on 30 NICE TAs of therapies for advanced/metastatic cancer. This review is limited to NICE appraisals to: (a) be reflective of trials encountered within the HTA context, and; (b) information which is publicly available. In addition, this rapid review is limited to key trials included in the company's submission for the intervention, in which full details are available on the number of events and censoring. In order to keep the review manageable, summary information relating to trials included in indirect comparisons/meta-analyses undertaken by the company are not considered.

A number of HTAs reported trial data from different data-cuts. Where possible, data from the most recent data-cut were extracted. Data from earlier data cut-offs were extracted when data on the number of observed PFS and OS events were not reported in the most up-to-date data-cut. In addition, a number of HTAs reported data from several trials for the intervention; these were extracted, where possible.

- ***Findings from the rapid review***

Key study characteristics are summarised in Figure 74.

Figure 74: Key characteristics of the 68 study arms included in the review



Abbreviations: OS: overall survival; PFS: progression-free survival

Thirty-five trials were included in the 30 HTAs included in this review, from which data were extracted from 25 two-arm trials, 4 three-arm trials, with the remaining 6 trials being single trial arms, corresponding to a total of 68-trial study arms. The 30 HTAs included in this review are reported in Appendix 7.

In summary:

- The mean number of patients enrolled/randomised per arm was 276.1 (median: 240; range: 42-872).
- Median PFS was not reached in 5 trial arms (median reached in 63 trial arms). In trial arms where the median PFS was reached, the mean PFS amongst enrolled/randomised patients was 7.16 months (median: 5.20; range: 0.9-42.20).
- Median OS was not reached in 19 trial arms (reached in 49 trial arms). In trial arms where the median OS was reached, mean OS amongst enrolled/randomised patients was 17.12 months (median: 17.30; range: 3.80-35.50).
- The mean proportion of observed PFS events amongst enrolled/randomised patients, where PFS is defined as either progression or death before progression was 68.34% (median: 70.78%, range: 16.81% - 98.51%).
- The mean proportion of observed OS events amongst enrolled/randomised patients was 42.05% (median: 40.23%, range: 5.82% - 84.62%).
- The ratio between the total number of deaths and PFS events (e.g. OS events divided by PFS events) was 59.15% (median: 61.91%, range: 13.33% - 91.76%).

These findings are used to inform the scenarios examined in the simulation study.

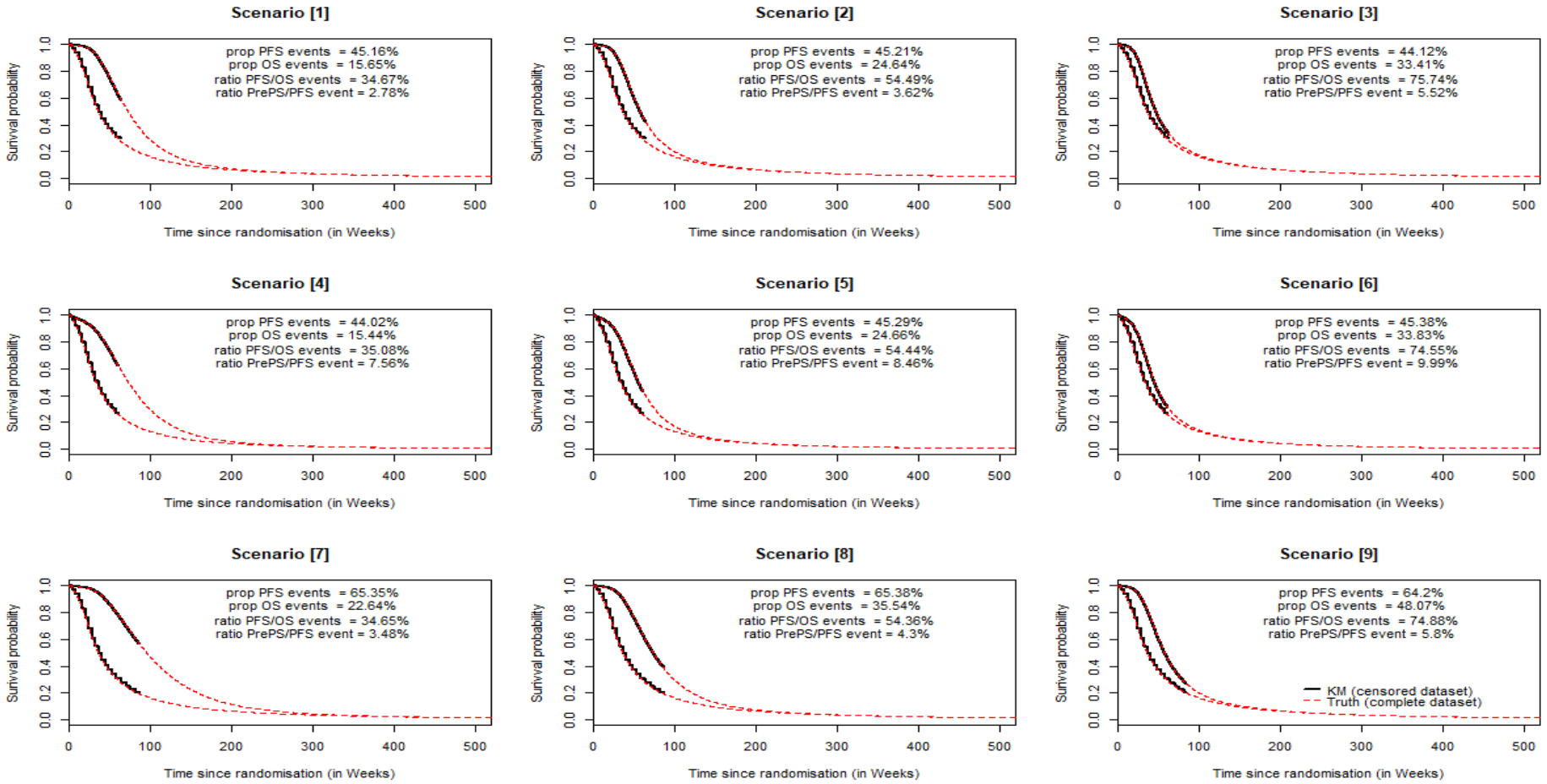
Appendix 8 : AIC for model fit to TTP in the breast, prostate and lung datasets

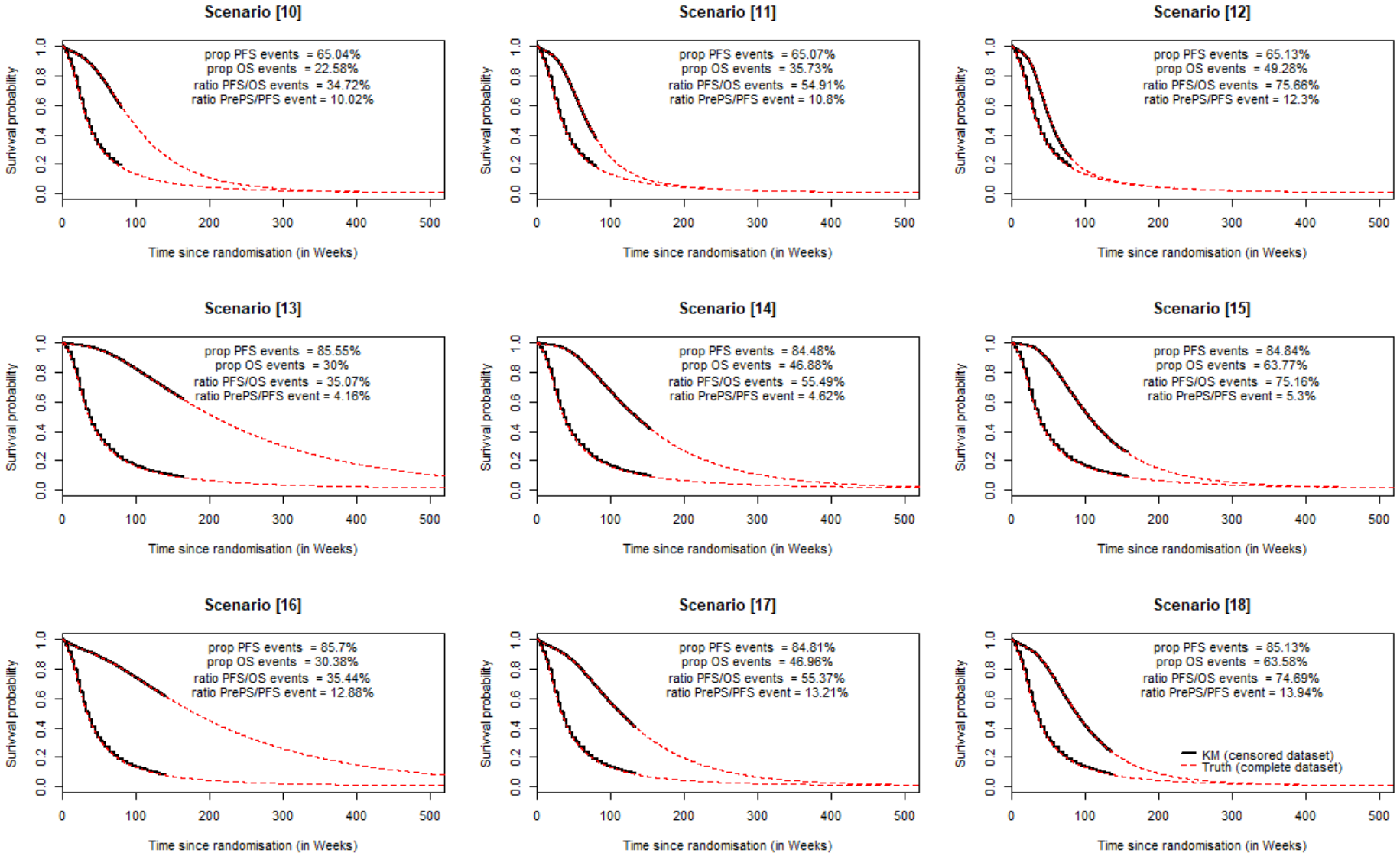
	Breast	Prostate	Lung
Exponential	2,088.8	3,374.3	2,953.1
Weibull	2,082.8	3,375.1	2,888.6
Gompertz	2,090.5	3,376.3	2,948.7
LogNormal	2,139.9	3,456.2	2,825.1
LogLogistic	2,070.8	3,385.5	2,808.2
Gamma	2,080.4	3,375.3	2,856.6
GenGamma	2,081.4	3,377.0	2,827.1
Spline (one knot)	2,060.3	3,348.0	2,808.8
Spline (two knot)	2,058.9	3,347.4	2,791.2
Spline (three knot)	2,060.9	3,346.7	2,793.8
Spline (four knot)	2,062.7	3,344.8	2,767.7

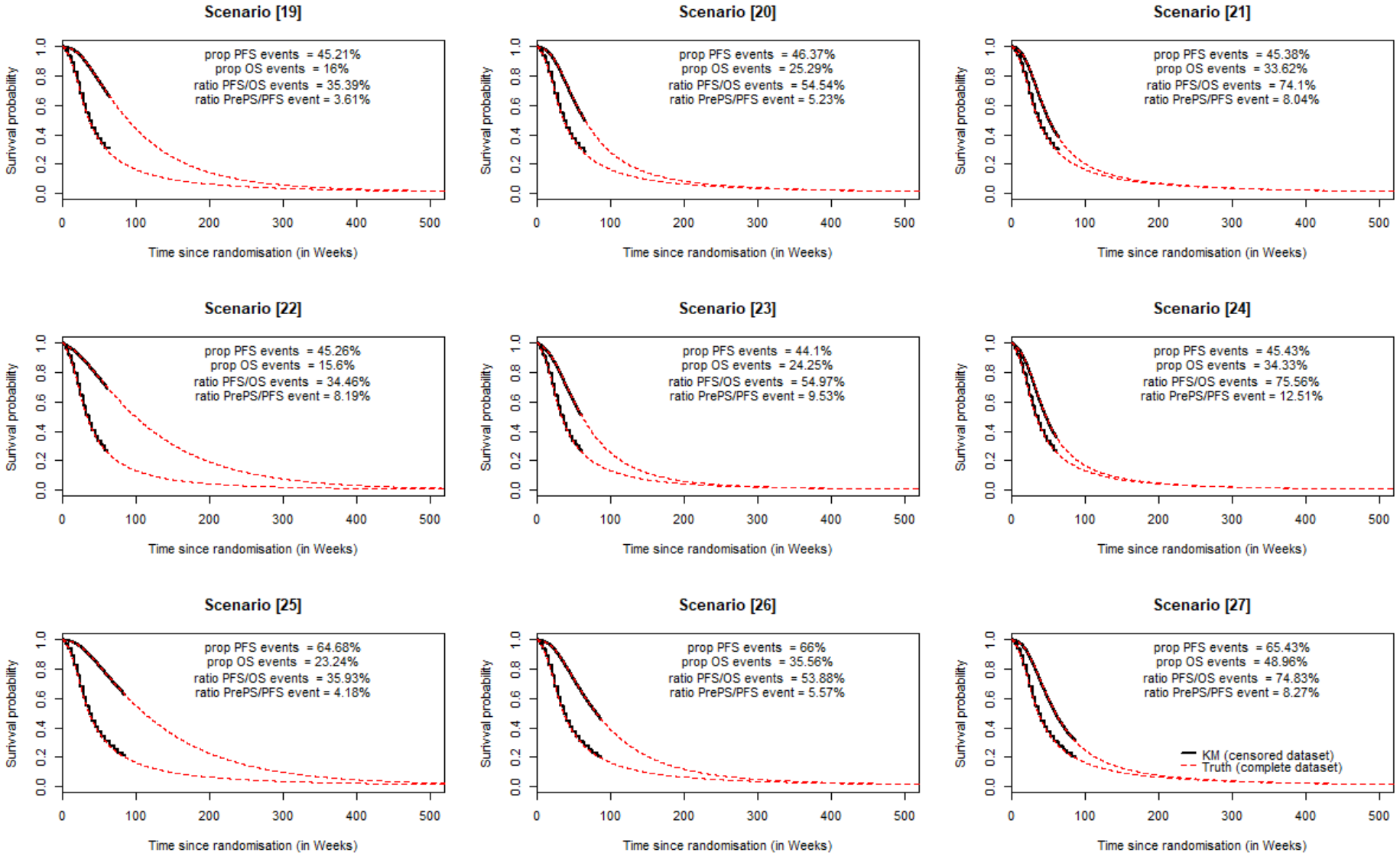
Appendix 9 : Calibrated parameters

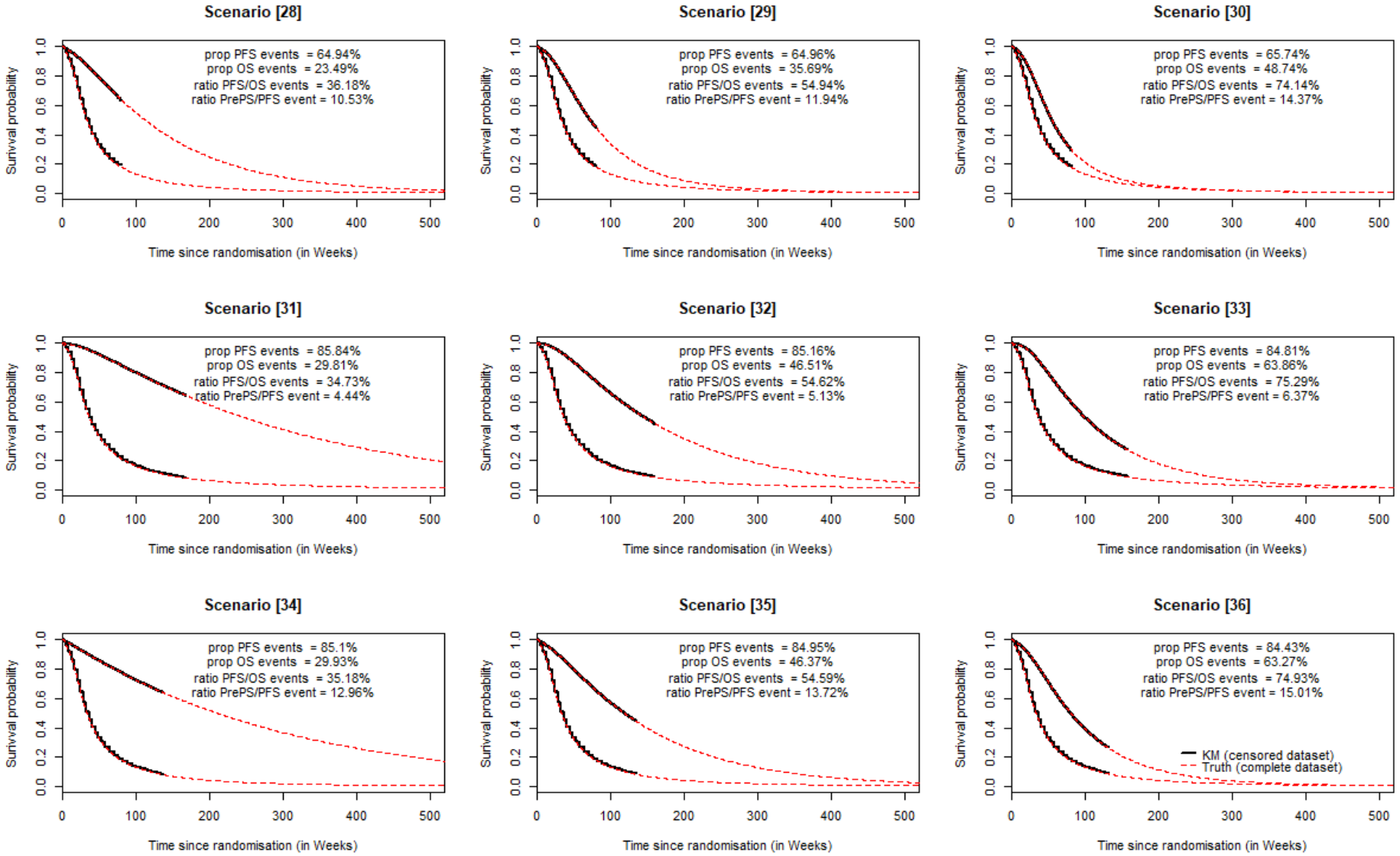
Scenario	adm.coeff	pps.coeff		Scenario	adm.coeff	pps.coeff
1	12.478035	-3.4235992		28	27.520411	-4.6928649
2	11.683833	-2.8708669		29	27.765966	-3.8389242
3	11.011459	-2.2365854		30	29.340254	-3.1454755
4	7.6568535	-3.6132046		31	114.66164	-5.6143098
5	9.1357667	-2.9604047		32	108.11197	-4.8619237
6	8.7772735	-2.3345744		33	105.24018	-4.2160072
7	33.99621	-3.9937388		34	83.774822	-5.6664283
8	33.542072	-3.4539686		35	83.16378	-4.7869728
9	32.235189	-2.8724304		36	79.682677	-4.0554359
10	27.878214	-4.1185211		37	11.876984	-4.6539933
11	28.246922	-3.4263333		38	12.468284	-3.7227728
12	28.066599	-2.8248406		39	11.520266	-2.980176
13	112.13568	-5.2723776		40	9.4177451	-5.1906546
14	102.3145	-4.6048834		41	8.0741347	-4.0171942
15	104.74938	-4.1732336		42	8.7592819	-3.1272833
16	88.350968	-5.2506728		43	34.095796	-4.9367239
17	82.139651	-4.4866779		44	34.218473	-4.1399848
18	83.566896	-4.007175		45	32.620622	-3.3111407
19	12.061512	-4.0212744		46	26.533994	-5.2350521
20	13.384992	-3.3275497		47	26.730997	-4.2340087
21	11.721155	-2.6146696		48	29.020159	-3.4263114
22	8.7270917	-4.467706		49	109.11841	-5.8010987
23	7.9006853	-3.4345884		50	102.25587	-5.0071826
24	8.5001327	-2.6346434		51	110.799	-4.4155695
25	32.503319	-4.4275459		52	87.368408	-6.0459865
26	34.650467	-3.8218342		53	79.572474	-5.0450773
27	33.775418	-3.1084309		54	85.960616	-4.3419768

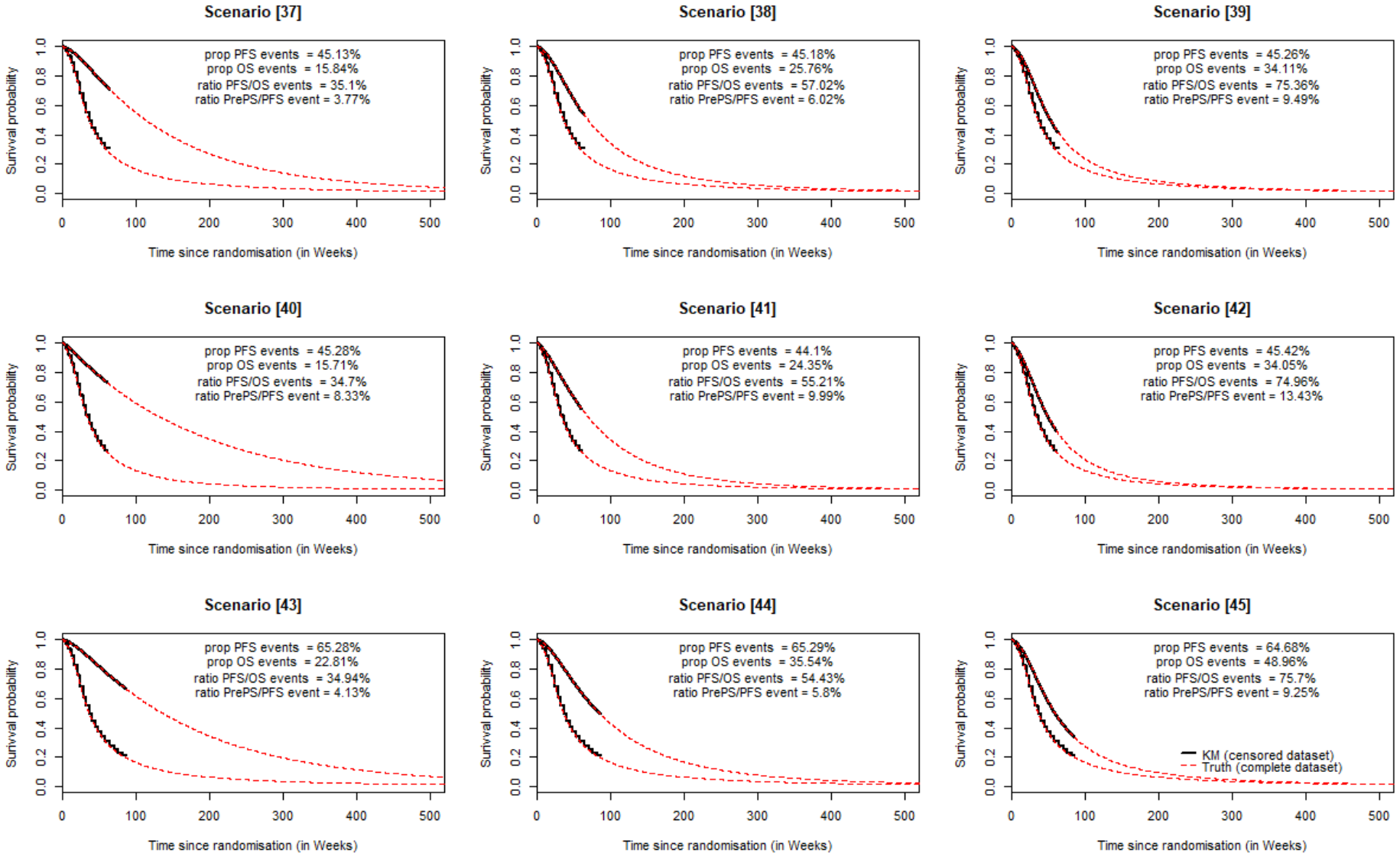
Appendix 10 : Simulated PFS and OS for the 54 Scenarios used in the Simulation study

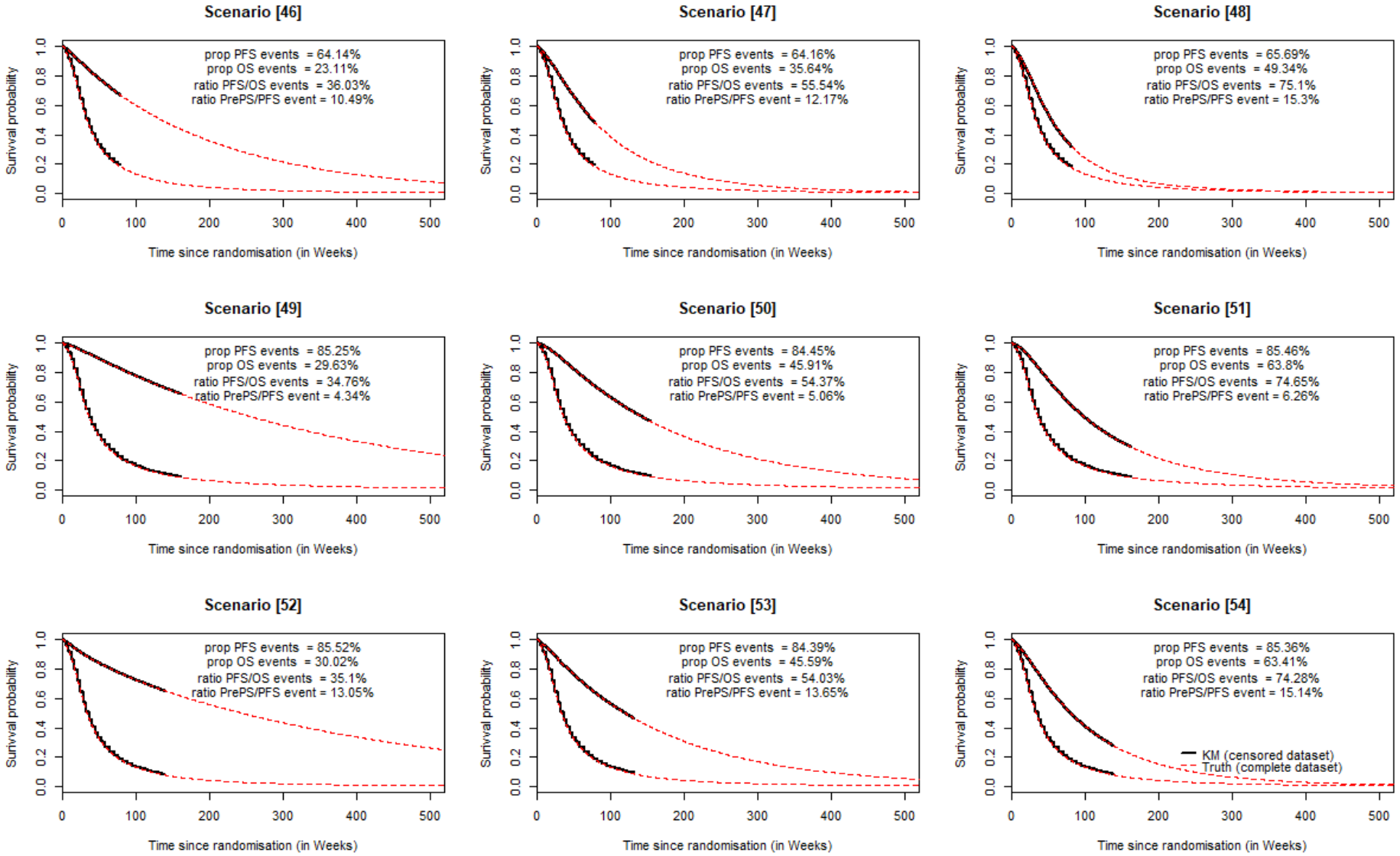



Scenario [10]
prop PFS events = 65.04%
prop OS events = 22.58%
ratio PFS/OS events = 34.72%
ratio PrePS/PFS event = 10.02%
Scenario [11]
prop PFS events = 65.07%
prop OS events = 35.73%
ratio PFS/OS events = 54.91%
ratio PrePS/PFS event = 10.8%
Scenario [12]
prop PFS events = 65.13%
prop OS events = 49.28%
ratio PFS/OS events = 75.66%
ratio PrePS/PFS event = 12.3%
Scenario [13]
prop PFS events = 85.55%
prop OS events = 30%
ratio PFS/OS events = 35.07%
ratio PrePS/PFS event = 4.16%
Scenario [14]
prop PFS events = 84.48%
prop OS events = 46.88%
ratio PFS/OS events = 55.49%
ratio PrePS/PFS event = 4.62%
Scenario [15]
prop PFS events = 84.84%
prop OS events = 63.77%
ratio PFS/OS events = 75.16%
ratio PrePS/PFS event = 5.3%
Scenario [16]
prop PFS events = 85.7%
prop OS events = 30.38%
ratio PFS/OS events = 35.44%
ratio PrePS/PFS event = 12.88%
Scenario [17]
prop PFS events = 84.81%
prop OS events = 46.96%
ratio PFS/OS events = 55.37%
ratio PrePS/PFS event = 13.21%
Scenario [18]
prop PFS events = 85.13%
prop OS events = 63.58%
ratio PFS/OS events = 74.69%
ratio PrePS/PFS event = 13.94%
KM (censored dataset)
Truth (complete dataset)
Survival probability
Time since randomisation (in Weeks)


Scenario [19]
prop PFS events = 45.21%
prop OS events = 16%
ratio PFS/OS events = 35.39%
ratio PrePS/PFS event = 3.61%
Scenario [20]
prop PFS events = 46.37%
prop OS events = 25.29%
ratio PFS/OS events = 54.54%
ratio PrePS/PFS event = 5.23%
Scenario [21]
prop PFS events = 45.38%
prop OS events = 33.62%
ratio PFS/OS events = 74.1%
ratio PrePS/PFS event = 8.04%
Scenario [22]
prop PFS events = 45.26%
prop OS events = 15.6%
ratio PFS/OS events = 34.46%
ratio PrePS/PFS event = 8.19%
Scenario [23]
prop PFS events = 44.1%
prop OS events = 24.25%
ratio PFS/OS events = 54.97%
ratio PrePS/PFS event = 9.53%
Scenario [24]
prop PFS events = 45.43%
prop OS events = 34.33%
ratio PFS/OS events = 75.56%
ratio PrePS/PFS event = 12.51%
Scenario [25]
prop PFS events = 64.68%
prop OS events = 23.24%
ratio PFS/OS events = 35.93%
ratio PrePS/PFS event = 4.18%
Scenario [26]
prop PFS events = 66%
prop OS events = 35.56%
ratio PFS/OS events = 53.88%
ratio PrePS/PFS event = 5.57%
Scenario [27]
prop PFS events = 65.43%
prop OS events = 48.96%
ratio PFS/OS events = 74.83%
ratio PrePS/PFS event = 8.27%
KM (censored dataset)
Truth (complete dataset)
Survival probability
0.0 0.2 0.4 0.6 0.8 1.0
0 100 200 300 400 500
Time since randomisation (in Weeks)


Scenario [28]
prop PFS events = 64.94%
prop OS events = 23.49%
ratio PFS/OS events = 36.18%
ratio PrePS/PFS event = 10.53%
Scenario [29]
prop PFS events = 64.96%
prop OS events = 35.69%
ratio PFS/OS events = 54.94%
ratio PrePS/PFS event = 11.94%
Scenario [30]
prop PFS events = 65.74%
prop OS events = 48.74%
ratio PFS/OS events = 74.14%
ratio PrePS/PFS event = 14.37%
Scenario [31]
prop PFS events = 85.84%
prop OS events = 29.81%
ratio PFS/OS events = 34.73%
ratio PrePS/PFS event = 4.44%
Scenario [32]
prop PFS events = 85.16%
prop OS events = 46.51%
ratio PFS/OS events = 54.62%
ratio PrePS/PFS event = 5.13%
Scenario [33]
prop PFS events = 84.81%
prop OS events = 63.86%
ratio PFS/OS events = 75.29%
ratio PrePS/PFS event = 6.37%
Scenario [34]
prop PFS events = 85.1%
prop OS events = 29.93%
ratio PFS/OS events = 35.18%
ratio PrePS/PFS event = 12.96%
Scenario [35]
prop PFS events = 84.95%
prop OS events = 46.37%
ratio PFS/OS events = 54.59%
ratio PrePS/PFS event = 13.72%
Scenario [36]
prop PFS events = 84.43%
prop OS events = 63.27%
ratio PFS/OS events = 74.93%
ratio PrePS/PFS event = 15.01%
KM (censored dataset)
Truth (complete dataset)
Survival probability
0.0 0.2 0.4 0.6 0.8 1.0
0 100 200 300 400 500
Time since randomisation (in Weeks)


Scenario [37]
prop PFS events = 45.13%
prop OS events = 15.84%
ratio PFS/OS events = 35.1%
ratio PrePS/PFS event = 3.77%
Scenario [38]
prop PFS events = 45.18%
prop OS events = 25.76%
ratio PFS/OS events = 57.02%
ratio PrePS/PFS event = 6.02%
Scenario [39]
prop PFS events = 45.26%
prop OS events = 34.11%
ratio PFS/OS events = 75.36%
ratio PrePS/PFS event = 9.49%
Scenario [40]
prop PFS events = 45.28%
prop OS events = 15.71%
ratio PFS/OS events = 34.7%
ratio PrePS/PFS event = 8.33%
Scenario [41]
prop PFS events = 44.1%
prop OS events = 24.35%
ratio PFS/OS events = 55.21%
ratio PrePS/PFS event = 9.99%
Scenario [42]
prop PFS events = 45.42%
prop OS events = 34.05%
ratio PFS/OS events = 74.96%
ratio PrePS/PFS event = 13.43%
Scenario [43]
prop PFS events = 65.28%
prop OS events = 22.81%
ratio PFS/OS events = 34.94%
ratio PrePS/PFS event = 4.13%
Scenario [44]
prop PFS events = 65.29%
prop OS events = 35.54%
ratio PFS/OS events = 54.43%
ratio PrePS/PFS event = 5.8%
Scenario [45]
prop PFS events = 64.68%
prop OS events = 48.96%
ratio PFS/OS events = 75.7%
ratio PrePS/PFS event = 9.25%
KM (censored dataset)
Truth (complete dataset)
Survival probability
Time since randomisation (in Weeks)


Scenario [46]
prop PFS events = 64.14%
prop OS events = 23.11%
ratio PFS/OS events = 36.03%
ratio PrePS/PFS event = 10.49%
Scenario [47]
prop PFS events = 64.16%
prop OS events = 35.64%
ratio PFS/OS events = 55.54%
ratio PrePS/PFS event = 12.17%
Scenario [48]
prop PFS events = 65.69%
prop OS events = 49.34%
ratio PFS/OS events = 75.1%
ratio PrePS/PFS event = 15.3%
Scenario [49]
prop PFS events = 85.25%
prop OS events = 29.63%
ratio PFS/OS events = 34.76%
ratio PrePS/PFS event = 4.34%
Scenario [50]
prop PFS events = 84.45%
prop OS events = 45.91%
ratio PFS/OS events = 54.37%
ratio PrePS/PFS event = 5.06%
Scenario [51]
prop PFS events = 85.46%
prop OS events = 63.8%
ratio PFS/OS events = 74.65%
ratio PrePS/PFS event = 6.26%
Scenario [52]
prop PFS events = 85.52%
prop OS events = 30.02%
ratio PFS/OS events = 35.1%
ratio PrePS/PFS event = 13.05%
Scenario [53]
prop PFS events = 84.39%
prop OS events = 45.59%
ratio PFS/OS events = 54.03%
ratio PrePS/PFS event = 13.65%
Scenario [54]
prop PFS events = 85.36%
prop OS events = 63.41%
ratio PFS/OS events = 74.28%
ratio PrePS/PFS event = 15.14%
KM (censored dataset)
Truth (complete dataset)
Survival probability
0.0 0.2 0.4 0.6 0.8 1.0
0 100 200 300 400 500
Time since randomisation (in Weeks)

Appendix 11 : Code used for the simulation study (to estimate health state sojourn time only)

```
################################################################################
# Please note that the code for the Li's model is available on request to the original authors
#  The code used to generate the simulation scenarios is available on Request
(r.rafia@sheffield.ac.uk)
################################################################################
rm(list=ls(all=TRUE)) # Clear console
set.seed(15) ; id<-as.numeric(Sys.getenv("SGE_TASK_ID")); print(id); # Dataset number and set
seed
#setwd("~/Documents/Dataset[49]"); id<-50 # Test on a Scenario 49; Dataset 49
Start <- Sys.time(); #Starting time

################################################################################
# Download required libraries and corrected function for mssample
################################################################################
library(robustbase);library(survival);library(flexsurv);library(copula);library(asaur);library
(mstate);library(MASS);library(optimx);library(coda);library(MCMCpack);library(boot);library(m
vtnorm);library(SurvCorr);library(triangle);library(data.table);
source("_mssamplenew.R") # Download corrected version of mssample function

################################################################################
# Select Scenario and define key inputs (number of bootstrap, distribution considered…)
################################################################################
.Sc<-1 # 1 = Base case scenario (i.e selection based on plausibility); 2 = AIC scenario
.nbBoot=1000; # Number of boostrap (probabilistic analysis)
.qpfs=0.8; .qpd=0.5; # Utility values for PFS and PD used to calculate QALYs
.mgin<-c(0.85,0.05) # Margins for visual fit and long term plausibility
.name<-c("exp","weibull","gompertz","lnorm","llogis","gamma","gengamma"); # Spline excluded
.np<-c(1,2,2,2,2,2,3); .np<-.np[1:length(.name)] # Spline excluded
.thin=20; .burn=5000 # thining interval and burning period for Copula model

################################################################################
# Download dataset (240,000 patients), life table and inputs for long term plausibility
################################################################################
dataset<-read.table(paste0("dataset.txt"),skip=1); dataset<-as.data.frame(dataset);
names(dataset)<-
c("empty","ID","age","response","pfs","pfstime","os","ostime","pfs.true","pfstime.true","os.tr
ue","ostime.true") ;

lifetable<- read.csv("lifetable.csv", header=TRUE);
lifetable$pool<-(lifetable$Male+lifetable$Female)/2;lifetable$pool<-1-exp(-lifetable$pool);
lifetable$pool<--(log(1-lifetable$pool))/52;lifetable$pool<-1-exp(-lifetable$pool);lifetab<-
NULL;
for    (i    in    0:100){lifetab2<-rep(lifetable$pool[i+1],52);lifetab<-c(lifetab,lifetab2)};
life_table<-c(lifetab,1);

.input<-read.table(paste0("Input.txt"),skip=1); .input<-as.data.frame(.input);
.plaus<-.input[,1:2];true.m<-.input[,3];    .truepfs<-true.m[1];.trueos<-true.m[2];.trueqaly<-
true.m[3]

################################################################################
# Select dataset, define cycle …
################################################################################
n_row_dat<-nrow(dataset);n_pat_dat<-n_row_dat/1000;
start.dat<-1+(n_pat_dat*(id-1));end.dat<-n_pat_dat*id;dataset<-dataset[start.dat:end.dat,];
med.age<-median(dataset$age) # median age
.LT<-life_table[round((med.age*52)):length(life_table)]; # life table to use (taken from median
age observed in the trial)
.lgp1<-length(.LT)+1; .lgp0<-length(.LT); # Define time horizon
.sq<-seq(from=0, to = length(.LT), by=1) # Define cycle accross time horizon (ex 0,1,2,3,4,...10)
.QTT<-c(t(matrix(rep(.sq,.nbBoot),nrow=.lgp1)))  # cycle repeat for number of bootsrap (ex
0,0,0,1,1,1,2,2,2,3,3,3,4,4,4....10,10,10)
```

```
##############################################################################################
# Generate datasets for TTP, PFS, OS, prePS and PPS (with or without covariate)
##############################################################################################
dataset$ttptime<-dataset$pfstime; dataset$ttp<-dataset$pfs
dataset$ttp[dataset$pfstime==dataset$ostime&dataset$os==1]<-0
dataset$prepstime<-dataset$pfstime; dataset$preps<-0
dataset$preps[dataset$pfstime==dataset$ostime&dataset$os==1]<-1
dataset$ppstime<-dataset$ostime-dataset$pfstime; dataset$pps<-dataset$os
pps_data<-subset(dataset,dataset$pfs==1&dataset$ppstime>0)

# Datasets are set up with 3 variables as time, event, covariate (if appropriate); # NA if no
covariate
pfs_data<-cbind(dataset$pfstime,dataset$pfs,rep(NA,length(dataset$pfs)))
ttp_data<-cbind(dataset$ttptime,dataset$ttp,rep(NA,length(dataset$pfs)))
os_data<-cbind(dataset$ostime,dataset$os,rep(NA,length(dataset$pfs)))
preps_data<-cbind(dataset$prepstime,dataset$preps,rep(NA,length(dataset$pfs)))
pps_data_NoCov<-cbind(pps_data$ppstime,pps_data$pps,rep(NA,length(pps_data$pps)))
pps_data_Log<-cbind(pps_data$ppstime,pps_data$pps,log(pps_data$pfstime))
pps_data_NoLog<-cbind(pps_data$ppstime,pps_data$pps,pps_data$pfstime)

##############################################################################################
# Create matrix for general population mortality (diagnole with NA) – to be used later …
##############################################################################################
.SQT<-c(t(matrix(rep(.sq,.lgp1),nrow=.lgp1)))   #   cycle   repeat   for   cycle   duration   (ex
0,0,0,1,1,1,2,2,2,3,3,3,4,4,4....10,10,10)
.Z<-sapply(1:.lgp0, function(expr){f<-.LT;y<-c(.LT[expr:.lgp0],rep(NA,(expr-1)))})
.Z<-cbind(.Z,rep(NA,.lgp0))

##############################################################################################
# Variables used within mssample function
##############################################################################################
ntrans<-3;tmat<- transMat(list(c(2,3), 3, c()),names = c("progression-free", "progression",
"death"))
tmat2 <- transMat(x = list(c(2, 4), c(3), c(), c()),names=c("PFS", "prog","death after prog",
"death without prog"))
tt.ms<-.lgp1;     tt.ms2<-seq(0,.lgp0,1);     newtrans<-rep(1:ntrans,each=tt.ms);     timeD<-
rep(tt.ms2,ntrans);
preps.name<-rep(.name,length(.name));
ttp.name<-c(t(matrix(preps.name,nrow=length(.name))));                    pfs.loop.msm.name<-
rbind(ttp.name,preps.name)
preps.N<-rep(1:length(.name),length(.name));  ttp.N<-c(t(matrix(preps.N,nrow=length(.name))));
pfs.loop.msm<-rbind(ttp.N,preps.N)

##############################################################################################
# Start simulation – For a given dataset
# Seven methods considered: 1) PSM, 2) unadjusted STM, 3) STM with PPS adjusted by log of TTP,
# 4) STM with PPS adjusted by TTP (non-log scale), 5) MSM using mssample function, 6) Li's model
and 7) Fu's model (Copula)
##############################################################################################

##############################################################################################
# myTryCatch - Function to identify when a function return an error (available online)
# custom tryCatch to return result and warnings -- http://stackoverflow.com/a/24569739/2271856
##############################################################################################
myTryCatch <- function(expr) {
  warn <- err <- NULL; value <- withCallingHandlers(
    tryCatch(expr, error=function(e) {
      err <<- e;NULL
    }), warning=function(w) {
      warn <<- w;invokeRestart("muffleWarning")})
  list(value=value, warning=warn, error=err)
}
```

```
#############################################################################################
# a.S.gp - Function to adjust a survival curve by general population mortality
# This function is applied to a number of column (for instance for 7 distributions or 1,000
boostrapped survival function)
#############################################################################################
a.S.gp<-function(expr){
  vector_s<-Z<-Y<-NULL; #Set variables to NULL
  vector_s<-expr
  Z<-matrix(rep(.LT,ncol(vector_s)),ncol=ncol(vector_s))
  Y<- -log(vector_s);Y<-diff(Y);
  Y[Y<0]<-0;Y[is.nan(Y)]<-0;Y[is.infinite(Y)]<-0;
  Y<-pmax(Y,Z);Y<-1-exp(-Y);
  Y<-1-Y
  Y<-apply(Y,2,cumprod)
  Y<-rbind(rep(1,ncol(vector_s)),Y)
  return(Y)
}

#############################################################################################
# zeroadd - Function to create diagonal for PPS matrix to use in STM
# Also use for tmp.matNA below (to generate matrix with NA and zeroes)
#############################################################################################
zeroadd<-function(x,mat){
  f<-mat[,x]
  y<-c(rep(0,x-1),f[1:(length(f)-x+1)])
}

tmp.matNA<-rbind(rep(NA,ncol(.Z)),.Z*NA) # Generate a matrix with only NA
tmp.matNA<-sapply(1:.lgp1, zeroadd, tmp.matNA) # Transform matrix with zero in diagonale


#############################################################################################
# Some general functions used for survival – For spline model (excluded)
# dmySpline, pmySpline, qmySpline, splinem
# Generate survival distribution for a given distribution
# Note: dat.S need to be of the form time, status, and covariate (NA if no covariate)
#############################################################################################
dmySpline <- function(.x, .gamma1, .gamma2, .gamma3, .knots1, .knots2, .knots3, ...)
  dsurvspline(.x,gamma = c(.gamma1, .gamma2, .gamma3), knots = c(.knots1, .knots2, .knots3),
...)

pmySpline <- function(.x,.gamma1, .gamma2, .gamma3, .knots1, .knots2, .knots3, ...)
  psurvspline(.x,gamma = c(.gamma1, .gamma2, .gamma3), knots = c(.knots1, .knots2, .knots3),...)

qmySpline <- function(.x,.gamma1, .gamma2, .gamma3, .knots1, .knots2, .knots3, ...)
  qsurvspline(.x,gamma = c(.gamma1, .gamma2, .gamma3), knots = c(.knots1, .knots2, .knots3),...)

splinem<-function(dat.S){
  .dat<-.cov<-NULL;
  .dat<-dat.S; .cov<-.dat[,3]
  if(is.na(.cov[1])){
    flexsurvspline(Surv(.dat[,1], .dat[,2]) ~1,k=1,scale="odds")
  }else{
    flexsurvspline(Surv(.dat[,1], .dat[,2]) ~.cov,k=1,scale="odds")
  }
}

#############################################################################################
# survm – Function to fit a specific survival distribution
#############################################################################################
survm<-function(.Distr, dat.S){
  .dat<-.cov<-.ds<-NULL
  .ds<-.Distr;.dat<-dat.S;.cov<-.dat[,3]
  if(is.na(.cov[1])){
    if(.ds=="mySpline"){splinem(.dat)}else{flexsurvreg(Surv(.dat[,1], .dat[,2]) ~ 1, dist=.ds)}
  }else{
    if(.ds=="mySpline"){splinem(.dat)}else{flexsurvreg(Surv(.dat[,1],   .dat[,2])   ~   .cov,
dist=.ds)}
  }
}
```

```
##############################################################################################
# Cdistr - Function to Catch single distributions that return an error that need to be rejected
because
# (a) lack of convergion for instance or (b) drop too quickly (inconsistency identified during
testing)
##############################################################################################
Cdistr<-function(.Distr, dat.S){
  .dat<-.ds<-X<-Y<-Z<-XX<-xx<-NULL
  .dat<-dat.S;.ds<-.Distr;
  X<-myTryCatch(survm(.ds, .dat))
  Y<-myTryCatch(normboot.flexsurvreg(X$value,B=1000, raw=T))
  if(is.null(Y$value)){Z="null"}else{Z=.ds}

  #Added constraint to avoid distribution that produce distribution that drop too quickly
(especially gengamma)

  if(Z!="null"){
    XX<-survm(Z, .dat)
    XX<-summary(object= XX,X = NULL,type = "survival", t = .sq, start = 0, ci = FALSE, tidy =
FALSE);
    XX<- XX[[1]]$est
    xx<-abs(diff(XX));  xx<-max(xx)
    if(xx<0.075){Z<-Z}else{Z="null"}
  }else{Z="null"}
  return(Z)
}

##############################################################################################
# myCdistr - Function to return valid distribution for a given dataset
##############################################################################################
myCdistr<-function(dat.S){
  X<-.dat<-NULL
  .dat<-dat.S
  X<-unlist(lapply(.name,Cdistr,.dat))
  return(X)
}

##############################################################################################
# sv.distr - Generate survival function for a given distribution
# sv.aic - Generate AIC for a given distribution
##############################################################################################
sv.distr<-function(.Distr,dat.S){
  .dat<-.ds<-Y<-Z<-NULL
  .dat<-dat.S;.ds<-.Distr;
  if(.ds=="null"){Y<-rep(NA,.lgp1)}else{
    Z<-survm(.ds,.dat);
    Y<-summary(object= Z,X = NULL,type = "survival", t = .sq, start = 0, ci = FALSE, tidy =
FALSE);
    Y<- Y[[1]]$est
  }
  return(Y)
}

sv.aic<-function(.Distr,dat.S){
  .dat<-.ds<-Y<-NULL
  .dat<-dat.S;.ds<-.Distr;
  if(.ds=="null"){Y<-NA}else{
    Y<-survm(.ds,.dat)$AIC
  }
  return(Y)
}


#ADDED
CdistrALL2<-function(.Distr, dat.S){
  .dat<-.ds<-X<-Y<-Z<-XX<-xx<-NULL
  .dat<-dat.S;.ds<-.Distr;
  X<-myTryCatch(survm(.ds, .dat))
  Y<-myTryCatch(normboot.flexsurvreg(X$value,B=1000, raw=T))
  if(is.null(Y$value)){Z="null"}else{Z=.ds}
  #Added constraint to avoid distribution that produce distribution that drop too quickly
(especially gengamma)
  return(Z)
```

```
}


myCdistr2<-function(dat.S){
  X<-.dat<-NULL
  .dat<-dat.S
  X<-unlist(lapply(.name,CdistrALL2,.dat))
  return(X)
}
##############################################################################################
# .fitdist - Return survival function and AIC for all possible distributions (list[surv, AIC])
# .adjT = used as a variable for whether to adjust for general population or not (as some
distributions dont need to)
##############################################################################################
.fitdist<-function(dat.S,.adjT){
  .dat<-.dd<-Y<-Z<-.adj<-NULL
  .adj<-.adjT
  .dat<-dat.S;
  .dd<-myCdistr(.dat)
  #ADDED FOLLOWING RUNNING
  if(sum(.dd=="null")==7){.dd<-myCdistr2(.dat)}else{.dd<-.dd}
Y<-sapply(.dd,sv.distr,.dat)
  Z<-sapply(.dd,sv.aic,.dat)
  if(.adj=="null"){Y<-Y}else{Y<-a.S.gp(Y)}
  return(list(Y,Z))
}


##############################################################################################
# m.select.fit - Function used to identify survival distribution based on (a) visual fit, (b)
plausibility and (c) AIC or AIC only
##############################################################################################
m.select.fit<-function(dat.S,.pred.sv,.clinpl,.bound){
.dat<-.pred.SV<-.clp<-.bd<-.pred.SV1<-.pred.SV2<-.tmp.time<-.rmNA<-.stp1<-.stp2<-.stp3<-.cr3<-
.cst<-.km<-.kmU<-.kmL<-n.with<-p.with<-r.with<-NULL;
.dat<-dat.S; .pred.SV<-.pred.sv;  .clp<-unlist(.clinpl);  .bd<-.bound;
.pred.SV1<-.pred.SV[[1]]; .pred.SV2<-.pred.SV[[2]]; nRowX<-nrow(.pred.SV1)
.clp2<-unlist(.clp[2]); .clp1<-unlist(.clp[1]); .bd2<-unlist(.bd[2]); .bd1<-unlist(.bd[1])

  .tmp.time<-min(.clp2, nRowX); #Constraint to avoid error (used for when debugging)
  #STEP 1 (a): Remove distribution that reaches 0 when expected survival around 10%
  .rmNA<-.pred.SV1[.tmp.time,];.rmNA[.rmNA<0.0001]<-NA
  #STEP 1 (b) If all distribution reaches 0, then remove constraint
  if (sum(is.na(.rmNA))==length(.rmNA)){.rmNA<-rep(1,length(.rmNA))};
  .rmNA[!is.na(.rmNA)]<-1;   .rmNA<-t(matrix(rep(.rmNA,nRowX),ncol=nRowX));   .stp1<-.pred.SV1
*.rmNA

  #STEP 2: Visual fit (based on fit to KM - within CI)
  .km<-survfit(Surv(.dat[,1],.dat[,2])~1,conf.int=TRUE);
  .surv<-.km$surv;  .kmU<-.km$upper; .kmL<-.km$lower;  .kmtime<-round(.km$time)
  # Constraint added because NA possible in KM at the end
  .cst<-min(sum(!is.na(.kmtime)),sum(!is.na(.kmL)),sum(!is.na(.kmU)));
  .kmU<-.kmU[1:.cst]; .kmL<-.kmL[1:.cst]; .kmtime<-.kmtime[1:.cst]

  .tmpkm<-.stp1[.kmtime+1,] #ADD +1 because start at 1 not 0
  .fL<-.tmpkm>=.kmL;  .fU<-.tmpkm<=.kmU;  n.with<-.fL+.fU;
  p.with<-colSums(n.with==2)/nrow(n.with);  r.with<-min(.mgin[1],max(p.with,na.rm=TRUE));
  if(r.with!=.mgin[1]){r.with<-r.with*0.9}
  s.with<-p.with>=r.with;                  s.with[s.with==FALSE]<-NA;                  s.with<-
t(matrix(rep(s.with,nRowX),ncol=nRowX))
  .stp2<-.stp1 *s.with

  #STEP 3: Selection based on long-term prediction
  .clp2<-min(.clp2,.lgp1); .tmp.pl<-.stp2[.clp2,];
  m.y<-abs(.tmp.pl-.clp1)-.mgin[2];m.y<-pmax(0,ceiling(m.y/0.02));m.y<-min(m.y,na.rm=TRUE)
  mgin.adj<-.mgin[2]+m.y*0.02;
  .cr3<-(between(.tmp.pl,(.clp1-mgin.adj),(.clp1+mgin.adj)));           .cr3[.cr3==FALSE]<-NA;
.cr3.tmp<-.cr3
  .cr3<-t(matrix(rep(.cr3.tmp,nRowX),ncol=nRowX))
  .stp3<-.stp2 *.cr3

  #Constraint added for PFS (make sure PFS lower than expected OS)
  if(is.null(.bd)){.tmpcr<-1; .tmpfs <-1}else{
```

```
    .tmpcr<-NULL; OS.t<- .stp3[min(.bd2,.lgp1),]; OS.p<-.bd1; .tmpcr<-OS.t<OS.p;
    if(sum(.tmpcr,na.rm=TRUE)==0){ .tmpcr<-(OS.t<OS.p+0.05)} # if none, increase range by 0.05
(constraint in case)
    if(sum(.tmpcr,na.rm=TRUE)==0){ .tmpcr<-(OS.t<OS.p+0.25)} # if none, then increase range by
extreme (constraint in case)
    if(sum(.tmpcr,na.rm=TRUE)==0){ .tmpcr<-(OS.t<OS.p+0.50)} # if none, then increase range by
extreme (constraint in case)
    .tmpfs<- .tmpcr
    .tmpfs[.tmpfs==0]<-NA  }
  .cr3.tmp<- .cr3.tmp*.tmpfs
 #STEP 4 (Final): Select distribution based on AIC
  .cr4<-.pred.SV2; .cr4[is.na(.cr3.tmp)]<-NA;

  .base<-which((min(.cr4,na.rm=TRUE)==.cr4)=="TRUE")           ;                        .aic<-
which((min(.pred.SV2,na.rm=TRUE)==.pred.SV2)=="TRUE");
  .curv<-c(.base,.aic); return(.curv)}

##############################################################################################
# p.function.boot - Generate survival distribution for 1,000 boostrapped sample for a given
distribution
# adjT = variable for adjustment for gen pop mortality?
##############################################################################################
p.function.boot<-function(.Distr,dat.S,adjT){
  .ds<-.param<-.p<-.c<-.dat<-.d<-.e<-.adj<-NULL
  .ds<-.Distr; .dat<-dat.S; .param<-par.boot(.ds,.dat); .adj<-adjT
  if(.ds=="gengamma"){
    .p<-do.call(paste0("p",    .ds),    list(.QTT,.param[,1],.param[,2],.param[,3]))}else    if
(.ds=="exp") {
      .p<-do.call(paste0("p", .ds), list(.QTT,.param[,1]))} else if (.ds=="mySpline"){
        .p<-do.call(paste0("p",                                                          .ds),
list(.QTT,.param[,1],.param[,2],.param[,3],.param[,4],.param[,5],.param[,6]))} else if
  (.ds=="null"){           .p<-rep(NA,length(.LT))}else{.p<-do.call(paste0("p",          .ds),
list(.QTT,.param[,1],.param[,2]))}

  .c<-1-t(matrix(.p, nrow = .nbBoot, ncol = .lgp1))
  .d<-cbind(sv.distr(.ds,.dat),.c)
    if(.adj=="null"){.e<-.d}else{.e<-a.S.gp(.d)}
  return(.e)
}


##############################################################################################
# par.boot - Generate parameters for 1,000 boostrapped sample for a given distribution
##############################################################################################
par.boot<-function(.Distr,dat.S){
  .ds<-.dat<-X<-Y<-NULL
  .ds<-.Distr;.dat<-dat.S
  X<-survm(.ds,.dat) ;   Y<-normboot.flexsurvreg(X,B=.nbBoot, raw=T)
  if(.ds=="mySpline"){Y<-cbind(Y,t(matrix(rep(X$knots,.nbBoot),ncol=.nbBoot)))}else{Y<-Y}
  return(Y)
}

##############################################################################################
# gen.tn - Function to generate survival distributions (deterministic + probabilitic) for PFS
and OS (mostly)
# Based on name of distribution, curve is selected based on plausility, and boostrapped curve
are generated
##############################################################################################
gen.tn<-function(Tname){
  X<-.dat<-.clinP<-.limX<-  Pred_x  <-  Selection_x  <-  dist_x<-.scenario1<-.scenario2<-.tname<-
.adjG<-NULL; .tname<-Tname

  if(.tname=="pfs"){.dat<-pfs_data;.clinP<-.plaus[1,];.limX<-.plaus[3,];.adjG<-1}else
if(.tname=="ttp"){
    .dat<-ttp_data;.clinP<-.plaus[2,];.limX<-NULL;.adjG<-"null"}else if(.tname=="os"){
      .dat<-os_data;.clinP<-.plaus[3,];.limX<-NULL;.adjG<-1}else if(.tname=="preps"){
        .dat<-preps_data;.clinP<-.plaus[4,];.limX<-NULL;.adjG<-1}else if(.tname=="pps"){
          .dat<-pps_data;.clinP<-.plaus[5,];.limX<-NULL;.adjG<-1}else{return(error)}

  Pred_x <-.fitdist(.dat,.adjG)
  # Select distribution for the base-case and scenario  analysis
  Selection_x <-m.select.fit(.dat, Pred_x,.clinP,.limX)
  dist_x<-.name[Selection_x];
```

```
  # Boostrapped distribution (option selected to adjusted for general population mortality)
  .scenario1<-p.function.boot(dist_x[1],.dat,.adjG);                              .scenario2<-
p.function.boot(dist_x[2],.dat,.adjG)
  return(list(.scenario1,.scenario2, dist_x))
}

##############################################################################################
# psm.mod – Run PSM (Approach 1)
##############################################################################################
psm.mod<-function(.i,os_mat,pfs_mat){
  pfst<-ost<-pfs.e<-os.e<-NULL
  pfst<-pfs_mat[[.Sc]][,.i];ost<-os_mat[[.Sc]][,.i]
  pfs.e<- pmin(pfst,ost)
  os.e<- ost
  psm<-cbind(pfs.e,os.e)
}

##############################################################################################
# Functions specific to STMs and MSM
##############################################################################################

##############################################################################################
# param.pp.mat - This function is used to transform parameters for PPS according to the covariate
of time
# .Distr = distribution
# .parX = parameter for survival function with the covariate
# .SQ_T = cycle
##############################################################################################
param.pp.mat<-function(.Distr,.parX,.SQ_T){
  .ds<-.param<-.SQ_T1<-NULL;
  .ds<-.Distr;.param<-.parX;.SQ_T1<-.SQ_T
  if(.ds=="exp"){
    par1<-.param[,1]*exp(.param[,2]*.SQ_T1);   par.s<-par1
  }else if (.ds=="lnorm"){
    par1<-.param[,1]+.param[,3]*.SQ_T1;    par2<-.param[,2];     par.s<-cbind(par1,par2)
  }else if(.ds=="gengamma"){
    par1<-.param[,1]+.param[,4]*.SQ_T1;     par2<-.param[,2];     par3<-.param[,3];     par.s<-
cbind(par1,par2,par3)
  }else if(.ds=="mySpline"){
    par1<-.param[,1]+.param[,4]*.SQ_T1;     par2<-.param[,2];     par3<-.param[,3];    par.s<-
cbind(par1,par2,par3)
  }else if(.ds=="null"){
    par<-NA
  }else{
    par1<-.param[,1];    par2<-.param[,2]*exp(.param[,3]*.SQ_T1) ;     par.s<-cbind(par1,par2)
  }
}


##############################################################################################
# svCOLmat  - Generate survival distribution for each column in matrix - for instance .n = 1
select 1:1500; .n=2 select 1501:3000
# . Used because otherwise fail because of Spline
##############################################################################################
svCOLmat<-function(.i,.Distr,.parX){
  .param<-.ds<-.p<-.c<-.n<-NULL
  .ds<-.Distr;   .n<-.i;   .param<-.parX[.n,]

  if(.ds=="gengamma"){
    .p<-do.call(paste0("p", .ds), list(.sq,.param[1],.param[2],.param[3]))}else if (.ds=="exp")
{
      .p<-do.call(paste0("p", .ds), list(.sq,.param))} else if (.ds=="mySpline"){
        .p<-do.call(paste0("p",                                                          .ds),
list(.sq,.param[1],.param[2],.param[3],.param[4],.param[5],.param[6]))} else if (.ds=="null"){
.p<-rep(NA,.lgp1)}else{ .p<-do.call(paste0("p", .ds), list(.sq,.param[1],.param[2]))}
  .c<-1-.p
}
```

```
##############################################################################################
# adj_gp_pps  -  Adjust for general population in matrix for PPS
# .xp = vector for survival to adjust
##############################################################################################
adj_gp_pps<-function(.xp){
  .x<-adjtrans<-NULL;
  .x<-.xp;
  adjtrans<- -log(.x);adjtrans<-diff(adjtrans);
  adjtrans[adjtrans<0]<-0;adjtrans[is.nan(adjtrans)]<-0;adjtrans[is.infinite(adjtrans)]<-0;
  adjtrans<-pmax(adjtrans,.Z);adjtrans<-1-exp(-adjtrans);
  adjtrans<-1-adjtrans
  adjtrans<-apply(adjtrans,2,cumprod)
  adjtrans<-rbind(rep(1,ncol(.x)),adjtrans)
  return(adjtrans)
}

##############################################################################################
# pps.matrix.gen - Generate matrix for PPS with zero in diagonale
# PPS adjusted according to covariate
##############################################################################################
pps.matrix.gen<-function(.Distr,dat.S){
  .ds<-.dat<-X<-Z<-X1<-Y1<-Z1<-XX2<-XX3<-XX4<-NULL
  .ds<-.Distr;.dat<-dat.S
  X<-Cdistr(.ds,.dat) # catch distribution that fail?
  if(X=="null"){Z="null"}else{Z=.ds}

  if(Z=="null"){XX4<-tmp.matNA}else{  #if distribution not valid (then exclude)
    X1<-survm(Z, .dat);
    if(X1$ncovs==0){X1<-c(X1$res[,1],0)}else{X1<-X1$res[,1]}

    X2<-rep(X1,.lgp1) # repeat parameter for length of LT * LT (for)
    X2<-t(matrix(X2,nrow=length(X1)))

    #Log covariate
    #replace inf by 0 when in log scale
    .cov<-.dat[,3]
    if(is.na(.cov[1])){
      a.SQT<-.sq}else if (max(.cov)<10){ #Log dataset cannot be more than 10 (log 15000)
        a.SQT<-log(.sq);a.SQT[is.infinite(a.SQT)]<-0}else{a.SQT<-.sq}

    PAR_STORE<-param.pp.mat(Z,X2,a.SQT)
    PAR_STORE<-as.matrix(PAR_STORE)
    .k<-NULL
    if(Z=="mySpline"){
      .k<-survm(Z,.dat)$knots
      .k<-rep(.k,.lgp1)
      .k<-t(matrix(.k,nrow=3))
      PAR_STORE<-cbind(PAR_STORE,.k)
    }

    XX2<-sapply(1:.lgp1,svCOLmat,Z,PAR_STORE)
    XX3<-adj_gp_pps(XX2)
    XX4<-sapply(1:.lgp1, zeroadd, XX3)
  }
}


##############################################################################################
# genOS.stm - Generate OS prediction for a given distribution, dataframe, PFS and PrePS
# pfs.t = survival distribution for pfs
# preps.t = survival distribution for preps
# pps.t = Matrix for PPS
##############################################################################################
genOS.stm<-function(.i,pfs.t,preps.t,pps.t){
  pfs.t1<-preps.t1<-pps.t1<-NULL
  pfs.t1<-pfs.t;preps.t1<-preps.t;pps.t1<-t(pps.t[[.i]])
  pps.t1[is.na(pps.t1)]<-1
  pfs.d1<-c(0,abs(diff(pfs.t1))) #PFS difference
    #STM1 (using prePS)
  pd.d1<--log(preps.t1);pd.d1[is.infinite(pd.d1)]<-0;pd.d1<-c(abs(diff(pd.d1)));pd.d1<-1-exp(-
pd.d1);
  pd.d1<-pfs.t1[1:(length(pfs.t1)-1)]*pd.d1;pd.d1<-c(0,pd.d1)
```

```
  pd.d1<-pmax(0,pfs.d1-pd.d1);pd.d1<-                          rep(pd.d1,length(pfs.t1));pd.d1<-
t(matrix(pd.d1,length(pfs.t1)));pd.d1<-t(pd.d1)
  n.pd1<-pps.t1*pd.d1;
  pd.e1<-(colSums(n.pd1));os.e1<-pfs.t1+pd.e1;
  return(os.e1)
}

##############################################################################################
# gen.pred.os -Generate STM for all distributions of OS
# Take the first column for PFS and prepS (deterministic)
# This function will be used to define which OS is plausible (to select the plausible PPS)
# pfs.t, preps.t = survival distribution for pfs, preps (deterministic)
# dat.S = PPS data
# Return predicted OS and AIC
##############################################################################################
gen.pred.os<-function(pfs.t,preps.t,dat.S){
  .dat<-pfs.t1<-preps.t1<-pps.t1<-NULL
  .dat<-dat.S;
  #If prediction are in in a list
  pfs.t1<-pfs.t[,1];  preps.t1<-preps.t[,1]
  #Generate matrix for PPS for 7 distribution
  #ADDED FOLLOWING RUNNING
  .nameAD<-unlist(lapply(.name,Cdistr,.dat))
  if(sum(.nameAD=="null")==7){.nameAD<-myCdistr2(.dat)}else{.nameAD<-.nameAD}

  mat.PPS<-lapply(.nameAD,pps.matrix.gen,.dat)
  os.e2<-sapply(1:length(.nameAD),genOS.stm,pfs.t1,preps.t1,mat.PPS)
  os.aic<-sapply(.nameAD,sv.aic,.dat)
  return(list(os.e2,os.aic))
}
##############################################################################################
# fun.os.pred - This function is used to identify which distribution for PPS is the most
plausible
# given PFS and prePS
##############################################################################################
fun.os.pred<-function(dat.S,pfs.t,preps.t){
  dat.<-.cov<-dat.use<-.pfsT<-.prepsT<- Pred_x<- Select_x<-os.data<- Dist_x<-NULL
  .dat<-dat.S; .cov<-.dat[,3]
  .pfsT<-pfs.t; .prepsT<-preps.t
  Pred_x<-gen.pred.os(.pfsT,.prepsT,.dat)

  #SELECT DISTRIBUTION FOR BASE AND SCENARIO
  Select_x<-m.select.fit(os_data, Pred_x,.plaus[3,],NULL)
  Dist_x<-.name[Select_x];
  return(Dist_x)
}

##############################################################################################
# stmgen - This function is used to generate prediction for the STM for for each bootstap sample
# for a selected distribution (which would have been selected beforehand)
# .cycle = value of the coavriate of cycle of time (for ex 1,2,3,4....)or in log scale
# parB = boostrapped parameters (for ex 1000 row of parameters)
##############################################################################################
stmgen<-function(.i,parX,cycle,pfs.t,preps.t,.Distr){
  pfs.t1<-preps.t1<-.ds<-parN<-g<-PAR_STORE<-.cyc<-pps.t1<-os.e<-NULL
  pfs.t1<-pfs.t[,.i]; preps.t1<-preps.t[,.i]; .ds<-.Distr; parN<-parX ; .cyc<-cycle
  XX<-parN[.i,] #Parameters
  X2<-rep(XX,.lgp1)
  X2<-t(matrix(X2,nrow=length(XX)))

  PAR_STORE<-param.pp.mat(.ds,X2,.cyc);  PAR_STORE<-as.matrix(PAR_STORE)
  .k<-NULL;   if(.ds=="mySpline"){  .k<-survm(.ds,.dat)$knots;       .k<-rep(.k,.lgp1);      .k<-
t(matrix(.k,nrow=3))  }

  pps.t1<-sapply(1:.lgp1,svCOLmat,.ds,PAR_STORE)
  pps.t1<-matrix(pps.t1,nrow=.lgp1,ncol=.lgp1)
  pps.t1<-adj_gp_pps(pps.t1); # Adjust general population mortality
  pps.t1<-sapply(1:.lgp1, zeroadd, pps.t1)
  os.e<-genOS.stm2(pfs.t1,preps.t1,pps.t1)
  return(os.e)
}
```

```
##############################################################################################
# genOS.stm2 - Generate a single STM for a given PFS, PrePS and PPS curve
##############################################################################################
genOS.stm2<-function(pfs.t,preps.t,pps.t){
  pps.t1<-preps.t1<-pfs.t1<-pd.d1<-pd.e1<-os.e1<-pd.d1<-NULL
  pfs.t1<-pfs.t;preps.t1<-preps.t;pps.t1<-t(pps.t)
  pps.t1[is.na(pps.t1)]<-1
  pfs.d1<-c(0,abs(diff(pfs.t1))) #PFS difference

  #STM1 (using prePS)
  pd.d1<--log(preps.t1);pd.d1[is.infinite(pd.d1)]<-0;pd.d1<-c(abs(diff(pd.d1)));pd.d1<-1-exp(-
pd.d1);
  pd.d1<-pfs.t1[1:(length(pfs.t1)-1)]*pd.d1;pd.d1<-c(0,pd.d1)
  pd.d1<-pmax(0,pfs.d1-pd.d1);pd.d1<-                    rep(pd.d1,length(pfs.t1));pd.d1<-
t(matrix(pd.d1,length(pfs.t1)));pd.d1<-t(pd.d1)
  n.pd1<-pps.t1*pd.d1;
  pd.e1<-(colSums(n.pd1));os.e1<-pfs.t1+pd.e1;
  return(cbind(pfs.t1,os.e1))
}

##############################################################################################
# stm.mod - Generate prediction for STM model (depending on data, PPS will be adjusted or not
by the covariate)
##############################################################################################
stm.mod<-function(dat.S,pfs.t,preps.t){
  .dat<-pfs.t1<-preps.t1<-NULL
  pfs.t1<-pfs.t[[.Sc]];   preps.t1<-preps.t[[.Sc]];   .dat<-dat.S
  #Step 1 identify distribution to select for base-case and scenario
  selectdist.pps<-fun.os.pred(.dat,pfs.t1,preps.t1)
  #SELECT DISTRIBUTION FOR GIVEN SCENARIO
  .ds<-selectdist.pps[.Sc]
  #Start with base-case
  Z<-survm(.ds,.dat);  B<-normboot.flexsurvreg(Z,B=.nbBoot, raw=T);   H<-rbind(Z$res[,1],B)
  if(Z$ncovs==0){H<-cbind(H,rep(0,nrow(H)))}else{H<-H}
  .cov<-.dat[,3] # Covariate
  if(is.na(.cov[1])){a.SQT<-.sq}else                 if                 (max(.cov)<10){a.SQT<-
log(.sq);a.SQT[is.infinite(a.SQT)]<-0}else{ a.SQT<-.sq}
  XX6<-sapply(1:nrow(H),stmgen,H,a.SQT,pfs.t1,preps.t1,.ds)
  return(XX6)
}

##############################################################################################
# Functions specific to MSMs
##############################################################################################

##############################################################################################
# approx_mssample - Generate approximation for MSM for a given PPS, TTP and prePS (similar to
STM)
##############################################################################################
approx_mssample<-function(pps.t,ttp.t,preps.t){
  LY_e <-OS_e <-PFS_e <-pd.e3<-n.pd3<-pfs.t3<-pd.d3<-preps.msm<-ttp.msm<-mat.pps<-NULL;
  preps.msm<-preps.t; preps.msm<-diff(preps.msm);
  ttp.msm<-ttp.t; ttp.msm<-diff(ttp.msm);
  pfs.t3<-ttp.msm+preps.msm;
  pfs.t3<-1-exp(-pfs.t3); pfs.t3<-cumprod(1 - pfs.t3); pfs.t3<-c(1,pfs.t3);mat.pps<-t(pps.t)
  pfs.d3<-c(0,abs(diff(pfs.t3)));ttp.dy<-ttp.msm; preps.dy<-preps.msm #Number of progressors
(usin prePS)
  pd.d3<-ttp.dy/(ttp.dy+preps.dy);pd.d3[is.nan(pd.d3)]<-0;pd.d3<-c(0,pd.d3);pd.d3<-
pd.d3*pfs.d3
  pd.d3<- rep(pd.d3,length(ttp.t)); pd.d3<-t(matrix(pd.d3,length(ttp.t))); pd.d3<-t(pd.d3)
  n.pd3<-mat.pps*pd.d3;pd.e3<-(colSums(n.pd3))
  OS_e<-pfs.t3+pd.e3; PFS_e<-pfs.t3
  LY_e<-cbind(PFS_e,OS_e)
  return(LY_e)
}
```

```
#############################################################################################
# genapprox - Generate predictions for approximation method (for probabilitic for a given
bootstrap) for 1,000 bootstrapped
#############################################################################################
genapprox<-function(.i,ttp.t,preps.t,pps.t){
  ttpt1<-prepst1<-ppst1<-XX3<-XX4<-mssamp_app1<-NULL
  ttpt1<-ttp.t[,.i]
  prepst1<-preps.t[,.i]
  ppst1<-exp(-pps.t[,.i])
  ppst1<-rep(ppst1,.lgp1)
  ppst1<-matrix(ppst1,ncol=.lgp1)

  XX3<-adj_gp_pps(ppst1)
  XX4<-sapply(1:.lgp1, zeroadd, XX3)
  mssamp_app1<-approx_mssample(XX4,ttpt1,prepst1)
  return(mssamp_app1)
}

#############################################################################################
# msm.pfs - Generate PFS to look at combination of TTP and prePS that are acceptable
# . d= combination of TTP and prePS (49 combination here)
# 1,000 sample used here as PFS is shortish
#############################################################################################

msm.pfs<-function(.i,.d,ttp.t,preps.t){
  .dS<-ttp.x<-preps.x<-pps.x<-cumHaz<-Haz<-stateprobsT1<-pfs.e<-NULL
  .dS<-.d[,.i];   ttp.x<-ttp.t[,.dS[1]];   preps.x<-preps.t[,.dS[2]]
  pps.x<-(1-pexp(.sq,exp(-2))) #Arbirtrary PPS assumed (short for speed)

  if(is.na(preps.x[2])|is.na(ttp.x[2])){pfs.e<-rep(NA,.lgp1)}else{
    cumHaz<-c(ttp.x,preps.x,preps.x); Haz<-cumHaz; Haz[!is.finite(Haz)]<-0;
    Haz<-cbind(time=as.vector(timeD),Haz=as.vector(Haz),trans=as.vector(newtrans));      Haz<-
as.data.frame(Haz)
    stateprobsT1 <- MSsample.new(Haz=Haz,trans=tmat2,tvec=tt.ms2,clock="reset", M=1000)
    pfs.e<- stateprobsT1$pstate1
  }
  return(pfs.e)
}

#############################################################################################
# msm.os - Generate OS prediction for 7 combinations
# generate OS to look at what PPS is acceptable
# 10,000 samples used (as otherwise some inconsistencies)
#############################################################################################
msm.os<-function(.i,.d,ttp.t,preps.t,pps.t){
  .dS<-ttp.x<-preps.x<-pps.x<-cumHaz<-Haz<-stateprobsT1<-pfs.e<-NULL
  .dS<-.i;   ttp.x<-ttp.t;   preps.x<-preps.t;   pps.x<-pps.t[,.dS]
    if(is.na(pps.x[2])){os.e<-rep(NA,.lgp1)}else{
    cumHaz<-c(ttp.x,preps.x, pps.x); Haz<-cumHaz; Haz[!is.finite(Haz)]<-0;
    Haz<-cbind(time=as.vector(timeD),Haz=as.vector(Haz),trans=as.vector(newtrans));      Haz<-
as.data.frame(Haz)
    stateprobsT1 <- MSsample.new(Haz=Haz,trans=tmat2,tvec=tt.ms2,clock="reset", M=10000)
    os.e <- 1-(stateprobsT1$pstate3+stateprobsT1$pstate4)
  }
  return(os.e)
}

#############################################################################################
# msm.all - Run the multistate (deterministic only)
# prediction for both PFS and OS
# 10,000 used (as otherwise some inconsistencies)
#############################################################################################
msm.all<-function(ttp.t,preps.t,pps.t){
  ttp.x<-preps.x<-pps.x<-cumHaz<-Haz<-stateprobsT1<-pfs.e<-NULL
  ttp.x<-ttp.t[,1];  preps.x<-preps.t[,1] ;   pps.x<-pps.t[,1]
  cumHaz<-c(ttp.x,preps.x,pps.x); Haz<-cumHaz; Haz[!is.finite(Haz)]<-0;
  Haz<-cbind(time=as.vector(timeD),Haz=as.vector(Haz),trans=as.vector(newtrans));        Haz<-
as.data.frame(Haz)
  stateprobsT1 <- MSsample.new(Haz=Haz,trans=tmat2,tvec=tt.ms2,clock="reset", M=10000)
  pfs.e<- stateprobsT1$pstate1
  os.e <- 1-(stateprobsT1$pstate3+stateprobsT1$pstate4)
  pred.e<-cbind(pfs.e,os.e)
  return(pred.e)}
```

```
################################################################################################
# MSM.mod – Wrapper MSM function
################################################################################################
MSM.mod<-function(){
  ttp.fit<-preps.fit<-pps.fit<-d_ttp.msm1<-d_preps.msm1<-d_pps.msm1<-d_ttp_boot<-
d_preps_boot<-d_pps_boot<-NULL
  ttp.fit<-.fitdist(ttp_data,"null") # no adjustment for gen pop
  preps.fit<-.fitdist(preps_data,1) # adjusted for gen pop
  pps.fit<-.fitdist(pps_data_NoCov,1) # adjusted for gen pop
#AIC scenario
  aic.ttp<-ttp.fit[[2]];   aic.preps<-preps.fit[[2]] ;   aic.pps<-pps.fit[[2]]
  ttp.aic<-c(t(matrix(rep(aic.ttp,length(aic.ttp)),nrow=length(aic.ttp))))
  preps.aic<-rep(aic.preps,length(aic.preps));     aic.msm.pfs<-rbind(ttp.aic,preps.aic)     ;
aic.msm.pfs<-colSums(aic.msm.pfs,na.rm = FALSE)

  if(.Sc==1){ # Base-case
    #adjustment to make sure no NA, and curve is decreasing (for Scenario 1 only)
    Trans1Cum1<- -log(ttp.fit[[1]]);Trans1Cum1<-diff(Trans1Cum1);Trans1Cum1[Trans1Cum1<0]<-0;
    Trans1Cum1[is.nan(Trans1Cum1)]<-0;Trans1Cum1[is.infinite(Trans1Cum1)]<-
0;Trans1Cum1[Trans1Cum1<0]<-0;
    Trans1Cum1<-rbind(rep(0,ncol(Trans1Cum1)),apply(Trans1Cum1, 2, cumsum));

    Trans2Cum1<-  -log(preps.fit[[1]]);Trans2Cum1<-diff(Trans2Cum1);Trans2Cum1[Trans2Cum1<0]<-
0;
    Trans2Cum1[is.nan(Trans2Cum1)]<-0;Trans2Cum1[is.infinite(Trans2Cum1)]<-
0;Trans2Cum1[Trans2Cum1<0]<-0;
    Trans2Cum1<-rbind(rep(0,ncol(Trans2Cum1)),apply(Trans2Cum1, 2, cumsum));

    Trans3Cum1<- -log(pps.fit[[1]]);Trans3Cum1<-diff(Trans3Cum1);Trans3Cum1[Trans3Cum1<0]<-0;
    Trans3Cum1[is.nan(Trans3Cum1)]<-0;Trans3Cum1[is.infinite(Trans3Cum1)]<-
0;Trans3Cum1[Trans3Cum1<0]<-0;
    Trans3Cum1<-rbind(rep(0,ncol(Trans3Cum1)),apply(Trans3Cum1, 2, cumsum));

    #Generate prediction for PFS (based on TPP and preps) for all combinations
    pred_pfs_msm<-sapply(1:length(ttp.N),msm.pfs,pfs.loop.msm,Trans1Cum1,Trans2Cum1)
    pred_pfs_msm<-list(pred_pfs_msm,aic.msm.pfs)
    select.pfs.msm<-m.select.fit(pfs_data,pred_pfs_msm,.plaus[1,],.plaus[3,])

    d_pfs.msm<-pfs.loop.msm.name[,select.pfs.msm];                d_ttp.msm1<-d_pfs.msm[1,1];
d_preps.msm1<-d_pfs.msm[2,1]
    # Select distribution
    ttp.msm.os<-Trans1Cum1[,d_ttp.msm1]
    preps.msm.os<-Trans2Cum1[,d_preps.msm1]
    #Now lets identify the distribution that provide OS given TTP and PREPS
    pred_os_msm<-sapply(1:length(.name),msm.os,.name,ttp.msm.os,preps.msm.os,Trans3Cum1)
    pred_os_msm<-list(pred_os_msm,aic.pps)
    select.os.msm<-m.select.fit(os_data,pred_os_msm,.plaus[3,],NULL)
    d_pps.msm1<-.name[select.os.msm[1]]
  }else{ # AIC scenario
    x1<-ttp.fit[[2]];       .aic1<-which((min(x1,na.rm=TRUE)==x1)=="TRUE");       d_ttp.msm1<-
.name[.aic1]
    x2<-preps.fit[[2]];     .aic2<-which((min(x2,na.rm=TRUE)==x2)=="TRUE");     d_preps.msm1<-
.name[.aic2]
    x3<-pps.fit[[2]] ;       .aic3<-which((min(x3,na.rm=TRUE)==x3)=="TRUE");      d_pps.msm1<-
.name[.aic3]
  }
  d_ttp_boot<-p.function.boot(d_ttp.msm1,ttp_data,"null")
  d_preps_boot<-p.function.boot(d_preps.msm1,preps_data,1)
  d_pps_boot<-p.function.boot(d_pps.msm1,pps_data_NoCov,1)
  tCum1<-NULL;tCum2<-NULL;tCum3<-NULL
  tCum1<- -log(d_ttp_boot);tCum1<-diff(tCum1);tCum1[tCum1<0]<-0;
  tCum1[is.nan(tCum1)]<-0;tCum1[is.infinite(tCum1)]<-0;tCum1[tCum1<0]<-0;
  tCum1<-rbind(rep(0,ncol(tCum1)),apply(tCum1, 2, cumsum));
  tCum2<- -log(d_preps_boot);tCum2<-diff(tCum2);tCum2[tCum2<0]<-0;
  tCum2[is.nan(tCum2)]<-0;tCum2[is.infinite(tCum2)]<-0;tCum2[tCum2<0]<-0;
  tCum2<-rbind(rep(0,ncol(tCum2)),apply(tCum2, 2, cumsum));
  tCum3<- -log(d_pps_boot);tCum3<-diff(tCum3);tCum3[tCum3<0]<-0;
  tCum3[is.nan(tCum3)]<-0;tCum3[is.infinite(tCum3)]<-0;tCum3[tCum3<0]<-0;
  tCum3<-rbind(rep(0,ncol(tCum3)),apply(tCum3, 2, cumsum));
  xy<-c(msm.all(tCum1,tCum2,tCum3)) #deterministic
  # probabilitic (2:1001)
  xy1<-sapply(2:ncol(tCum1),genapprox,tCum1,tCum2,tCum3);   xy2<-cbind(xy,xy1)
  return(xy2)}
```

```
#############################################################################################
# Functions specific to Li's model (2015)
#############################################################################################
```
[The original code is available on request to the Authors (Li et al, 2015)]
The code was adapted for this thesis

```
#############################################################################################
# Functions specific to Fu's model (2013) - Copula model
#############################################################################################

#############################################################################################
# fun.transform.init.param - Function to transform initial parameter (to help with convergence)
#############################################################################################
fun.transform.init.param<-function(.dist,parm){
  .ds<-.par<-x<-NULL ;   .ds<-.dist;.par<-parm
  if(.ds=="gengamma"|.ds=="mySpline"){x<-1/exp(.par)}else if(.ds=="gompertz"){
    par1<-.par[1];par2<-.par[2];x<-c(exp(par1),1/par2)}else{x<-1/.par}
}

#############################################################################################
# fun.transform.param - Function to transform back parameters
#############################################################################################
fun.transform.param<-function(.dist,parm){
  .ds<-.par<-x<-NULL
  .ds<-.dist;.par<-parm
  if(.ds=="gengamma"){x<-log(1/.par)}else if(.ds=="gompertz"){
    if(is.null(ncol(.par))){par1<-.par[1];par2<-.par[2];x<-c(log(par1),1/par2)}else{par1<-
.par[,1];par2<-.par[,2];
x<-cbind(log(par1),1/par2)}}else if (.ds=="mySpline"){
      if(is.null(ncol(.par))){par1<-.par[1];par2<-.par[2];par3<-.par[3];par4<-.par[4];par5<-
.par[5];par6<-.par[6];
x<-c(log(1/par1),log(1/par2),log(1/par3),1/par4,1/par5,1/par6)}else{par1<-.par[,1];par2<-
.par[,2];par3<-.par[,3];
par4<-.par[,4];par5<-.par[,5];par6<-.par[,6];x<-
cbind(log(1/par1),log(1/par2),log(1/par3),1/par4,1/par5,1/par6)}
    }else{x<-1/.par}}

#############################################################################################
# fun.init.parm - Function to estimate initial parameters
#############################################################################################
fun.init.parm<-function(.dist,time,status){
  knot<-x<-par<-par.T<-.ds<-.time<-.status<-NULL
  .ds<-.dist;.time<-time;.status<-status
  if(.ds=="mySpline"){
    x<-     flexsurvspline(Surv(.time,.status)~1,k=1,scale="odds");   knot<-x$knots}else{   x<-
flexsurvreg(Surv(.time,.status)~1,dist=.ds)}
  par<-fun.transform.init.param(.ds,x$res[,1]);   par.T<-c(par,1/knot)}

#############################################################################################
# param.i - Function to define the initial parameters for the marginal selected - Otherwise
error message
#############################################################################################
param.i<-function(.dist){
  .ds<-x<-NULL;   .ds<-.dist
  if(.ds=="exp"){
    x<-list(rate=0.05)}else if(.ds=="weibull"){
      x<-list(shape=0.1, scale=1)}else if (.ds=="gompertz"){
        x<-list(shape=0, rate=1)}else if (.ds=="lnorm"){
          x<-list(meanlog=1, sdlog=1)}else if (.ds=="llogis"){
            x<-list(shape=1, scale=1)}else if (.ds=="gamma"){
              x<-list(shape=0, rate=1)}else if (.ds=="gengamma"){
                x<-list(mu=0, sigma=1, Q=1)}else if (.ds=="mySpline"){
                  x<-list(.gamma1=0.05,  .gamma2=0.05,  .gamma3=0.05,  .knots1=1,  .knots2=1,
.knots3=1)}else {
                    x<-"error"}  }
```

```
##############################################################################################
# dfunction – Density function to use
##############################################################################################
dfunction<-function(.dist,p.ar,ve.c){
  .ds<-NULL;   .ds<-.dist
  if(.ds=="gengamma"){
    x<-do.call(paste0("d", .ds), list(ve.c,p.ar[1],p.ar[2],p.ar[3]))}else if (.ds=="exp") {
      x<-do.call(paste0("d", .ds), list(ve.c,p.ar[1]))} else if (.ds=="mySpline"){
        x<-do.call(paste0("d",                                                          .ds),
list(ve.c,p.ar[1],p.ar[2],p.ar[3],p.ar[4],p.ar[5],p.ar[6]))} else {
          x<-do.call(paste0("d", .ds), list(ve.c,p.ar[1],p.ar[2]))
        }
}


##############################################################################################
# pfunction – probability function to use
##############################################################################################
pfunction<-function(.dist,p.ar,ve.c){
  .ds<-NULL
  .ds<-.dist
  if(.ds=="gengamma"){
    x<-do.call(paste0("p", .ds), list(ve.c,p.ar[1],p.ar[2],p.ar[3]))}else if (.ds=="exp") {
      x<-do.call(paste0("p", .ds), list(ve.c,p.ar[1]))} else if (.ds=="mySpline"){
        x<-do.call(paste0("p",                                                          .ds),
list(ve.c,p.ar[1],p.ar[2],p.ar[3],p.ar[4],p.ar[5],p.ar[6]))} else {
          x<-do.call(paste0("p", .ds), list(ve.c,p.ar[1],p.ar[2]))
        }
}

##############################################################################################
# qfunction – probability function to use
##############################################################################################
qfunction<-function(.dist,p.ar,ve.c){
  .ds<-NULL
  .ds<-.dist
  if(.ds=="gengamma"){
    x<-do.call(paste0("q", .ds), list(ve.c,p.ar[1],p.ar[2],p.ar[3]))}else if (.ds=="exp") {
      x<-do.call(paste0("q", .ds), list(ve.c,p.ar[1]))} else if (.ds=="mySpline"){
        x<-do.call(paste0("q",                                                          .ds),
list(ve.c,p.ar[1],p.ar[2],p.ar[3],p.ar[4],p.ar[5],p.ar[6]))} else {
          x<-do.call(paste0("q", .ds), list(ve.c,p.ar[1],p.ar[2]))
        }
}

##############################################################################################
# wrapper.copula – Generate copula paremeters and parameters for marginals for TTP and OS
# Adapted from Fu et al (2013
##############################################################################################
wrapper.copula<-function(ttp.dist,os.dist){
.dttp<-.dos<-.dat<-init.pam<-dnX<-npar1<-npar1<-npar2<-par1<-par2<-par3<-par.c1<-par.c2<-
par.c<-par.all<- E.myCop.norm<-NULL
dnX<-npar1<-npar2<-.npar1<-.npar2<-.npar3<-.tpar1<-.tpar2<-.tpar<-cop.est<-.dat<-init.pam<-
dn1<-dn2<-dn<-p1<-p2<- cop.est<-NULL
  .dttp<-ttp.dist;.dos<-os.dist
  #Set data up (censor for TTP)
  .dat<-dataset
  .dat<-cbind(.dat$pfstime,.dat$ostime,.dat$pfs,.dat$os)   ;        .dat<-as.data.frame(.dat);
names(.dat)<-c("PD","OS","delta","xi")
  .dat$delta[.dat$PD==.dat$OS &.dat$xi==1]<-0; #censor for TTP

  init.pam<-c(fun.init.parm(.dttp,.dat$PD,.dat$delta),fun.init.parm(.dos,.dat$OS,.dat$xi),0.5)
  p1<-param.i(.dttp);p2<-param.i(.dos)
  E.myCop.norm <- normalCopula(0,dim=2,dispstr="ex");
  E.myMvd <- mvdc(copula=E.myCop.norm, margins =c(.dttp,.dos),paramMargins = list(p1,p2));

  postloglikelihood <- function(para,dat,E.myMvd){ eps <- 1e-50;
  ff<-length(para)-1
  if (ff==2){
    if  ((para[1]  <=  eps)  |  (para[2]  <=  eps)  |  (para[length(para)]  <=  -(1-eps))  |
(para[length(para)]>= (1-eps))) return(-Inf)}else if((ff==3)){
      if ((para[1] <= eps) | (para[2] <= eps)| (para[3] <= eps) | (para[length(para)] <= -(1-
eps)) | (para[length(para)]>= (1-eps))) return(-Inf)}else if((ff==4)){
```

```
        if ((para[1] <= eps) | (para[2] <= eps)| (para[3] <= eps) | (para[4] <= eps)|
(para[length(para)] <= -(1-eps)) | (para[length(para)]>= (1-eps))) return(-Inf)}else
if((ff==5)){
          if ((para[1] <= eps) | (para[2] <= eps) | (para[3] <= eps) | (para[4] <= eps) |
(para[5] <= eps)| (para[length(para)] <= -(1-eps)) | (para[length(para)]>= (1-eps))) return(-
Inf)}else if((ff==6)){
            if ((para[1] <= eps) | (para[2] <= eps) | (para[3] <= eps)| (para[4] <= eps)|
(para[5] <= eps)| (para[6] <= eps)| (para[length(para)] <= -(1-eps)) | (para[length(para)]>=
(1-eps))) return(-Inf)}else if((ff==7)){
              if ((para[1] <= eps) | (para[2] <= eps) | (para[3] <= eps)| (para[4] <= eps)|
(para[5] <= eps)| (para[6] <= eps)| (para[7] <= eps)| (para[length(para)] <= -(1-eps)) |
(para[length(para)]>= (1-eps))) return(-Inf)}else if((ff==8)){
                if ((para[1] <= eps) | (para[2] <= eps) | (para[3] <= eps)| (para[4] <= eps)|
(para[5] <= eps)| (para[6] <= eps)| (para[7] <= eps)| (para[8] <= eps)| (para[length(para)] <=
-(1-eps)) | (para[length(para)]>= (1-eps))) return(-Inf)}else if((ff==9)){
                  if ((para[1] <= eps) | (para[2] <= eps) | (para[3] <= eps)| (para[4] <= eps)|
(para[5] <= eps)| (para[6] <= eps)| (para[7] <= eps)| (para[8] <= eps)| (para[9] <= eps)|
(para[length(para)] <= -(1-eps)) | (para[length(para)]>= (1-eps))) return(-Inf)}else
if((ff==12)){
                    if ((para[1] <= eps) | (para[2] <= eps) | (para[3] <= eps)| (para[4] <=
eps)| (para[5] <= eps)| (para[6] <= eps)| (para[7] <= eps)| (para[8] <= eps)| (para[9] <= eps)|
(para[10] <= eps) | (para[11] <= eps) | (para[12] <= eps)| (para[length(para)] <= -(1-eps)) |
(para[length(para)]>= (1-eps))) return(-Inf)}

  dnX<-E.myMvd@margins ;   npar1<-.np[.name==dnX[[1]]] ;   npar2<-.np[.name==dnX[[2]]]

  par1<-para[1:npar1] ;   par2<-para[(npar1+1):(npar1+npar2)] ;   par3<-tail(para,n=1)

  par.c1<-fun.transform.param(dnX[1],par1)  ;      par.c2<-fun.transform.param(dnX[2],par2)  ;
par.c<-par3 ;   par.all<-c(par.c1,par.c2,par.c)

  delta <- dat[,3]; xi <- dat[,4];    c1.indx <- delta & xi; c2.indx <- delta & (!xi);   c3.indx
<- (!delta)&xi; c4.indx <- (!delta)&(!xi);
  dn1<-.dttp;dn2<-.dos

  # decompose the loglikelihood function to 4 parts according to the manuscript
  # First Component
  if(sum(c1.indx)>0){
    loglik.c1 <- loglikMvdc(par.all,as.matrix(dat[c1.indx,1:2]),E.myMvd);
  } else {    loglik.c1 <- 0;
  }
  #Second
  if(sum(c2.indx)>0){
    loglik.c2 <- sum(log(pnorm(qnorm(pfunction(dn2,par.c2,dat[c2.indx,2])),
                               mean=par.c*qnorm(pfunction(dn1,par.c1,dat[c2.indx,1])),
                               sd=sqrt(1-par.c^2),
                               lower.tail=F)*dfunction(dn1,par.c1,dat[c2.indx,1])));
  } else {    loglik.c2 <- 0;
  }
  #Third Component
  if(sum(c3.indx) >0 ){
      loglik.c3 <- sum(log(pnorm(qnorm(pfunction(dn1,par.c1,dat[c3.indx,1])),
                               mean=par.c*qnorm(pfunction(dn2,par.c2,dat[c3.indx,2])),
                               sd=sqrt(1-par.c^2),
                               lower.tail=F)*dfunction(dn2,par.c2,dat[c3.indx,2])));
  } else {    loglik.c3 <- 0;
  }
  # Fourth Component
  if(sum(c4.indx)>0){
    sigma <- matrix(c(1,par.c,par.c,1),nrow=2);
    CDF <- function(V,sigma){
      return(pmvnorm(lower = V,upper=Inf, sigma=sigma,mean=c(0,0))[1])
    }
    loglik.c4<-
sum(log(apply(qnorm(cbind(pfunction(dn1,par.c1,dat[c4.indx,1]),pfunction(dn2,par.c2,dat[c4.ind
x,1]))),1,CDF,sigma)));
  } else {    loglik.c4 <- 0;
  }
  loglik <- loglik.c1+loglik.c2+loglik.c3+loglik.c4; return(loglik);
  }
cop.est<-myTryCatch(MCMCmetrop1R(postloglikelihood,theta.init=init.pam,              dat=.dat,
E.myMvd=E.myMvd,thin=.thin,
mcmc=(.nbBoot*.thin), burnin=.burn,tune=1, logfun=TRUE,optim.method = "Nelder-Mead"))
```

```
post.all1<-NULL;post.all1<-cop.est$value

if(is.null(post.all1)){cop.est<-myTryCatch(MCMCmetrop1R(postloglikelihood,theta.init=init.pam,
dat=.dat, E.myMvd=E.myMvd,
thin=.thin,mcmc=(.nbBoot*.thin), burnin=.burn,tune=1, logfun=TRUE,optim.method = "BFGS"))}
post.all2<-NULL;post.all2<-cop.est$value

if(is.null(post.all2)){cop.est<-myTryCatch(MCMCmetrop1R(postloglikelihood,theta.init=init.pam,
dat=.dat, E.myMvd=E.myMvd,
thin=.thin,mcmc=(.nbBoot*.thin), burnin=.burn,tune=1, logfun=TRUE,optim.method = "CG"))}
post.all3<-NULL;post.all3<-cop.est$value

if(is.null(post.all1) & is.null(post.all2) & is.null(post.all3)){cop.est<-NULL}

if(is.null(cop.est$value)){.tpar<-NA}else{
    mat_par_cop<-cop.est$value
    dnX<-E.myMvd@margins
    npar1<-.np[.name==dnX[[1]]]
    npar2<-.np[.name==dnX[[2]]]

    .npar1<-mat_par_cop[,1:npar1]
    .npar2<-mat_par_cop[,(npar1+1):(npar1+npar2)]
    .npar3<-mat_par_cop[,ncol(mat_par_cop)]

    #Return parameter back to normal values
    .tpar1<-fun.transform.param(dnX[[1]],.npar1)
    .tpar2<-fun.transform.param(dnX[[2]],.npar2)
    .tpar<-cbind(.tpar1,.tpar2,.npar3)
  }
  return(.tpar)
}


##############################################################################################
# gen_pred_cop_pfs - Generate prediction for PFS
##############################################################################################
gen_pred_cop_pfs<-function(.i,mat_copula){
  .matcop<-.ds<-prediction.i<-.est<-NULL
  .matcop<-mat_copula[[.i]] ;   .ds<-.name[.i]

  if(is.na(.matcop[1])){prediction.i<-rep(NA,.lgp1*2)}else{
    .matcop<-colMeans(.matcop)
    npar1<-.np[.i] ;     npar2<-.np[.name==.ds.osC] # TTP and OS
    par1<-.matcop[1:npar1]  ;         par2<-.matcop[(npar1+1):(npar1+npar2)]  ;         par3<-
tail(.matcop,n=1)

    #Generate random sample from copula parameters (15,000 sample)
    .coprnd = rCopula(15000, normalCopula(par3));

    pfs.p<-qfunction_Vector(.coprnd[,1],.ds,par1)
    os.p<-qfunction_Vector(.coprnd[,2],.ds.osC,par2)
    pfs.p<-pmin(pfs.p,os.p)

    #round pfs and os to higher value
    pfs.p<-ceiling(pfs.p) ;     os.p<-ceiling(os.p)
    breaks = c(.sq,tail(.sq,1)+1)

    duration.cut.pfs = cut(pfs.p, breaks, right=FALSE)
    duration.freq.pfs = table(duration.cut.pfs)

    duration.cut.os = cut(os.p, breaks, right=FALSE)
    duration.freq.os = table(duration.cut.os)

    duration.cumfreq.pfs = cumsum(duration.freq.pfs)
    duration.cumfreq.os = cumsum(duration.freq.os)

    duration.cumfreq.pfs<-as.numeric(duration.cumfreq.pfs)
    duration.cumfreq.os<-as.numeric(duration.cumfreq.os)

  prediction.i<-c(duration.cumfreq.pfs,duration.cumfreq.os)}
  return(prediction.i) # Only intrested in PFS at the model

}
```

```
##############################################################################################
# gen_pred_cop – Generate predictions for OS and OS for a given distribution for TTP
##############################################################################################
gen_pred_cop<-function(.i,mat_copula){
  .matcop<-.ds<-prediction.i<-npar1<-npar2<-par1<-par2<-par3<-.coprnd<-NULL
  .matcop<-mat_copula[.i,]
  npar1<-.np[.name==.ds.pfsC] # TTP
  npar2<-.np[.name==.ds.osC] # OS

  par1<-.matcop[1:npar1] ;   par2<-.matcop[(npar1+1):(npar1+npar2)] ;   par3<-tail(.matcop,n=1)

  #Generate random sample from copula parameters
  .coprnd = rCopula(15000, normalCopula(par3));

  pfs.p<-qfunction_Vector(.coprnd[,1],.ds.pfsC,par1)
  os.p<-qfunction_Vector(.coprnd[,2],.ds.osC,par2)
  pfs.p<-pmin(pfs.p,os.p)

  #round pfs and os to higher value
  pfs.p<-ceiling(pfs.p)
  os.p<-ceiling(os.p)

  breaks = c(.sq,tail(.sq,1)+1)

  duration.cut.pfs = cut(pfs.p, breaks, right=FALSE)
  duration.freq.pfs = table(duration.cut.pfs)
   duration.cut.os = cut(os.p, breaks, right=FALSE)
  duration.freq.os = table(duration.cut.os)
  duration.cumfreq.pfs = cumsum(duration.freq.pfs)
  duration.cumfreq.os = cumsum(duration.freq.os)
  duration.cumfreq.pfs<-as.numeric(duration.cumfreq.pfs)
  duration.cumfreq.os<-as.numeric(duration.cumfreq.os)

  prediction.i<-cbind(duration.cumfreq.pfs,duration.cumfreq.os)
  return(prediction.i)
}

##############################################################################################
# myAIC_cop – Calculate AIC
##############################################################################################
myAIC_cop<-function(.i,mat_copula) {
  .ds<-np<-.mCop<-.nparC<-aic.cop<-NULL
  .ds<-.name[.i] ;  .matcop<-mat_copula[[.i]]
  if(is.na(.matcop[1])){aic.cop<-NA}else{

    .mCop<-colMeans(.matcop) ;  .mCop<-.mCop[1:.np[.i]] ;  .ds<-.name[.i] ;  .nparC<-.np[.i]

    aic.cop1<-dfunction(.ds,.mCop,pfs_data[,1])*pfs_data[,2]+((1-pfs_data[,2])*(1-
pfunction(.ds,.mCop,pfs_data[,1])))
    aic.cop<--log(aic.cop1);   aic.cop[which(aic.cop==Inf)]<-0;
    aic.cop<-sum(aic.cop);
    aic.cop<-2*(aic.cop)+2*(.nparC)  }
  return(aic.cop)
}

##############################################################################################
# qfunction_Vector – Generate survival time given a distribution, random number and parameters
##############################################################################################
qfunction_Vector<-function(RnD,.Distr,parX){
  .ds<-.param<-.rnd<-NULL;
  .ds<-.Distr; .param<-parX;.rnd<-RnD;
  if(.ds=="gengamma"){
    .q<-do.call(paste0("q",     .ds),     list(.rnd,.param[1],.param[2],.param[3]))}else     if
(.ds=="exp") {
      .q<-do.call(paste0("q", .ds), list(.rnd,.param[1]))} else if (.ds=="mySpline"){
        .q<-do.call(paste0("q",                                                          .ds),
list(.rnd,.param[1],.param[2],.param[3],.param[4],.param[5],.param[6]))} else if
  (.ds=="null"){           .q<-rep(NA,length(.LT))}else{.q<-do.call(paste0("q",          .ds),
list(.rnd,.param[1],.param[2]))}
  return(.q)}
```

```
##############################################################################################
# m.select.fit.copOS - Addition function to make sure that for the combination of TTP and OS -
OS is still within the data
##############################################################################################
m.select.fit.copOS<-function(dat.S,.pred.sv,.clinpl,.bound){
.dat<-.pred.SV<-.clp<-.bd<-.pred.SV1<-.pred.SV2<-.tmp.time<-.rmNA<-.stp1<-.stp2<-.stp3<-.cr3<-
.cst<-.km<-.kmU<-.kmL<-n.with<-p.with<-r.with<-NULL;
  .dat<-dat.S ;   .pred.SV<-.pred.sv ;   .clp<-unlist(.clinpl) ;   .bd<-.bound;   .pred.SV1<-
.pred.SV;  nRowX<-nrow(.pred.SV1)
   .clp2<-unlist(.clp[2]) ;     .clp1<-unlist(.clp[1]) ;   .bd2<-unlist(.bd[2]) ;     .bd1<-
unlist(.bd[1])
   .tmp.time<-min(.clp2, nRowX); #Constraint to avoid error (used for when debugging)
  #STEP 1 (a): Remove distribution that reaches 0 when expected survival around 10%
  .rmNA<-.pred.SV1[.tmp.time,];.rmNA[.rmNA<0.0001]<-NA
  #STEP 1 (b) If all distribution reaches 0, then remove constraint
  if (sum(is.na(.rmNA))==length(.rmNA)){.rmNA<-rep(1,length(.rmNA))};
  .rmNA[!is.na(.rmNA)]<-1;
  .rmNA<-t(matrix(rep(.rmNA,nRowX),ncol=nRowX))
  .stp1<-.pred.SV1 *.rmNA

  #STEP 2: Visual fit (based on fit to KM - within CI)
  .km<-survfit(Surv(.dat[,1],.dat[,2])~1,conf.int=TRUE);
  .surv<-.km$surv;   .kmU<-.km$upper;   .kmL<-.km$lower ;   .kmtime<-round(.km$time)
  # Constraint added because NA possible in KM at the end
  .cst<-min(sum(!is.na(.kmtime)),sum(!is.na(.kmL)),sum(!is.na(.kmU))) ;
  .kmU<-.kmU[1:.cst]; .kmL<-.kmL[1:.cst]; .kmtime<-.kmtime[1:.cst]
   .tmpkm<-.stp1[.kmtime+1,] #ADD +1 because start at 1 not 0
  .fL<-.tmpkm>=.kmL;   .fU<-.tmpkm<=.kmU;    n.with<-.fL+.fU;

  p.with<-colSums(n.with==2)/nrow(n.with);   r.with<-min(.mgin[1],max(p.with,na.rm=TRUE));
  if(r.with!=.mgin[1]){r.with<-r.with*0.9}
  s.with<-p.with>=r.with;                  s.with[s.with==FALSE]<-NA;                 s.with<-
t(matrix(rep(s.with,nRowX),ncol=nRowX))
  .stp2<-.stp1 *s.with

#STEP 3: Selection based on long-term prediction
  .clp2<-min(.clp2,.lgp1);   .tmp.pl<-.stp2[.clp2,];
  m.y<-abs(.tmp.pl-.clp1)-.mgin[2];m.y<-pmax(0,ceiling(m.y/0.02));m.y<-min(m.y,na.rm=TRUE)
  mgin.adj<-.mgin[2]+m.y*0.02;
  .cr3<-(between(.tmp.pl,(.clp1-mgin.adj),(.clp1+mgin.adj)));           .cr3[.cr3==FALSE]<-NA;
.cr3.tmp<-.cr3
  .cr3<-t(matrix(rep(.cr3.tmp,nRowX),ncol=nRowX));   .stp3<-.cr3
  return(.stp3)
}

##############################################################################################
# E.performance
##############################################################################################
E.performance <-function(m.dat=NULL,m.true=NULL){
  perform.e<-m.mean1<- m.mean2<- m.bias1<- m.bias2<- m.rmse1<-m.rmse2<-m.cover<-m.se<-NULL;
  s.dat<-colSums(m.dat)
  m.mean1<-  s.dat[1] ; #Mean (first value)
  m.mean2<-  mean(s.dat) ; #Mean (across all values)
  m.bias1<-  m.mean1- m.true ; #absolute bias around mean
  m.bias2<- m.mean2- m.true ; #absolute bias around probabilistic mean
  m.rmse1<- abs(m.true- m.mean1)^2; #RMSE around mean
  m.rmse2<- abs(m.true- m.mean2)^2 ; #RMSE around probabilistic mean
  m.lci1<-min(s.dat);m.uci1<-max(s.dat);m.cover<-between(m.true,m.lci1,m.uci1) ; #Coverage
  m.se<-sd(s.dat)/sqrt(length(s.dat)) ; #SE
  perform.e<-c(m.mean1, m.mean2, m.bias1, m.bias2, m.rmse1, m.rmse2, m.cover,m.se)
  return(perform.e)
}

##############################################################################################
# summaryperformance
##############################################################################################
summaryperformance<-function(.datO){
  .dat.output<- pfs.pef<- os.pef<- qaly.pef<- performanceall <-NULL;
  .dat.output<-.datO
  pfs.pef<-E.performance(.dat.output[[1]],                                  .truepfs);os.pef<-
E.performance(.dat.output[[2]],.trueos);
  qaly.pef<- E.performance(.dat.output[[3]], .trueqaly);
  performanceall <-c(pfs.pef ,os.pef ,qaly.pef);
```

```
  return(performanceall)
}

#############################################################################################
# sumLY - Smmarise LY and QALY - adjusted for general population mortality (in case)
#############################################################################################
sumLY<-function(x){
  a<-b<-c<-d<-NULL
  a<-x[1:.lgp1,] ;    b<-x[(.lgp1+1):(.lgp1*2),] ;   a<-a.S.gp(a) ;    b<-a.S.gp(b) ;    c<-
a*.qpfs+(b-a)*.qpd ;
d<-list(a,b,c)
}

#############################################################################################
# Start to Run the different nodels
#############################################################################################
Start <- Sys.time()
#Run PSM
Start1 <- Sys.time()
pfs_matrix<-gen.tn("pfs");       os_matrix<-gen.tn("os");       preps_matrix<-gen.tn("preps");
ttp_matrix<-gen.tn("ttp");
pfs_mod<-sapply(1:ncol(os_matrix[[1]]),psm.mod,os_matrix,pfs_matrix)
Stop1 <- Sys.time();RunTime1<-Stop1-Start1;
RunTime1

#Run 3 versions of STM
Start2 <- Sys.time()
STM1_mod<-stm.mod(pps_data_NoCov,pfs_matrix,preps_matrix)
Stop2 <- Sys.time();RunTime2<-Stop2-Start2; RunTime2
Start3 <- Sys.time()
STM2_mod<-stm.mod(pps_data_Log,pfs_matrix,preps_matrix)
Stop3 <- Sys.time();RunTime3<-Stop3-Start3; RunTime3
Start4 <- Sys.time()
STM3_mod<-stm.mod(pps_data_NoLog,pfs_matrix,preps_matrix)
Stop4 <- Sys.time();RunTime4<-Stop4-Start4; RunTime4

#Run MSM
Start5 <- Sys.time()
MSM_Mod<-MSM.mod()
Stop5 <- Sys.time();RunTime5<-Stop5-Start5; RunTime5

#Run Li model
Start6 <- Sys.time()
Li_Output<-sapply(1:(.nbBoot+1),bootLI,dataset)
Stop6 <- Sys.time();RunTime6<-Stop6-Start6; RunTime6

Start7 <- Sys.time()

#####################
# Only run when .Scen = 1
#####################

if (.Sc==1){
.dat1<-.clinP1<-.limX1<-PREDICTION1<-SELECTION<-copula_output<-predpfs<-pfs.i<-os.i<-
aic.cop.pfs<-predpfs2<-pfs_mat_select<-cop_output<-.ds.pfsC<-NULL;
.clinP1<-.plaus[3,]; .limX1<-NULL; .dat1<-os_data ;   PREDICTION1<-.fitdist(.dat1,1)
  SELECTION<-m.select.fit(.dat1,PREDICTION1,.clinP1,.limX1)
  .ds.osC<-.name[SELECTION[[1]]] # Select distribution for OS (based on independent fit)

  #Extract parameters and copula for the 7 distribution given a OS distribution
  copula_output<-sapply(.name[1:length(.name)],wrapper.copula,.ds.osC)

  if(sum(is.na(copula_output))==length(.name)){cop_output<-rep(NA,.lgp1*2);cop_output<-
rep(cop_output,.nbBoot+1);
cop_output<-t(matrix(cop_output,nrow=.nbBoot+1))}else{
    predpfs<-1-(sapply(1:length(.name),gen_pred_cop_pfs,copula_output)/15000)
    pfs.i<-predpfs[1:.lgp1,]
    os.i<-predpfs[(.lgp1+1):(.lgp1*2),]
    #Select OS first - to make sure OS still fit the data
    os.i2<-m.select.fit.copOS(.dat1,os.i,.clinP1,.limX1)
    pfs.i2<-pfs.i *os.i2
```

```
    aic.cop.pfs<-sapply(1:length(.name),myAIC_cop,copula_output);                aic.cop.pfs<-
aic.cop.pfs*pfs.i2[1,]
    predpfs2<-list(pfs.i2,aic.cop.pfs)
    select.pfs.copula<-m.select.fit(pfs_data,predpfs2,.plaus[1,],.plaus[3,])

    .ds.pfsC<-.name[select.pfs.copula[[.Sc]]]
    pfs_mat_select<-copula_output[[.ds.pfsC]];                                   pfs_mat_select<-
rbind(colMeans(pfs_mat_select),pfs_mat_select)
    cop_output<-1-(sapply(1:nrow(pfs_mat_select),gen_pred_cop,pfs_mat_select)/15000)
  } #End if function for NA
}else{
.dat1<-.dat2<-.ds.osC<-.ds.pfsC<- PREDICTION1<- PREDICTION2<- os.distT <- ttp.distT <- copuGEN
<- cop_output <-NULL
.dat1<-os_data; .dat2<-ttp_data
PREDICTION1<-.fitdist(.dat1,2); PREDICTION2<-.fitdist(.dat2,2)

os.distT<-which((min(PREDICTION1[[2]],na.rm=TRUE)==PREDICTION1[[2]])=="TRUE") ;
ttp.distT<-which((min(PREDICTION2[[2]],na.rm=TRUE)==PREDICTION2[[2]])=="TRUE") ;
.ds.osC<-.name[os.distT]; .ds.pfsC<-.name[ttp.distT]
copuGEN<-wrapper.copula(.ds.pfsC,.ds.osC); copuGEN<-rbind(colMeans(copuGEN),copuGEN)
cop_output<-1-(sapply(1:nrow(copuGEN),gen_pred_cop,copuGEN)/15000)
}

Stop7 <- Sys.time();RunTime7<-Stop7-Start7; RunTime7

##############################################################################################
# Summarise and extract outputs
##############################################################################################
.mod1<-sumLY(pfs_mod); .mod2<-sumLY(STM1_mod); .mod3<-sumLY(STM2_mod); .mod4<-sumLY(STM3_mod);
.mod5<-sumLY(MSM_Mod); .mod6<-sumLY(Li_Output);.mod7<-sumLY(cop_output)
#Extract Results
.perfomance1<-summaryperformance(.mod1);              .perfomance2<-summaryperformance(.mod2);
.perfomance3<-summaryperformance(.mod3);              .perfomance4<-summaryperformance(.mod4);
.perfomance5<-summaryperformance(.mod5); .perfomance6<-summaryperformance(.mod6);
.perfomance7<-summaryperformance(.mod7)
#Combine performance
.modall<-rbind(.perfomance1,.perfomance2,.perfomance3,.perfomance4,.perfomance5,.perfomance6,
.perfomance7);

#Extract fitted curves
lpfs1<-.mod1[[1]][,1];los1<-.mod1[[2]][,1];lpfs2<-.mod2[[1]][,1];los2<-.mod2[[2]][,1];lpfs3<-
.mod3[[1]][,1];
los3<-.mod3[[2]][,1];  lpfs4<-.mod4[[1]][,1];los4<-.mod4[[2]][,1];lpfs5<-.mod5[[1]][,1];los5<-
.mod5[[2]][,1];
lpfs6<-.mod6[[1]][,1];los6<-.mod6[[2]][,1]; lpfs7<-.mod7[[1]][,1];los7<-.mod7[[2]][,1];

km.pfs<-survfit(Surv(pfs_data[,1],        pfs_data[,2])~1,conf.int=TRUE);        km.surv.pfs<-
summary(km.pfs,time=.sq)$surv;
km.surv.pfs<-c(km.surv.pfs,rep(NA,.lgp1-length(km.surv.pfs)));
km.os<-survfit(Surv(os_data[,1],                    os_data[,2])~1,conf.int=TRUE);km.surv.os<-
summary(km.os,time=.sq)$surv;
km.surv.os<-c(km.surv.os,rep(NA,.lgp1-length(km.surv.os)));
#Extract Graph
graphX<-cbind(.sq,km.surv.pfs,lpfs1,lpfs2,lpfs3,lpfs4,lpfs5,                              lpfs6,
lpfs7,km.surv.os,los1,los2,los3,los4, los5, los6, los7)
write.table(.modall, paste0("performance", id, ".txt"), row.names = FALSE);
write.table(graphX,paste0("graph", id, ".txt"), row.names = FALSE); Stop <- Sys.time();RunTime<-
Stop-Start; RunTime
```

Appendix 12 : Further details regarding the implementation of methods under investigation (Chapter 8)

10.6.1.1.1 Implementation of the partitioned survival approach in this simulation study

The implementation of the PSM was relatively straightforward, with parametric distributions fitted directly to the data for PFS and OS within the censored dataset.

The process for selecting the parametric distributions for PFS and OS is described in Chapter 2. The minimum between PFS and OS was taken for PFS to avoid logical inconsistency.

10.6.1.1.2 Implementation of the simplified state-transition approach in this simulation study

Three variations are explored in this simulation study:

- STM using unadjusted PPS
- STM using PPS adjusted using log of TTP as a covariate
- STM using PPS adjusted using the TTP as a covariate (non-log scale)

The implementation of the STM in this simulation study follows the steps described in Chapter 5 for what was referred to as "Approach 3" (which encompasses "Approach 2"), i.e. the PFS curve is used directly, with the proportion of events that are deaths or progression in each cycle being calculated based on the number of people that are progression-free multiplied by the probability of death in a given cycle from the PrePS curve (assumed to be an exponential for simplicity).

As described in Chapter 5, an alternative approach to determine the proportion of PFS events that are death or progression (referred to as "Approach 1" in Chapter 5) could have been used. Whilst little difference is expected, it is recognised that it would have been interesting to explore all possible formulations. However, for pragmatic reasons, and model run-time, only one approach was explored.

In brief,

1. parametric distributions are first fitted to the PFS data to determine the proportion of patients who are progression-free at each cycle in the model,
2. a parametric distribution (exponential) is then fitted to the PrePS data to determine the probability of dying for patients who are progression-free in each model cycle,
3. from the PFS and PrePS curve, I calculate in each cycle how many patients progress or die prior to progression,

4. a parametric distribution is then fitted to the PPS data to determine to probability of dying after progression,
5. given that non-constant transition (i.e. time-varying parametric distributions) are considered for PPS, the PD health state is implemented in the model using a series of tunnel states to allow the death probability to be conditioned on the time since experiencing progression.
6. the PPS curve is then applied to the proportion of PFS events that are progression to determine the time to death in people who progressed.

Similar to the unadjusted PPS, whereby a parametric distribution is fitted to data for PPS, for the methods using the adjusted PPS, a parametric distribution is fitted to the data for PPS, but with TTP included as a covariate in the statistical model. Given that the general STM process included tunnel states for the progressive-disease health state, including the dependence between PPS and TTP in the economic model was straightforward, as the probability of death was allowed to vary according to time. For new patients progressing, the PPS curve according to the time at which the patient progressed was used. For instance, assuming the PPS distribution follows an exponential distribution (constant probability) for simplicity, an individual who progressed at cycle 1 will have a different probability of dying compared with an individual who progressed at cycle 20.

10.6.1.1.3 Implementation of the multi-state model in this simulation study

The implementation of the MSM in this simulation study follows the steps described in Chapter 4 using the `mssample` function (part of the `mstate` package). Compared with the simplified STM, PFS is not used directly, but estimated by modelling the two competing transitions (TTP and PrePS) and combining them under a competing risk framework.

The implementation was relatively straightforward. Parametric distributions were fitted to the data for TTP, PrePS and PPS. These were then transformed into the cumulative hazard form and used directly within the `mssample` function (corrected version) as described in Chapter 4 to generate predictions.

As described in Chapter 4, the multi-state approach using the `mssample` function uses a simulation approach, and thus, the more patients are sampled, the more accurate/stable the prediction. However, increasing the number of patients sampled, increases model run-time. Similarly, the time horizon and time step used (cycle length) both have an effect on the model run-time. 10,000 patients were assumed in the `mssample` function in order to increase precision.

In the simulation study, results are generated both deterministically and probabilistically. Therefore, it is important to consider model run-time. To estimate results probabilistically, 1,000 iterations of the MSM must be generated for each 1,000 datasets within each of the 54 scenarios investigated. Considering 5,000 patients (to reduce sampling variation in predictions), a cycle length of a week and a time horizon of about 19 years (corresponding to 1,000 weeks), a single model takes slightly less than half a minute to run, although this could vary depending on the performance of the computer. Based on the above estimation, using the `mssample` function to estimate the probabilistic prediction in a single dataset for a single scenario takes approximately 8-9 hours. It is therefore impractical to use the `mssample` function when generating prediction probabilistically in this simulation study given the large number of model runs required (1,000 datasets for each 54 scenarios). It should be noted that it is possible to reduce the number of patients sampled to 1,000 to reduce model run-time. Using this approach, a single model run takes less than 10 seconds, and coverage can be estimated in around 2 hours. Whilst this is perhaps more manageable, this still leads to an unmanageable model run-time. Furthermore, reducing the number of patients decreases precision. An alternative approach was used to approximate the coverage of the MSM to provide a trade-off between running time and precision.

Consequently, an alternative approach (referred to here as the "approximation method") is used when running the MSM probabilistically, derived from the steps used in the `mssample` function described in Chapter 4.

Compared with the `mssample` function, this approximation method follows a cohort approach, similar to the simplified STM previously described, with:

- PFS estimated by combining the hazards from the two cause-specific events (TTP and PrePS) rather than using PFS directly from the trial.
- The proportion of PFS events that are death or progression in each cycle given by the contribution of the two competing events to the cumulative hazard. In the `mssample` function, a sampling approach is used to determine whether an event is death or progression. Because a cohort approach is used here, in each cycle, the proportion of PFS events which are death or progression was determined by comparing: (1) the sum of the hazards from the two cause-specific events (TTP and PrePS) and (2) the individual hazards for each events at this time point. For instance, assuming the hazards for TTP and PrePS are 0.02 and 0.001 for a given cycle respectively, I assumed that 95% of events would be progression and 5% would be death.

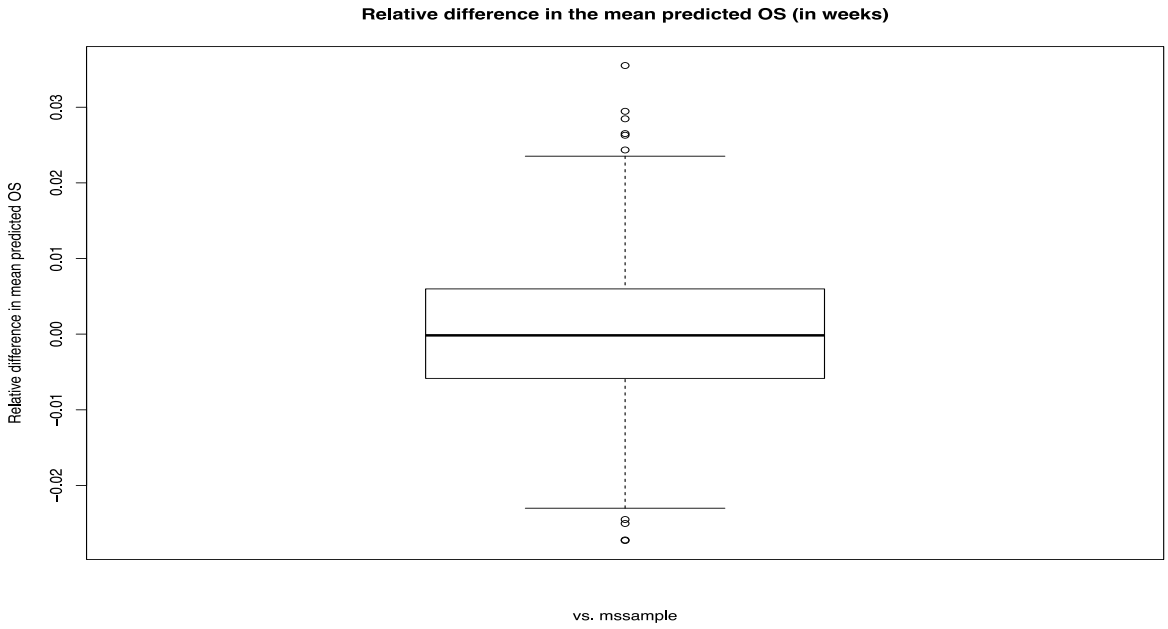
In order to validate that the approximation method described here provides broadly similar results when compared with the `mssample` function, predictions for OS were generated and compared using both

approaches when applied to the BC dataset . For simplicity, all transitions (TTP, PrePS and PPS) were assumed to follow Weibull distributions. Input parameter values were drawn by random sampling for each distribution repeated 1,000 times using the `normboot.flexsurvreg` function in R.

Five thousand patients were used in the `mssample` function to increase precision, with the mean calculated OS ranging from 110.5 to 161.4 weeks. Estimates from the approximation method were compared with the generated mean OS taken from the `mssample` function for the 1,000 iterations. Overall, absolute differences in predicted OS using the `mssample` function and the approximation method were very small (median: 0.007; Range: -4.18 to 5.41 weeks). When considering relative differences, estimates using the approximation method were generally within +/- 0.5% of the estimates generated using the `mssample` function (Figure 75).

In terms of model run-time, in this particular example, the probabilistic results were estimated in less than 5 minutes using the approximation method compared with more than 8 hours using the `mssample` function (this time was however reduced to 2 hours if 1,000 patients were sampled instead of 5,000; however, some precision was lost). It should be noted that this approximation method is not and cannot be a replacement for the `mssample` function given that it can only provide an approximation of the mean lifetime survival (which is intended for the aim of this thesis).

Figure 75 : Comparison of prediction for OS using the multi-state function and approximation method

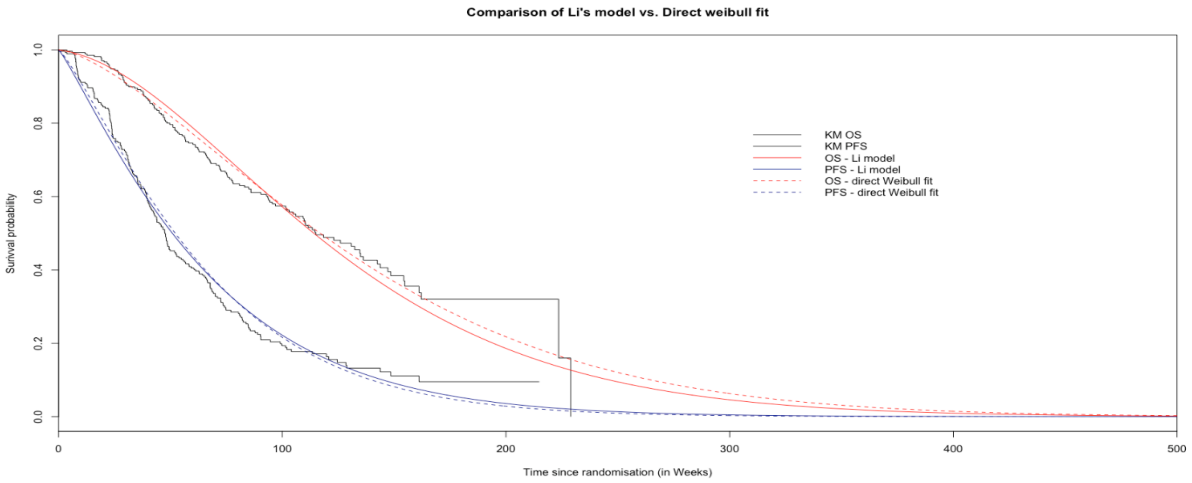


10.6.1.1.4 Implementation of the Li model in this simulation study

As described in Chapter 6, the model developed by Li *et al* (2015) uses Weibull distributions for every transition. Whilst the R program developed by the authors was not available in the public domain, it was kindly provided by the authors following a request. Because the model is restricted to the Weibull distribution, little to no adaptation is required, with the only amendment relating to the extraction of information on PFS as well as OS. The function reports only point estimates. The implementation of the Li model is relatively straightforward.

It should be noted that the Li model returns an error when a person has a recorded OS event but not a recorded PFS event (censored). Whilst this is rare, it is possible, because progression is recorded at the time of the visit and patients could be lost of follow-up for progression, whilst the time of death is recorded and known with certainty. Consequently, when using the Li model, an assumption had to be made that TTP was equal to OS when PFS was not recorded, but OS was. This is a possible limitation. For illustration, predictions using the Li model when applied to the BC dataset are shown in Figure 76 against assuming independent Weibull distributions for each transition.

Figure 76 : Estimation using the Li model



In the simulation study, results are generated both deterministically and probabilistically. Only the point estimate is reported when running the model proposed by Li *et al* (2015). Consequently, probabilistic results in this simulation study were generated by bootstrapping the sample data and fitting the model to each bootstrapped dataset. This is different to other approaches where the probabilistic estimate is generated by sampling parameters. Whilst this is a possible limitation, the impact on the interpretation of results is likely to be minimal.

10.6.1.1.5 Implementation of the Fu model in this simulation study

As described in Chapter 6, the model developed by Fu *et al* (2013) allows the joint modelling of PFS and OS under a semi-competing risk framework using a Gaussian copula. As described in Chapter 6, whilst there exist other copulas and a Bayesian framework was used, a key benefit of the approach proposed by Fu *et al* (2013) is that the program is directly available in the appendices to the paper, and therefore can be used directly in health economics in a reproducible and transparent way. A Gaussian copula is also considered generally appropriate when modelling PFS and OS as it is tail independent.

In this section, I will describe how the Fu model was implemented in this simulation study using the BC dataset.

- Step 1: Setting up the data

Prior to the implementation of the Fu model, data must be set up in the appropriate format. Notably, data are required on TTP and OS, rather than PFS. In other words, in our dataset, a new variable needs to be created for TTP whereby in patients with a PFS event for whom PFS is equal to OS, the PFS event needs to be censored.

It is also easier if variables for TTP and OS are renamed in a similar way as suggested by the authors and data follow the same order. Failing that, the variable names need to be replaced in the R program to ensure that the appropriate variables are selected. Similarly, additional changes would be required in the R program to ensure that the correct variables are picked up if the order of variables is different. Consequently, for each dataset examined in the simulation, a new dataset was created to mimic the format used by the author.

- Step 2: Select the marginal distribution for TTP and OS

The authors assumed that the marginal survival distributions for both TTP and OS follow exponential distributions and the R code provided reflects this. The exponential distribution can be restrictive and therefore analysts may wish to explore alternative marginal distributions. Any combination of marginal distribution is possible; however, this may not be feasible or practical given run-time and may be inappropriate if some distributions are clearly implausible. Considering the seven distributions for survival endpoints considered in the simulation study (exponential, Weibull, Gompertz, log-normal, log-logistic, gamma, and generalised gamma), 49 possible combinations of TTP and OS are possible.

The estimation of the values for both the copula and marginal survival distribution parameters can be time-consuming (as the samples of the posterior are reported, rather than point estimate) for each

possible combination, with the run-time for a given combination of marginal survival distribution for TTP and OS conditional on the sample size of the data, the number of burn-in, thinning interval, optimisation method, number of samples of the posterior distribution and marginal survival distribution used. Exploratory analysis indicated a model run-time ranging from 5-10 minutes to more than 30 minutes for a given combination of marginal survival distribution for TTP and OS. Assuming a model run-time of 10 minutes and 49 combinations (based on seven parametric distributions), it would take around 8 hours to estimate all possible combinations of marginal survival distribution for TTP and OS for a single sample dataset. Given that 54 scenarios (which contain 1,000 datasets) are examined, examining all possible combination was not considered to be practical for this simulation study. Consequently, I limited the number of combinations examined to reduce model run-time.

Given that OS (terminal event) censors TTP (non-terminal event) under a semi-competing risk framework, the estimate for OS is likely to be close to the estimate of OS using the independent fit to the data. The distribution for OS should not be significantly affected by TTP (unless there are lot of deaths pre-progression). Consequently, to limit the number of combinations of TTP and OS, the marginal distribution for OS in the Fu model could be narrowed down to the OS distributions that are considered plausible when looking at the independent fit to the data, if model running-time is an issue. It should be noted that this is a simplification.

This is also consistent with the objective in this simulation study - to assess whether jointly modelling PFS and OS meaningfully improves the estimation of health state sojourn time and QALYs compared with the independent model.

To limit the number of distributions examined, the marginal survival distribution for OS was assumed to be the same as that selected for the independent model (the PSM). Conversely, no assumption was made regarding the marginal survival distribution for TTP. Consequently, prior to selecting the most plausible model, seven combinations are run in each dataset for this simulation study (for the seven marginal survival distribution of TTP for the selected marginal survival distribution for OS; e.g. Weibull). This is a possible limitation and therefore results need to be interpreted with some caution.

- Step 3: Amending the R program to use a specified distribution

The R program provided by the authors can be adapted, so that the marginal survival distributions for TTP and OS other forms than the exponential distribution.

For the purpose of this thesis, and for the sake of efficiency, the R program was amended to allow for the following forms: exponential, Weibull, Gompertz, log-normal, log-logistic, gamma, and generalised gamma. This involved changing the density and distribution function. Whilst a function could have been created for all 49 possible combinations of TTP and OS, this would have been less efficient, less transparent, and more difficult to debug. Furthermore, although not presented here, the marginal distribution could also take more complex form, such as a spline model; the inclusion of a spline model with one knot was explored for feasibility but was not included in the final simulation.

In addition to key amendments to the density and distribution function for the marginal distributions when calculating the likelihood function (e.g. using the Gompertz instead of the exponential), it was necessary to transform the parameters of the marginal distribution when using distributions others than the original exponential distributions. Indeed, some parametric distributions allow negative parameters, which is incompatible with the algorithm used to estimate parameters. Similarly, the values taken by some parameters could also be high, making the estimation difficult. Consequently, supportive functions were created in order to transform parameters when used in conjunction with the main function.

- Step 4: Ensure that the model converges

The Fu model uses the random walk Metropolis Hastings algorithm to estimate parameters which maximise the likelihood. It should also be noted that the model may sometimes fail to converge, but that using a different optimisation method may facilitate convergence (for example, Nelder-Mead). The analyst can also modify the number of posterior samples, the burn-in number and thinning interval.

Furthermore, the choice of initial parameters is important, as the model will fail to converge if inappropriate or unrealistic initial values are used. For simplicity, initial values could be those estimated using the independent model.

- Step 5: Extract output

Once the data have been correctly set up, and the R program has been amended for the marginal distributions for TTP and OS to take on the specified form, the R program can be run. As the approach uses a Bayesian framework, the function reports only the samples of the posterior distribution for both: (a) the parameters of the marginal distributions of TTP and OS, and (b) the copula parameter, rather than the point estimate. Both deterministic and probabilistic results are considered in this simulation study. For the purpose of this simulation study, the point estimate was approximated by taking the median from the posterior distribution to account for potential skewness in the sample. It should be noted that when considering marginal distributions with more than one parameter (such as the Weibull distribution), parameters are correlated within each sample of the posterior distribution.

- Step 6: Generate TTP and OS

Once samples of the posterior distribution have been estimated for the parameters of the marginal distribution and the copula, it is then possible to generate predictions for TTP and OS by sampling TTP and OS given the copula parameter. As recommended by the authors, a large number of samples (e.g. 10,000) is required in order to reduce the influence of Monte Carlo sampling error. Consequently, predictions were generated by sampling 10,000 individuals.

- Step 7: Define PFS time

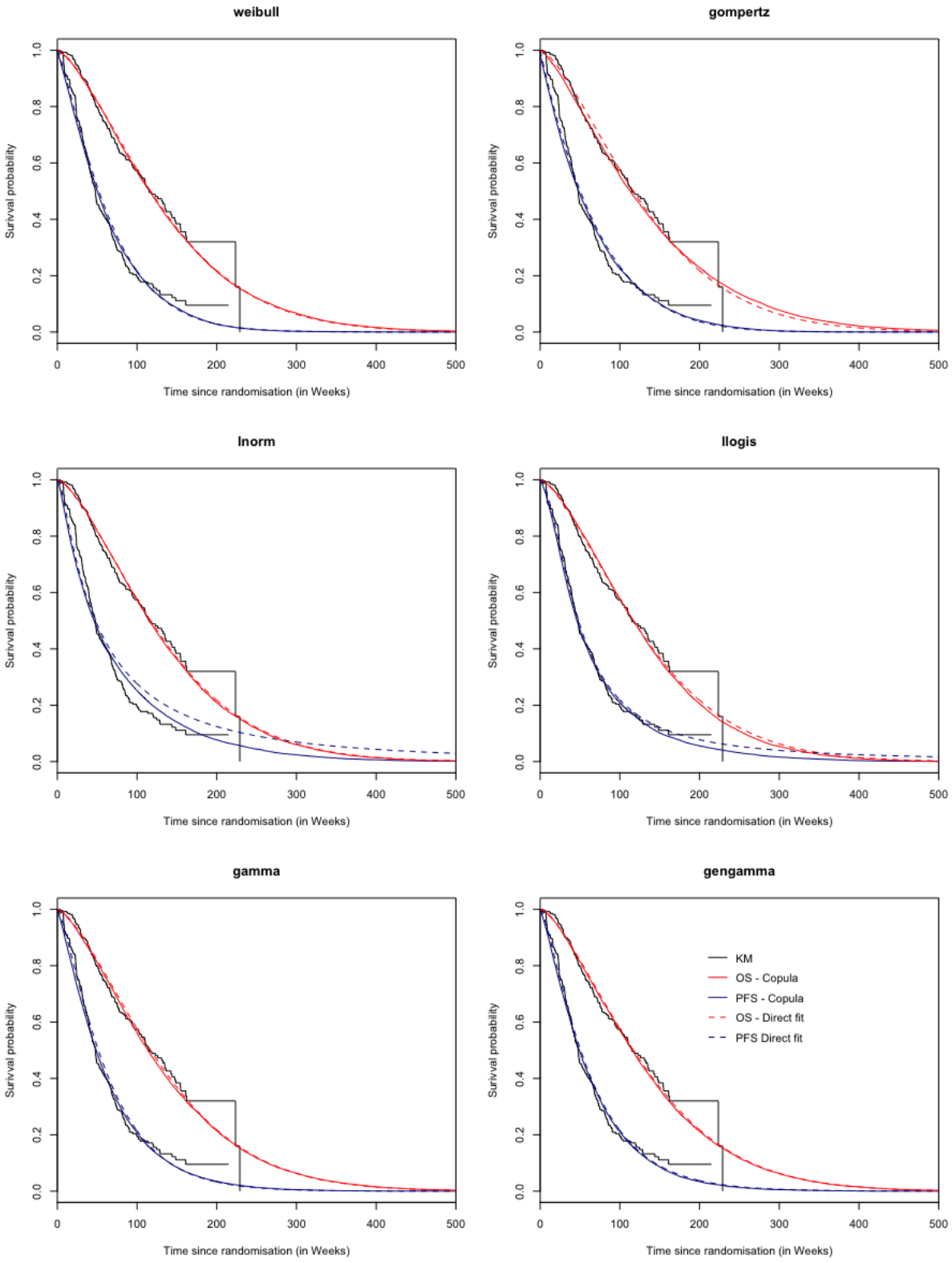
PFS is defined as the minimum of sampled time between TTP and OS.

- Step 8: Summarise information

The time to PFS and OS can then be summarised in terms of health state occupancy.

An example of the Fu model implementation is presented using the BC dataset. The marginal distribution for OS was assumed to follow a Weibull distribution, with the marginal distribution for TTP following 6 possible forms (Weibull, Gompertz, log-normal, log-logistic, gamma, and generalised gamma). The exponential is not shown here. Predictions are summarised in Figure 77. For transparency, the independent fit to OS (Weibull) is plotted alongside prediction from the Fu model. The independent fit assuming PFS follows the same distribution as TTP is also presented. However, there a limitation with this comparison as assuming that the marginal distribution for TTP follows a particular distribution does not necessarily mean that PFS would follow the same distribution.

Figure 77 : Estimation using Fu model assuming the marginal distribution for OS to follow Weibull



Appendix 13 : Full result simulation study

Table 23 : Results for simulated scenario 1

	Scenario1	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	57.07	59.45	59.57	59.22	56.33	46.30	49.23	56.75	60.21	60.32	59.97	55.60	46.50	49.56
PFS	Bias	-12.11	-9.73	-9.61	-9.96	-12.85	-22.88	-19.95	-12.43	-8.97	-8.86	-9.21	-13.58	-22.68	-19.62
PFS	MC SE	0.16	0.21	0.21	0.20	0.18	0.11	0.14	0.15	0.22	0.22	0.21	0.20	0.11	0.15
PFS	Rel Bias	-17.5%	-14.1%	-13.9%	-14.4%	-18.6%	-33.1%	-28.8%	-18.0%	-13.0%	-12.8%	-13.3%	-19.6%	-32.8%	-28.4%
PFS	empSE	4.93	6.40	6.51	6.17	5.67	3.33	4.36	4.73	6.64	6.74	6.39	6.08	3.37	4.48
PFS	MC SE	0.11	0.15	0.15	0.14	0.13	0.08	0.10	0.11	0.15	0.15	0.15	0.14	0.08	0.10
PFS	MSE	170.99	135.52	134.73	137.29	197.29	534.45	416.84	176.83	124.41	123.84	125.70	221.33	525.56	405.04
PFS	MC SE	3.71	3.82	3.85	3.87	4.37	4.86	5.50	3.65	3.68	3.70	3.73	5.23	4.88	5.52
PFS	Rel P	0.00%	-40.65%	-42.59%	-36.17%	-24.37%	119.85%	27.76%	0.00%	-49.24%	-50.83%	-45.33%	-39.52%	96.56%	11.28%
PFS	ModelSE	5.24	6.82	6.87	6.72	7.32	3.48	4.56	5.24	6.82	6.87	6.72	7.32	3.48	4.56
PFS	Cover2	0.35	0.59	0.59	0.58	0.44	0.00	0.05	0.33	0.62	0.63	0.62	0.42	0.00	0.06
PFS	Cover1	0.80	0.93	0.93	0.93	0.86	0.13	0.51	0.80	0.93	0.93	0.93	0.86	0.13	0.51
OS	Mean	90.54	103.81	93.26	108.39	101.13	83.14	94.37	93.47	114.62	98.84	114.29	110.45	83.91	97.04
OS	Bias	-8.49	4.78	-5.77	9.37	2.10	-15.89	-4.66	-5.55	15.60	-0.19	15.26	11.42	-15.12	-1.99
OS	MC SE	0.25	0.30	0.31	0.51	0.30	0.18	0.30	0.71	1.13	1.02	0.74	1.26	0.19	0.34
OS	Rel Bias	- 0.09	0.05 -	0.06	0.09	0.02 -	0.16 -	0.05	- 0.06	0.16 -	0.00	0.15	0.12 -	0.15 -	0.02
OS	empSE	7.72	9.14	9.56	15.80	9.14	5.61	9.10	21.86	34.82	31.48	22.91	38.86	5.82	10.32
OS	MC SE	0.18	0.21	0.22	0.36	0.21	0.13	0.21	0.50	0.80	0.72	0.53	0.89	0.13	0.24
OS	MSE	131.57	106.29	124.46	337.06	87.93	283.86	104.35	508.15	1454.43	990.09	757.10	1638.96	262.42	110.41
OS	MC SE	0.18	0.21	0.22	0.36	0.21	0.13	0.21	245.94	326.14	297.81	112.98	401.20	5.59	4.21
OS	Rel P	0.00%	-28.61%	-34.71%	-76.11%	-28.69%	89.50%	-27.94%	0.00%	-60.59%	-51.79%	-8.95%	-68.36%	1312.50%	348.43%
OS	ModelSE	24.70	50.78	34.64	32.51	52.45	7.03	16.50	24.70	50.78	34.64	32.51	52.45	7.03	16.50
OS	Cover2	0.90	0.99	0.83	0.98	0.99	0.38	0.95	0.92	0.98	0.87	0.99	0.99	0.40	0.96
OS	Cover1	0.98	1.00	1.00	1.00	1.00	0.91	0.99	0.98	1.00	1.00	1.00	1.00	0.91	0.99
QALYs	Mean	62.39	69.74	64.50	71.96	67.46	55.46	61.95	63.76	75.38	67.52	75.13	71.90	55.90	63.39
QALYs	Bias	-7.88	-0.53	-5.77	1.69	-2.80	-14.81	-8.31	-6.51	5.11	-2.75	4.87	1.63	-14.36	-6.88
QALYs	MC SE	0.15	0.19	0.18	0.29	0.18	0.11	0.17	0.37	0.58	0.52	0.40	0.64	0.12	0.19
QALYs	Rel Bias	- 0.11 -	0.01 -	0.08	0.02 -	0.04 -	0.21 -	0.12	- 0.09	0.07 -	0.04	0.07	0.02 -	0.20 -	0.10
QALYs	empSE	4.63	5.76	5.59	8.90	5.55	3.48	5.24	11.41	17.94	16.12	12.18	19.84	3.59	5.86
QALYs	MC SE	0.11	0.13	0.13	0.20	0.13	0.08	0.12	0.26	0.41	0.37	0.28	0.46	0.08	0.13
QALYs	MSE	83.48	33.37	64.52	82.07	38.63	231.38	96.54	172.36	347.53	267.27	171.93	395.94	219.17	81.65
QALYs	MC SE	2.41	1.41	2.09	3.82	1.54	3.30	2.87	61.87	79.44	72.86	26.44	95.98	3.30	2.64
QALYs	Rel P	0.00%	-35.22%	-31.41%	-72.93%	-30.33%	76.99%	-21.85%	0.00%	-59.55%	-49.93%	-12.29%	-66.94%	910.34%	279.26%
QALYs	ModelSE	12.82	25.72	17.81	16.99	26.73	4.23	9.07	12.82	25.72	17.81	16.99	26.73	4.23	9.07
QALYs	Cover2	0.78	0.97	0.80	0.96	0.96	0.11	0.80	0.80	0.97	0.83	0.97	0.95	0.13	0.83
QALYs	Cover1	0.96	0.99	1.00	1.00	1.00	0.70	0.99	0.96	0.99	1.00	1.00	1.00	0.70	0.99

Table 24 : Results for simulated scenario 2

	Scenario2	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	54.84	59.22	58.94	59.02	55.58	46.97	48.19	54.38	60.05	59.77	59.82	54.75	47.15	48.66
PFS	Bias	-14.34	-9.96	-10.24	-10.16	-13.60	-22.21	-20.99	-14.80	-9.13	-9.41	-9.36	-14.43	-22.03	-20.52
PFS	MC SE	0.15	0.19	0.19	0.20	0.16	0.11	0.15	0.15	0.20	0.20	0.21	0.17	0.11	0.15
PFS	Rel Bias	-20.7%	-14.4%	-14.8%	-14.7%	-19.7%	-32.1%	-30.3%	-21.4%	-13.2%	-13.6%	-13.5%	-20.9%	-31.8%	-29.7%
PFS	empSE	4.72	5.78	5.86	6.01	5.07	3.27	4.65	4.51	6.16	6.23	6.33	5.35	3.32	4.74
PFS	MC SE	0.11	0.13	0.13	0.14	0.12	0.08	0.11	0.10	0.14	0.14	0.15	0.12	0.08	0.11
PFS	MSE	227.73	132.61	139.10	139.39	210.55	504.01	462.05	239.22	121.13	127.43	127.67	236.73	496.20	443.48
PFS	MC SE	4.31	3.75	3.88	3.91	4.28	4.66	6.33	4.27	3.62	3.75	3.78	4.99	4.69	6.30
PFS	Rel P	0.00%	-33.52%	-35.23%	-38.44%	-13.58%	107.80%	3.02%	0.00%	-46.40%	-47.72%	-49.37%	-29.10%	84.36%	-9.54%
PFS	ModelSE	4.46	7.19	7.14	6.96	7.45	3.58	4.27	4.46	7.19	7.14	6.96	7.45	3.58	4.27
PFS	Cover2	0.18	0.59	0.57	0.57	0.38	0.00	0.03	0.16	0.63	0.61	0.61	0.36	0.00	0.04
PFS	Cover1	0.55	0.93	0.93	0.93	0.84	0.19	0.44	0.55	0.93	0.93	0.93	0.84	0.19	0.44
OS	Mean	71.58	84.39	76.75	91.22	80.78	68.28	74.25	72.75	87.19	80.63	94.33	83.45	68.60	75.89
OS	Bias	-14.78	-1.96	-9.60	4.87	-5.57	-18.07	-12.10	-13.60	0.84	-5.72	7.98	-2.90	-17.75	-10.46
OS	MC SE	0.20	0.22	0.19	0.50	0.20	0.13	0.21	0.23	0.76	0.81	0.62	0.91	0.13	0.24
OS	Rel Bias	- 0.17	- 0.02	- 0.11	0.06	- 0.06	- 0.21	- 0.14	- 0.16	0.01	- 0.07	0.09	- 0.03	- 0.21	- 0.12
OS	empSE	6.21	6.78	5.82	15.38	6.27	3.86	6.57	7.04	23.53	24.85	19.15	28.07	3.95	7.47
OS	MC SE	0.14	0.16	0.13	0.35	0.14	0.09	0.15	0.16	0.54	0.57	0.44	0.64	0.09	0.17
OS	MSE	256.81	49.71	126.06	259.91	70.31	341.40	189.67	234.45	553.88	649.51	430.07	795.22	330.78	165.18
OS	MC SE	0.14	0.16	0.13	0.35	0.14	0.09	0.15	5.94	188.67	178.86	66.88	215.07	4.51	4.90
OS	Rel P	0.00%	-16.13%	13.64%	-83.71%	-2.10%	158.11%	-10.83%	0.00%	-91.05%	-91.97%	-86.49%	-93.71%	217.83%	-11.21%
OS	ModelSE	9.80	30.33	33.30	26.10	38.33	4.57	9.92	9.80	30.33	33.30	26.10	38.33	4.57	9.92
OS	Cover2	0.46	0.92	0.66	0.91	0.89	0.05	0.61	0.49	0.92	0.70	0.92	0.88	0.06	0.67
OS	Cover1	0.87	1.00	1.00	0.99	0.99	0.56	0.94	0.87	1.00	1.00	0.99	0.99	0.56	0.94
QALYs	Mean	52.24	59.96	56.06	63.32	57.07	48.23	51.58	52.69	61.61	58.24	65.11	58.15	48.45	52.54
QALYs	Bias	-11.69	-3.97	-7.87	-0.61	-6.86	-15.70	-12.35	-11.24	-2.32	-5.69	1.18	-5.78	-15.49	-11.39
QALYs	MC SE	0.14	0.16	0.14	0.29	0.14	0.09	0.14	0.15	0.40	0.42	0.34	0.47	0.09	0.15
QALYs	Rel Bias	- 0.18	- 0.06	- 0.12	- 0.01	- 0.11	- 0.25	- 0.19	- 0.18	- 0.04	- 0.09	0.02	- 0.09	- 0.24	- 0.18
QALYs	empSE	4.28	4.79	4.28	8.79	4.30	2.77	4.36	4.55	12.38	12.87	10.52	14.42	2.83	4.77
QALYs	MC SE	0.10	0.11	0.10	0.20	0.10	0.06	0.10	0.10	0.28	0.30	0.24	0.33	0.06	0.11
QALYs	MSE	154.90	38.69	80.23	77.54	65.59	254.09	171.46	147.03	158.50	197.82	111.98	241.13	247.76	152.42
QALYs	MC SE	3.24	1.42	2.00	3.00	1.86	2.80	3.48	3.25	45.69	42.78	15.39	50.65	2.82	3.42
QALYs	Rel P	0.00%	-20.41%	0.05%	-76.33%	-1.20%	138.32%	-3.72%	0.00%	-86.47%	-87.48%	-81.26%	-90.02%	159.89%	-8.99%
QALYs	ModelSE	5.58	15.79	17.21	14.01	19.80	3.23	5.87	5.58	15.79	17.21	14.01	19.80	3.23	5.87
QALYs	Cover2	0.29	0.85	0.62	0.85	0.75	0.01	0.33	0.31	0.86	0.65	0.87	0.72	0.01	0.38
QALYs	Cover1	0.72	0.98	0.99	0.99	0.97	0.38	0.87	0.72	0.98	0.99	0.99	0.97	0.38	0.87

Table 25 : Result for simulated scenario 3

	Scenario3	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	51.51	58.70	58.80	58.75	54.36	47.47	46.50	51.11	59.55	59.65	59.59	53.97	47.59	46.73
PFS	Bias	-17.67	-10.48	-10.38	-10.43	-14.82	-21.71	-22.68	-18.07	-9.63	-9.53	-9.59	-15.21	-21.59	-22.45
PFS	MC SE	0.15	0.18	0.18	0.20	0.13	0.11	0.18	0.14	0.19	0.19	0.21	0.14	0.11	0.18
PFS	Rel Bias	-25.5%	-15.2%	-15.0%	-15.1%	-21.4%	-31.4%	-32.8%	-26.1%	-13.9%	-13.8%	-13.9%	-22.0%	-31.2%	-32.5%
PFS	empSE	4.58	5.56	5.55	6.02	4.16	3.32	5.56	4.41	5.91	5.88	6.34	4.27	3.35	5.61
PFS	MC SE	0.11	0.13	0.13	0.14	0.10	0.08	0.13	0.10	0.14	0.13	0.15	0.10	0.08	0.13
PFS	MSE	333.27	140.83	138.53	145.06	236.96	482.22	545.31	345.88	127.56	125.30	132.11	249.60	477.17	535.60
PFS	MC SE	5.15	3.87	3.80	4.02	4.05	4.64	7.91	5.08	3.73	3.66	3.88	4.21	4.66	7.86
PFS	Rel P	0.00%	-32.07%	-31.85%	-42.03%	21.70%	90.37%	-32.13%	0.00%	-44.27%	-43.68%	-51.58%	6.65%	72.96%	-38.15%
PFS	ModelSE	3.90	7.22	7.19	7.13	6.85	3.71	4.02	3.90	7.22	7.19	7.13	6.85	3.71	4.02
PFS	Cover2	0.07	0.58	0.59	0.58	0.27	0.00	0.04	0.06	0.62	0.62	0.61	0.27	0.00	0.05
PFS	Cover1	0.29	0.92	0.92	0.92	0.84	0.22	0.36	0.29	0.92	0.92	0.92	0.84	0.22	0.36
OS	Mean	58.58	70.43	67.36	80.31	66.22	58.24	61.66	59.44	71.48	69.46	82.11	66.30	58.38	62.43
OS	Bias	-19.71	-7.85	-10.92	2.03	-12.06	-20.04	-16.63	-18.84	-6.81	-8.83	3.82	-11.99	-19.90	-15.86
OS	MC SE	0.17	0.18	0.17	0.51	0.13	0.11	0.18	0.19	0.30	0.66	0.52	0.28	0.11	0.19
OS	Rel Bias	-0.25	-0.10	0.14	0.03	-0.15	-0.26	0.21	-0.24	-0.09	0.11	0.05	-0.15	0.25	0.20
OS	empSE	5.21	5.39	5.14	15.67	4.02	3.43	5.52	5.94	9.29	20.32	15.88	8.53	3.47	5.91
OS	MC SE	0.12	0.12	0.12	0.36	0.09	0.08	0.13	0.14	0.21	0.47	0.36	0.20	0.08	0.14
OS	MSE	415.66	90.72	145.73	249.51	161.63	413.43	306.97	390.34	132.64	490.51	266.66	216.46	408.16	286.44
OS	MC SE	0.12	0.12	0.12	0.36	0.09	0.08	0.13	6.78	43.99	326.25	16.13	42.94	4.45	6.02
OS	Rel P	0.00%	-6.60%	2.93%	-88.94%	68.42%	130.64%	-10.71%	0.00%	-59.17%	-91.46%	-86.03%	-51.58%	193.48%	0.80%
OS	ModelSE	7.14	14.02	19.15	18.10	14.23	3.93	6.86	7.14	14.02	19.15	18.10	14.23	3.93	6.86
OS	Cover2	0.16	0.69	0.54	0.82	0.51	0.00	0.26	0.17	0.72	0.58	0.85	0.49	0.00	0.29
OS	Cover1	0.51	0.98	0.94	0.98	0.95	0.33	0.75	0.51	0.98	0.94	0.98	0.95	0.33	0.75
QALYs	Mean	44.74	52.83	51.32	57.78	49.42	43.36	44.78	45.05	53.60	52.62	58.93	49.34	43.47	45.23
QALYs	Bias	-15.16	-7.07	-8.58	-2.12	-10.48	-16.53	-15.12	-14.84	-6.29	-7.27	-0.97	-10.56	-16.43	-14.67
QALYs	MC SE	0.13	0.14	0.14	0.29	0.10	0.09	0.13	0.14	0.19	0.35	0.30	0.16	0.09	0.14
QALYs	Rel Bias	-0.25	-0.12	0.14	0.04	-0.17	-0.28	0.25	-0.25	-0.11	0.12	0.02	-0.18	0.27	0.24
QALYs	empSE	3.92	4.26	4.17	8.95	3.12	2.68	4.05	4.19	5.82	10.75	9.14	4.94	2.70	4.28
QALYs	MC SE	0.09	0.10	0.10	0.21	0.07	0.06	0.09	0.10	0.13	0.25	0.21	0.11	0.06	0.10
QALYs	MSE	245.11	68.13	90.90	84.57	119.53	280.48	245.00	237.85	73.46	168.35	84.41	135.83	277.15	233.40
QALYs	MC SE	3.79	1.88	2.28	2.96	2.07	2.85	3.85	3.89	10.37	80.53	3.73	10.00	2.86	3.89
QALYs	Rel P	0.00%	-15.10%	-11.28%	-80.79%	58.09%	115.26%	-6.23%	0.00%	-48.18%	-84.80%	-78.98%	-28.02%	140.76%	-4.22%
QALYs	ModelSE	4.35	8.28	10.55	10.39	8.35	3.05	4.44	4.35	8.28	10.55	10.39	8.35	3.05	4.44
QALYs	Cover2	0.11	0.63	0.55	0.79	0.40	0.00	0.15	0.12	0.66	0.59	0.81	0.38	0.00	0.15
QALYs	Cover1	0.36	0.97	0.93	0.97	0.92	0.28	0.57	0.36	0.97	0.93	0.97	0.92	0.28	0.57

Table 26 : Result for simulated scenario 4

	Scenario4	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	50.98	56.86	56.91	56.70	50.23	42.34	44.33	52.41	57.73	57.78	57.56	49.91	42.53	44.71
PFS	Bias	-4.98	0.90	0.95	0.74	-5.73	-13.62	-11.63	-3.55	1.77	1.82	1.60	-6.05	-13.44	-11.25
PFS	MC SE	0.12	0.20	0.20	0.20	0.16	0.10	0.13	0.15	0.21	0.21	0.21	0.17	0.10	0.14
PFS	Rel Bias	-8.9%	1.6%	1.7%	1.3%	-10.2%	-24.3%	-20.8%	-6.4%	3.2%	3.2%	2.9%	-10.8%	-24.0%	-20.1%
PFS	empSE	3.82	6.23	6.25	6.16	5.05	3.07	3.91	4.59	6.50	6.52	6.42	5.13	3.11	4.04
PFS	MC SE	0.09	0.14	0.14	0.14	0.12	0.07	0.09	0.11	0.15	0.15	0.15	0.12	0.07	0.10
PFS	MSE	39.41	39.54	39.87	38.48	58.26	194.95	150.52	33.64	45.28	45.71	43.67	62.92	190.20	142.79
PFS	MC SE	1.39	1.87	1.90	1.84	1.75	2.64	3.02	1.39	2.19	2.23	2.15	1.95	2.65	2.94
PFS	Rel P	0.00%	-62.31%	-62.55%	-61.53%	-42.60%	55.31%	-4.49%	0.00%	-50.17%	-50.48%	-48.92%	-20.07%	116.80%	28.84%
PFS	ModelSE	5.71	6.87	6.88	6.78	5.90	3.24	4.12	5.71	6.87	6.88	6.78	5.90	3.24	4.12
PFS	Cover2	0.83	0.94	0.94	0.94	0.71	0.04	0.23	0.85	0.95	0.95	0.95	0.69	0.05	0.25
PFS	Cover1	0.99	0.99	0.99	0.99	0.97	0.51	0.80	0.99	0.99	0.99	0.99	0.97	0.51	0.80
OS	Mean	88.83	105.39	92.42	104.95	101.26	82.97	94.30	184.66	133.50	106.05	115.85	129.20	84.42	100.04
OS	Bias	-0.65	15.92	2.95	15.48	11.79	-6.50	4.82	95.19	44.03	16.58	26.38	39.73	-5.05	10.57
OS	MC SE	0.22	0.44	0.30	0.45	0.44	0.23	0.30	5.24	2.12	1.35	0.81	2.18	0.25	0.45
OS	Rel Bias	- 0.01	0.18	0.03	0.17	0.13	- 0.07	0.05	1.06	0.49	0.19	0.29	0.44	- 0.06	0.12
OS	empSE	6.72	13.71	9.20	13.91	13.69	7.14	8.92	161.42	65.35	41.74	24.91	67.21	7.65	13.24
OS	MC SE	0.15	0.31	0.21	0.32	0.31	0.16	0.21	3.71	1.50	0.96	0.57	1.54	0.18	0.32
OS	MSE	45.55	441.19	93.15	432.99	326.06	93.24	102.83	35090.28	6205.41	2015.56	1315.70	6091.48	84.06	286.84
OS	MC SE	0.15	0.31	0.21	0.32	0.31	0.16	0.21	2549.68	851.25	447.18	154.39	865.09	2.99	27.52
OS	Rel P	0.00%	-75.97%	-46.57%	-76.66%	-75.88%	-11.41%	-43.28%	0.00%	510.06%	1395.41%	4098.32%	476.77%	44380.83%	14765.31%
OS	ModelSE	173.47	81.81	48.90	37.89	85.84	9.74	33.12	173.47	81.81	48.90	37.89	85.84	9.74	33.12
OS	Cover2	1.00	0.92	0.99	1.00	0.97	0.84	1.00	0.99	0.90	1.00	0.99	0.96	0.85	1.00
OS	Cover1	1.00	0.94	1.00	1.00	0.98	0.99	1.00	1.00	0.94	1.00	1.00	0.98	0.99	1.00
QALYs	Mean	59.71	69.76	63.28	69.49	65.70	54.19	60.45	108.05	84.07	70.36	75.19	79.57	54.97	63.44
QALYs	Bias	-1.82	8.23	1.76	7.96	4.17	-7.34	-1.08	46.53	22.55	8.84	13.67	18.05	-6.56	1.91
QALYs	MC SE	0.13	0.25	0.17	0.25	0.24	0.13	0.17	2.64	1.07	0.68	0.42	1.10	0.14	0.24
QALYs	Rel Bias	- 0.03	0.13	0.03	0.13	0.07	- 0.12	- 0.02	0.76	0.37	0.14	0.22	0.29	- 0.11	0.03
QALYs	empSE	3.96	7.62	5.26	7.81	7.32	4.06	4.90	81.32	32.98	21.04	12.95	33.90	4.31	7.09
QALYs	MC SE	0.09	0.17	0.12	0.18	0.17	0.09	0.12	1.87	0.76	0.48	0.30	0.78	0.10	0.17
QALYs	MSE	18.96	125.75	30.71	124.36	70.96	70.31	25.12	8769.98	1594.71	520.43	354.43	1473.91	61.55	53.84
QALYs	MC SE	0.72	5.58	1.22	5.16	3.85	1.95	1.04	638.64	215.22	112.08	38.73	214.21	1.86	5.93
QALYs	Rel P	0.00%	-73.01%	-43.30%	-74.32%	-70.75%	-4.82%	-34.65%	0.00%	508.00%	1393.14%	3839.93%	475.29%	35506.92%	13059.80%
QALYs	ModelSE	87.31	41.07	24.78	19.45	43.12	5.39	17.12	87.31	41.07	24.78	19.45	43.12	5.39	17.12
QALYs	Cover2	1.00	0.93	1.00	1.00	0.99	0.64	0.99	0.99	0.92	1.00	1.00	0.98	0.69	0.99
QALYs	Cover1	1.00	0.96	1.00	1.00	0.99	0.97	1.00	1.00	0.96	1.00	1.00	0.99	0.97	1.00

Table 27 : Result for simulated scenario 5

	Scenario5	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	51.15	55.22	55.13	55.06	49.63	43.05	44.68	50.94	56.12	56.02	55.95	49.47	43.23	45.14
PFS	Bias	-4.81	-0.74	-0.83	-0.90	-6.33	-12.91	-11.28	-5.02	0.16	0.06	-0.02	-6.49	-12.74	-10.82
PFS	MC SE	0.13	0.17	0.17	0.17	0.14	0.10	0.11	0.12	0.17	0.18	0.18	0.14	0.10	0.12
PFS	Rel Bias	-8.6%	-1.3%	-1.5%	-1.6%	-11.3%	-23.1%	-20.2%	-9.0%	0.3%	0.1%	0.0%	-11.6%	-22.8%	-19.3%
PFS	empSE	3.88	5.11	5.25	5.33	4.23	3.01	3.50	3.68	5.30	5.46	5.52	4.35	3.05	3.57
PFS	MC SE	0.09	0.12	0.12	0.12	0.10	0.07	0.08	0.08	0.12	0.13	0.13	0.10	0.07	0.08
PFS	MSE	38.20	26.67	28.26	29.16	57.96	175.75	139.50	38.75	28.09	29.76	30.46	60.98	171.50	129.85
PFS	MC SE	1.23	0.95	1.08	1.11	1.72	2.50	2.61	1.22	1.03	1.15	1.17	1.82	2.49	2.55
PFS	Rel P	0.00%	-42.40%	-45.44%	-46.95%	-15.97%	65.67%	22.76%	0.00%	-51.83%	-54.57%	-55.61%	-28.32%	45.17%	6.34%
PFS	ModelSE	4.34	6.69	6.68	6.62	5.20	3.29	4.03	4.34	6.69	6.68	6.62	5.20	3.29	4.03
PFS	Cover2	0.74	0.95	0.94	0.93	0.66	0.06	0.24	0.73	0.95	0.94	0.94	0.64	0.07	0.28
PFS	Cover1	0.96	1.00	1.00	1.00	0.96	0.58	0.83	0.96	1.00	1.00	1.00	0.96	0.58	0.83
OS	Mean	71.97	79.40	72.63	82.76	75.06	64.14	74.60	74.85	85.96	76.90	86.94	80.15	64.52	76.41
OS	Bias	-1.44	5.99	-0.78	9.36	1.66	-9.26	1.19	1.45	12.55	3.49	13.53	6.75	-8.88	3.00
OS	MC SE	0.18	0.17	0.19	0.36	0.17	0.13	0.20	0.78	1.10	0.78	0.52	1.05	0.13	0.22
OS	Rel Bias	- 0.02	0.08 -	0.01	0.13	0.02 -	0.13	0.02	0.02	0.17	0.05	0.18	0.09 -	0.12	0.04
OS	empSE	5.60	5.20	5.85	11.04	5.22	3.92	6.09	24.02	33.79	23.97	15.88	32.43	4.02	6.71
OS	MC SE	0.13	0.12	0.13	0.25	0.12	0.09	0.14	0.55	0.78	0.55	0.36	0.74	0.09	0.15
OS	MSE	33.40	62.89	34.83	209.25	29.92	101.16	38.45	578.34	1298.35	586.21	435.01	1095.82	95.03	53.99
OS	MC SE	0.13	0.12	0.13	0.25	0.12	0.09	0.14	229.00	297.51	185.77	61.34	294.84	2.31	2.32
OS	Rel P	0.00%	16.02%	-8.48%	-74.26%	15.22%	103.80%	-15.42%	0.00%	-49.49%	0.39%	128.84%	-45.14%	3476.83%	1180.91%
OS	ModelSE	29.36	41.80	30.50	24.23	39.76	4.63	10.13	29.36	41.80	30.50	24.23	39.76	4.63	10.13
OS	Cover2	0.97	0.98	0.94	0.99	0.99	0.46	1.00	0.97	0.97	0.96	0.99	0.99	0.49	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00
QALYs	Mean	51.33	56.26	52.85	57.90	52.42	44.99	50.70	52.71	59.81	55.26	60.25	54.92	45.23	51.74
QALYs	Bias	-2.16	2.77	-0.64	4.41	-1.07	-8.50	-2.79	-0.78	6.32	1.76	6.76	1.43	-8.26	-1.75
QALYs	MC SE	0.12	0.12	0.13	0.21	0.11	0.09	0.12	0.40	0.56	0.40	0.28	0.53	0.09	0.13
QALYs	Rel Bias	- 0.04	0.05 -	0.01	0.08 -	0.02 -	0.16 -	0.05	- 0.01	0.12	0.03	0.13	0.03 -	0.15 -	0.03
QALYs	empSE	3.60	3.82	4.15	6.56	3.51	2.69	3.62	12.18	17.27	12.28	8.73	16.46	2.74	3.93
QALYs	MC SE	0.08	0.09	0.10	0.15	0.08	0.06	0.08	0.28	0.40	0.28	0.20	0.38	0.06	0.09
QALYs	MSE	17.64	22.24	17.65	62.44	13.44	79.54	20.89	148.79	337.81	153.83	121.88	272.56	75.78	18.48
QALYs	MC SE	0.64	0.91	0.62	2.81	0.54	1.48	0.82	55.78	75.49	45.72	15.54	72.14	1.47	0.75
QALYs	Rel P	0.00%	-10.96%	-24.87%	-69.88%	5.43%	79.55%	-1.22%	0.00%	-50.25%	-1.68%	94.58%	-45.23%	1873.71%	860.37%
QALYs	ModelSE	14.91	21.27	15.77	12.92	20.16	3.14	5.90	14.91	21.27	15.77	12.92	20.16	3.14	5.90
QALYs	Cover2	0.92	1.00	0.95	0.98	0.96	0.26	0.93	0.92	1.00	0.96	0.99	0.96	0.28	0.94
QALYs	Cover1	0.99	1.00	1.00	1.00	1.00	0.81	1.00	0.99	1.00	1.00	1.00	1.00	0.81	1.00

Table 28 : Result for simulated scenario 6

	Scenario6	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	51.25	54.43	54.32	54.41	48.90	43.59	46.00	50.70	55.40	55.28	55.37	48.70	43.72	46.49
PFS	Bias	-4.71	-1.54	-1.64	-1.55	-7.06	-12.37	-9.96	-5.26	-0.56	-0.68	-0.59	-7.27	-12.24	-9.47
PFS	MC SE	0.13	0.16	0.15	0.17	0.12	0.10	0.12	0.12	0.16	0.16	0.17	0.13	0.10	0.13
PFS	Rel Bias	-8.4%	-2.7%	-2.9%	-2.8%	-12.6%	-22.1%	-17.8%	-9.4%	-1.0%	-1.2%	-1.1%	-13.0%	-21.9%	-16.9%
PFS	empSE	3.94	4.84	4.76	5.09	3.71	2.98	3.83	3.71	4.96	4.89	5.22	3.95	3.01	3.90
PFS	MC SE	0.09	0.11	0.11	0.12	0.09	0.07	0.09	0.09	0.11	0.11	0.12	0.09	0.07	0.09
PFS	MSE	37.69	25.80	25.31	28.25	63.63	161.99	113.92	41.38	24.92	24.33	27.58	68.34	158.81	104.91
PFS	MC SE	1.28	0.96	0.92	1.04	1.71	2.38	2.58	1.33	0.93	0.88	1.04	1.80	2.38	2.51
PFS	Rel P	0.00%	-33.85%	-31.44%	-40.00%	12.71%	74.87%	6.00%	0.00%	-44.16%	-42.42%	-49.54%	-11.64%	52.02%	-9.67%
PFS	ModelSE	4.24	6.85	6.81	6.81	5.00	3.37	4.31	4.24	6.85	6.81	6.81	5.00	3.37	4.31
PFS	Cover2	0.73	0.94	0.94	0.93	0.63	0.08	0.37	0.70	0.95	0.95	0.94	0.59	0.09	0.41
PFS	Cover1	0.95	1.00	1.00	1.00	0.95	0.62	0.90	0.95	1.00	1.00	1.00	0.95	0.62	0.90
OS	Mean	61.10	66.43	62.90	73.79	61.02	54.39	63.33	61.67	67.58	64.57	76.24	61.22	54.55	64.53
OS	Bias	-4.20	1.14	-2.39	8.50	-4.27	-10.90	-1.96	-3.62	2.28	-0.72	10.95	-4.07	-10.74	-0.76
OS	MC SE	0.16	0.16	0.15	0.35	0.13	0.10	0.18	0.17	0.24	0.27	0.41	0.23	0.10	0.18
OS	Rel Bias	- 0.06	0.02 -	0.04	0.13 -	0.07 -	0.17 -	0.03	- 0.06	0.03 -	0.01	0.17 -	0.06 -	0.16 -	0.01
OS	empSE	4.99	4.86	4.68	10.77	3.95	3.13	5.40	5.15	7.44	8.18	12.63	7.10	3.17	5.70
OS	MC SE	0.11	0.11	0.11	0.25	0.09	0.07	0.12	0.12	0.17	0.19	0.29	0.16	0.07	0.13
OS	MSE	42.45	24.94	27.61	188.21	33.81	128.71	32.97	39.64	60.52	67.28	279.13	66.96	125.32	32.98
OS	MC SE	0.11	0.11	0.11	0.25	0.09	0.07	0.12	1.40	28.89	17.26	36.26	28.84	2.20	1.22
OS	Rel P	0.00%	5.11%	13.37%	-78.57%	59.44%	153.63%	-14.69%	0.00%	-52.03%	-60.26%	-83.34%	-47.34%	164.43%	-18.14%
OS	ModelSE	6.06	11.81	15.21	18.35	11.15	3.66	6.96	6.06	11.81	15.21	18.35	11.15	3.66	6.96
OS	Cover2	0.83	0.99	0.92	0.97	0.87	0.19	0.93	0.85	1.00	0.94	0.98	0.86	0.21	0.95
OS	Cover1	0.98	1.00	1.00	1.00	1.00	0.76	1.00	0.98	1.00	1.00	1.00	1.00	0.76	1.00
QALYs	Mean	45.92	49.54	47.75	53.22	45.18	40.27	45.46	46.05	50.41	48.87	54.73	45.22	40.39	46.21
QALYs	Bias	-3.51	0.11	-1.69	3.78	-4.25	-9.16	-3.97	-3.39	0.97	-0.56	5.30	-4.22	-9.04	-3.22
QALYs	MC SE	0.11	0.12	0.12	0.21	0.10	0.08	0.12	0.11	0.15	0.16	0.23	0.14	0.08	0.12
QALYs	Rel Bias	- 0.07	0.00 -	0.03	0.08 -	0.09 -	0.19 -	0.08	- 0.07	0.02 -	0.01	0.11 -	0.09 -	0.18 -	0.07
QALYs	empSE	3.43	3.79	3.72	6.39	2.97	2.41	3.61	3.48	4.75	5.04	7.21	4.26	2.44	3.77
QALYs	MC SE	0.08	0.09	0.09	0.15	0.07	0.06	0.08	0.08	0.11	0.12	0.17	0.10	0.06	0.09
QALYs	MSE	24.09	14.39	16.64	55.13	26.92	89.80	28.79	23.59	23.50	25.71	80.01	35.92	87.66	24.59
QALYs	MC SE	0.86	0.53	0.58	2.31	0.80	1.43	0.99	0.84	6.99	4.32	8.89	6.75	1.43	0.89
QALYs	Rel P	0.00%	-18.21%	-14.72%	-71.18%	33.17%	102.53%	-9.69%	0.00%	-46.27%	-52.29%	-76.67%	-33.24%	104.00%	-14.83%
QALYs	ModelSE	3.80	7.23	8.67	10.31	6.41	2.80	4.57	3.80	7.23	8.67	10.31	6.41	2.80	4.57
QALYs	Cover2	0.76	0.97	0.93	0.97	0.79	0.14	0.80	0.76	0.98	0.94	0.97	0.76	0.15	0.84
QALYs	Cover1	0.94	1.00	1.00	1.00	0.99	0.69	0.99	0.94	1.00	1.00	1.00	0.99	0.69	0.99

Table 29 : Result for simulated scenario 7

	Scenario7	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	59.42	60.96	61.11	60.87	57.25	49.08	51.30	59.43	61.54	61.70	61.45	56.89	49.20	51.43
PFS	Bias	-9.76	-8.22	-8.07	-8.31	-11.93	-20.10	-17.88	-9.75	-7.64	-7.48	-7.73	-12.29	-19.98	-17.75
PFS	MC SE	0.16	0.19	0.19	0.18	0.17	0.10	0.12	0.16	0.20	0.20	0.19	0.18	0.10	0.12
PFS	Rel Bias	-14.1%	-11.9%	-11.7%	-12.0%	-17.2%	-29.1%	-25.8%	-14.1%	-11.0%	-10.8%	-11.2%	-17.8%	-28.9%	-25.7%
PFS	empSE	4.92	5.89	5.98	5.68	5.27	3.09	3.62	4.80	6.17	6.26	5.94	5.69	3.11	3.66
PFS	MC SE	0.11	0.14	0.14	0.13	0.12	0.07	0.08	0.11	0.14	0.14	0.14	0.13	0.07	0.08
PFS	MSE	119.55	102.33	100.83	101.24	170.16	413.55	332.75	118.12	96.36	95.07	94.94	183.44	408.72	328.40
PFS	MC SE	2.97	2.97	2.94	2.92	3.64	3.98	4.20	2.90	2.93	2.91	2.86	4.34	3.99	4.20
PFS	Rel P	0.00%	-30.26%	-32.19%	-24.90%	-12.84%	153.85%	84.69%	0.00%	-39.45%	-41.22%	-34.67%	-28.93%	137.36%	72.16%
PFS	ModelSE	4.70	5.80	5.86	5.73	6.77	3.09	3.70	4.70	5.80	5.86	5.73	6.77	3.09	3.70
PFS	Cover2	0.42	0.58	0.59	0.59	0.34	0.00	0.01	0.42	0.61	0.62	0.61	0.34	0.00	0.01
PFS	Cover1	0.84	0.92	0.92	0.92	0.78	0.08	0.35	0.84	0.92	0.92	0.92	0.78	0.08	0.35
OS	Mean	117.53	132.87	121.04	128.05	129.77	111.21	120.67	119.13	148.18	127.29	132.44	142.74	112.11	122.67
OS	Bias	-4.43	10.90	-0.92	6.09	7.80	-10.76	-1.30	-2.83	26.21	5.32	10.48	20.77	-9.86	0.70
OS	MC SE	0.27	0.40	0.38	0.37	0.40	0.23	0.30	0.35	1.48	0.97	0.71	1.36	0.24	0.32
OS	Rel Bias	- 0.04	0.09	- 0.01	0.05	0.06	- 0.09	- 0.01	- 0.02	0.21	0.04	0.09	0.17	- 0.08	0.01
OS	empSE	8.47	12.41	11.74	11.42	12.26	7.13	9.20	10.94	45.58	29.83	21.86	41.94	7.35	9.92
OS	MC SE	0.19	0.28	0.27	0.26	0.28	0.16	0.21	0.25	1.05	0.68	0.50	0.96	0.17	0.23
OS	MSE	91.35	272.54	138.55	167.26	211.02	166.49	86.28	127.53	2762.62	917.23	587.00	2188.52	151.16	98.73
OS	MC SE	0.19	0.28	0.27	0.26	0.28	0.16	0.21	39.80	456.66	225.61	161.08	380.81	4.80	3.38
OS	Rel P	0.00%	-53.37%	-47.94%	-44.95%	-52.26%	41.20%	-15.25%	0.00%	-94.24%	-86.56%	-74.96%	-93.20%	121.32%	21.64%
OS	ModelSE	16.74	61.83	39.49	29.67	60.17	9.11	16.09	16.74	61.83	39.49	29.67	60.17	9.11	16.09
OS	Cover2	0.96	0.94	0.93	1.00	0.97	0.71	0.99	0.97	0.93	0.95	1.00	0.97	0.73	1.00
OS	Cover1	1.00	0.99	1.00	1.00	0.99	0.98	1.00	1.00	0.99	1.00	1.00	0.99	0.98	1.00
QALYs	Mean	76.59	84.72	78.85	82.29	82.06	70.33	75.72	77.39	92.55	82.16	84.66	88.44	70.82	76.76
QALYs	Bias	-5.15	2.98	-2.88	0.55	0.32	-11.41	-6.01	-4.34	10.82	0.42	2.92	6.70	-10.92	-4.98
QALYs	MC SE	0.16	0.22	0.20	0.21	0.22	0.13	0.16	0.20	0.75	0.49	0.37	0.69	0.13	0.17
QALYs	Rel Bias	- 0.06	0.04	- 0.04	0.01	0.00	- 0.14	- 0.07	- 0.05	0.13	0.01	0.04	0.08	- 0.13	- 0.06
QALYs	empSE	4.83	6.93	6.23	6.43	6.76	4.01	5.00	6.02	23.12	15.08	11.39	21.38	4.12	5.37
QALYs	MC SE	0.11	0.16	0.14	0.15	0.16	0.09	0.11	0.14	0.53	0.35	0.26	0.49	0.09	0.12
QALYs	MSE	49.77	56.93	47.02	41.62	45.78	146.25	61.19	55.05	651.04	227.45	138.02	501.31	136.27	53.53
QALYs	MC SE	1.78	2.80	1.62	1.48	1.97	2.94	2.18	10.08	111.92	53.76	38.97	91.88	2.88	2.02
QALYs	Rel P	0.00%	-51.53%	-39.86%	-43.65%	-49.04%	44.66%	-6.96%	0.00%	-93.22%	-84.08%	-72.05%	-92.07%	112.97%	25.74%
QALYs	ModelSE	8.73	31.10	20.04	15.29	30.53	5.02	8.59	8.73	31.10	20.04	15.29	30.53	5.02	8.59
QALYs	Cover2	0.88	0.98	0.90	0.98	0.99	0.38	0.88	0.89	0.97	0.93	0.99	0.99	0.41	0.90
QALYs	Cover1	0.99	0.99	1.00	1.00	1.00	0.91	1.00	0.99	0.99	1.00	1.00	1.00	0.91	1.00

Table 30 : Result for simulated scenario 8

	Scenario8	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.42	60.68	60.81	60.53	55.93	49.29	50.18	58.20	61.23	61.37	61.08	55.98	49.41	50.30
PFS	Bias	-10.76	-8.50	-8.37	-8.65	-13.25	-19.89	-19.00	-10.98	-7.95	-7.81	-8.10	-13.20	-19.77	-18.88
PFS	MC SE	0.14	0.17	0.18	0.17	0.15	0.10	0.12	0.14	0.18	0.19	0.18	0.16	0.10	0.12
PFS	Rel Bias	-15.6%	-12.3%	-12.1%	-12.5%	-19.1%	-28.8%	-27.5%	-15.9%	-11.5%	-11.3%	-11.7%	-19.1%	-28.6%	-27.3%
PFS	empSE	4.40	5.27	5.66	5.29	4.53	2.97	3.58	4.28	5.44	5.87	5.48	4.81	3.01	3.59
PFS	MC SE	0.10	0.12	0.13	0.12	0.10	0.07	0.08	0.10	0.12	0.13	0.13	0.11	0.07	0.08
PFS	MSE	135.21	100.05	102.09	102.75	195.92	404.51	373.89	138.89	92.79	95.45	95.53	197.37	399.88	369.27
PFS	MC SE	3.07	2.87	2.97	2.94	3.67	3.82	4.37	3.04	2.77	2.88	2.85	4.03	3.85	4.35
PFS	Rel P	0.00%	-30.27%	-39.55%	-30.82%	-5.57%	119.78%	51.17%	0.00%	-38.07%	-46.79%	-38.87%	-20.65%	102.33%	42.42%
PFS	ModelSE	4.14	5.53	5.62	5.53	6.04	3.12	3.42	4.14	5.53	5.62	5.53	6.04	3.12	3.42
PFS	Cover2	0.30	0.58	0.58	0.57	0.26	0.00	0.00	0.28	0.61	0.61	0.60	0.27	0.00	0.00
PFS	Cover1	0.71	0.92	0.92	0.92	0.74	0.10	0.23	0.71	0.92	0.92	0.92	0.74	0.10	0.23
OS	Mean	89.58	101.62	92.18	100.23	96.56	85.61	91.57	90.35	104.94	95.47	102.64	98.78	85.97	92.51
OS	Bias	-10.37	1.68	-7.77	0.28	-3.39	-14.34	-8.37	-9.60	4.99	-4.48	2.70	-1.17	-13.98	-7.44
OS	MC SE	0.19	0.20	0.25	0.28	0.20	0.14	0.21	0.20	0.67	0.75	0.50	0.58	0.14	0.22
OS	Rel Bias	-0.10	0.02	-0.08	0.00	-0.03	-0.14	-0.08	-0.10	0.05	-0.04	0.03	-0.01	-0.14	-0.07
OS	empSE	5.89	6.26	7.55	8.59	6.22	4.19	6.55	6.11	20.69	23.12	15.31	17.73	4.27	6.86
OS	MC SE	0.14	0.14	0.17	0.20	0.14	0.10	0.15	0.14	0.47	0.53	0.35	0.41	0.10	0.16
OS	MSE	142.17	42.01	117.38	73.79	50.17	223.07	112.95	129.41	452.46	553.86	241.42	315.43	213.50	102.34
OS	MC SE	0.14	0.14	0.17	0.20	0.14	0.10	0.15	3.64	139.05	159.27	81.51	99.55	3.88	3.22
OS	Rel P	0.00%	-11.67%	-39.23%	-53.03%	-10.54%	97.07%	-19.16%	0.00%	-91.29%	-93.02%	-84.09%	-88.14%	104.95%	-20.82%
OS	ModelSE	7.80	31.59	32.08	23.47	27.83	4.98	8.56	7.80	31.59	32.08	23.47	27.83	4.98	8.56
OS	Cover2	0.61	0.99	0.65	0.96	0.93	0.21	0.73	0.64	0.99	0.69	0.97	0.93	0.23	0.75
OS	Cover1	0.96	1.00	0.97	1.00	1.00	0.77	0.99	0.96	1.00	0.97	1.00	1.00	0.77	0.99
QALYs	Mean	62.31	69.01	64.33	68.27	65.06	57.59	60.84	62.63	70.84	66.14	69.65	66.18	57.81	61.35
QALYs	Bias	-8.41	-1.71	-6.40	-2.45	-5.67	-13.14	-9.89	-8.09	0.11	-4.59	-1.08	-4.54	-12.92	-9.38
QALYs	MC SE	0.13	0.14	0.16	0.18	0.13	0.09	0.13	0.13	0.35	0.39	0.27	0.30	0.09	0.13
QALYs	Rel Bias	-0.12	-0.02	-0.09	-0.03	-0.08	-0.19	-0.14	-0.11	0.00	-0.06	-0.02	-0.06	-0.18	-0.13
QALYs	empSE	3.86	4.28	4.78	5.40	4.05	2.73	3.96	3.94	10.74	11.99	8.43	9.31	2.77	4.11
QALYs	MC SE	0.09	0.10	0.11	0.12	0.09	0.06	0.09	0.09	0.25	0.28	0.19	0.21	0.06	0.09
QALYs	MSE	85.70	21.27	63.77	35.20	48.48	179.96	113.41	81.03	115.21	164.74	72.22	107.27	174.57	104.91
QALYs	MC SE	2.06	0.87	1.85	1.18	1.56	2.32	2.48	2.01	32.82	38.70	20.28	22.67	2.32	2.43
QALYs	Rel P	0.00%	-18.74%	-34.80%	-48.93%	-9.00%	100.74%	-4.79%	0.00%	-86.53%	-89.21%	-78.17%	-82.09%	102.18%	-8.19%
QALYs	ModelSE	4.43	16.15	16.39	12.30	14.53	3.18	4.93	4.43	16.15	16.39	12.30	14.53	3.18	4.93
QALYs	Cover2	0.45	0.92	0.64	0.89	0.78	0.02	0.41	0.45	0.93	0.66	0.92	0.79	0.03	0.44
QALYs	Cover1	0.80	0.99	0.96	1.00	0.98	0.50	0.91	0.80	0.99	0.96	1.00	0.98	0.50	0.91

Table 31 : Result for simulated scenario 9

	Scenario9	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	55.14	60.07	59.80	59.89	54.87	49.19	49.04	54.86	60.65	60.36	60.47	54.77	49.29	49.08
PFS	Bias	-14.04	-9.11	-9.38	-9.29	-14.31	-19.99	-20.14	-14.32	-8.53	-8.82	-8.71	-14.41	-19.89	-20.10
PFS	MC SE	0.12	0.16	0.16	0.16	0.13	0.09	0.13	0.12	0.16	0.16	0.17	0.13	0.09	0.13
PFS	Rel Bias	-20.3%	-13.2%	-13.6%	-13.4%	-20.7%	-28.9%	-29.1%	-20.7%	-12.3%	-12.8%	-12.6%	-20.8%	-28.7%	-29.1%
PFS	empSE	3.75	4.85	4.80	4.96	3.89	2.88	3.93	3.64	5.02	4.95	5.13	4.04	2.90	3.89
PFS	MC SE	0.09	0.11	0.11	0.11	0.09	0.07	0.09	0.08	0.12	0.11	0.12	0.09	0.07	0.09
PFS	MSE	211.12	106.40	110.99	110.77	219.85	407.95	420.97	218.36	97.92	102.34	102.22	223.91	403.86	419.03
PFS	MC SE	3.42	2.95	3.06	3.06	3.50	3.75	5.17	3.40	2.87	2.99	2.98	3.82	3.75	5.11
PFS	Rel P	0.00%	-40.33%	-39.03%	-42.91%	-7.27%	68.64%	-9.05%	0.00%	-47.34%	-45.94%	-49.70%	-18.59%	57.45%	-12.55%
PFS	ModelSE	3.35	5.54	5.45	5.51	5.58	3.16	3.23	3.35	5.54	5.45	5.51	5.58	3.16	3.23
PFS	Cover2	0.06	0.56	0.55	0.55	0.18	0.00	0.00	0.05	0.59	0.58	0.58	0.19	0.00	0.00
PFS	Cover1	0.38	0.92	0.92	0.92	0.68	0.09	0.18	0.38	0.92	0.92	0.92	0.68	0.09	0.18
OS	Mean	70.44	81.55	76.24	84.47	76.01	69.03	72.61	70.98	82.01	77.27	85.75	76.07	69.18	73.09
OS	Bias	-15.94	-4.83	-10.14	-1.90	-10.37	-17.35	-13.77	-15.40	-4.37	-9.11	-0.63	-10.31	-17.20	-13.29
OS	MC SE	0.13	0.17	0.14	0.27	0.14	0.10	0.14	0.14	0.20	0.34	0.32	0.18	0.10	0.14
OS	Rel Bias	-0.18	-0.06	-0.12	-0.02	-0.12	-0.20	-0.16	-0.18	-0.05	-0.11	-0.01	-0.12	-0.20	-0.15
OS	empSE	4.07	5.17	4.41	8.33	4.29	3.08	4.23	4.34	6.11	10.52	9.94	5.66	3.10	4.41
OS	MC SE	0.09	0.12	0.10	0.19	0.10	0.07	0.10	0.10	0.14	0.24	0.23	0.13	0.07	0.10
OS	MSE	270.62	50.03	122.17	72.96	125.91	310.60	207.38	256.02	56.37	193.56	99.13	138.21	305.49	196.12
OS	MC SE	0.09	0.12	0.10	0.19	0.10	0.07	0.10	4.31	7.74	65.80	18.75	7.71	3.45	3.70
OS	Rel P	0.00%	-38.06%	-14.76%	-76.16%	-10.00%	74.67%	-7.63%	0.00%	-49.52%	-82.97%	-80.93%	-41.08%	95.88%	-3.22%
OS	ModelSE	5.26	8.56	15.66	14.27	9.16	3.61	5.53	5.26	8.56	15.66	14.27	9.16	3.61	5.53
OS	Cover2	0.13	0.80	0.50	0.86	0.48	0.00	0.25	0.14	0.81	0.53	0.88	0.50	0.00	0.27
OS	Cover1	0.65	0.97	0.92	0.99	0.93	0.31	0.85	0.65	0.97	0.92	0.99	0.93	0.31	0.85
QALYs	Mean	51.76	58.80	56.06	60.21	54.47	49.27	51.02	51.95	59.20	56.74	61.02	54.47	49.38	51.27
QALYs	Bias	-12.18	-5.15	-7.88	-3.74	-9.48	-14.67	-12.92	-12.00	-4.74	-7.20	-2.93	-9.48	-14.57	-12.67
QALYs	MC SE	0.10	0.13	0.11	0.17	0.10	0.07	0.10	0.10	0.14	0.19	0.20	0.12	0.08	0.10
QALYs	Rel Bias	-0.19	-0.08	-0.12	-0.06	-0.15	-0.23	-0.20	-0.19	-0.07	-0.11	-0.05	-0.15	-0.23	-0.20
QALYs	empSE	3.03	3.88	3.47	5.31	3.12	2.31	3.09	3.11	4.29	5.91	6.02	3.75	2.33	3.17
QALYs	MC SE	0.07	0.09	0.08	0.12	0.07	0.05	0.07	0.07	0.10	0.14	0.14	0.09	0.05	0.07
QALYs	MSE	157.56	41.51	74.20	42.17	99.53	220.66	176.59	153.57	40.88	86.80	44.72	103.83	217.59	170.67
QALYs	MC SE	2.41	1.39	1.75	1.50	1.83	2.20	2.58	2.43	2.18	15.43	4.26	2.68	2.20	2.57
QALYs	Rel P	0.00%	-38.98%	-24.05%	-67.51%	-5.57%	71.61%	-4.11%	0.00%	-47.56%	-72.40%	-73.32%	-31.23%	78.11%	-3.80%
QALYs	ModelSE	3.22	5.46	8.50	8.09	5.86	2.67	3.53	3.22	5.46	8.50	8.09	5.86	2.67	3.53
QALYs	Cover2	0.08	0.72	0.51	0.81	0.34	0.00	0.08	0.08	0.73	0.53	0.83	0.34	0.00	0.09
QALYs	Cover1	0.32	0.96	0.92	0.97	0.87	0.19	0.58	0.32	0.96	0.92	0.97	0.87	0.19	0.58

Table 32 : Result for simulated scenario 10

	Scenario10	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	52.98	56.91	57.00	56.93	50.50	44.29	45.70	53.93	57.53	57.62	57.55	50.48	44.39	45.76
PFS	Bias	-2.99	0.95	1.04	0.96	-5.46	-11.68	-10.26	-2.03	1.57	1.66	1.58	-5.48	-11.57	-10.20
PFS	MC SE	0.12	0.18	0.18	0.17	0.14	0.09	0.11	0.14	0.18	0.18	0.18	0.14	0.09	0.11
PFS	Rel Bias	-5.3%	1.7%	1.9%	1.7%	-9.8%	-20.9%	-18.3%	-3.6%	2.8%	3.0%	2.8%	-9.8%	-20.7%	-18.2%
PFS	empSE	3.75	5.47	5.49	5.34	4.18	2.73	3.34	4.19	5.64	5.66	5.51	4.17	2.75	3.30
PFS	MC SE	0.09	0.13	0.13	0.12	0.10	0.06	0.08	0.10	0.13	0.13	0.13	0.10	0.06	0.08
PFS	MSE	23.01	30.77	31.14	29.41	47.24	143.83	116.42	21.67	34.24	34.72	32.79	47.42	141.52	114.93
PFS	MC SE	0.90	1.47	1.50	1.40	1.47	2.04	2.31	0.89	1.67	1.70	1.59	1.53	2.04	2.21
PFS	Rel P	0.00%	-52.84%	-53.14%	-50.54%	-19.12%	89.37%	26.33%	0.00%	-44.78%	-45.13%	-42.07%	0.93%	131.44%	61.24%
PFS	ModelSE	4.65	5.26	5.28	5.27	4.42	2.76	3.20	4.65	5.26	5.28	5.27	4.42	2.76	3.20
PFS	Cover2	0.87	0.93	0.93	0.94	0.62	0.04	0.15	0.89	0.92	0.92	0.93	0.63	0.04	0.15
PFS	Cover1	0.99	1.00	1.00	1.00	0.93	0.42	0.72	0.99	1.00	1.00	1.00	0.93	0.42	0.72
OS	Mean	113.14	131.14	117.39	121.33	127.83	109.22	117.75	179.51	158.37	134.24	132.92	158.40	110.65	122.04
OS	Bias	1.63	19.63	5.88	9.82	16.32	-2.29	6.24	68.00	46.86	22.73	21.41	46.89	-0.86	10.53
OS	MC SE	0.28	0.55	0.36	0.35	0.55	0.28	0.32	4.05	2.01	1.59	1.08	2.22	0.29	0.42
OS	Rel Bias	0.01	0.18	0.05	0.09	0.15	-0.02	0.06	0.61	0.42	0.20	0.19	0.42	-0.01	0.09
OS	empSE	8.59	16.96	10.97	10.68	16.97	8.62	9.68	124.80	62.00	49.05	33.30	68.48	9.01	12.66
OS	MC SE	0.20	0.39	0.25	0.25	0.39	0.20	0.23	2.86	1.42	1.13	0.76	1.57	0.21	0.30
OS	MSE	76.37	672.91	154.73	210.22	553.86	79.42	132.53	20183.67	6036.25	2920.45	1566.03	6882.34	81.84	270.91
OS	MC SE	0.20	0.39	0.25	0.25	0.39	0.20	0.23	1761.60	731.33	496.77	262.82	806.27	3.02	18.69
OS	Rel P	0.00%	-74.36%	-38.68%	-35.26%	-74.38%	-0.63%	-21.26%	0.00%	305.16%	547.31%	1304.63%	232.18%	19083.96%	9625.49%
OS	ModelSE	144.17	78.55	58.99	45.42	86.69	11.71	28.21	144.17	78.55	58.99	45.42	86.69	11.71	28.21
OS	Cover2	1.00	0.88	1.00	1.00	0.93	0.95	1.00	1.00	0.86	1.00	1.00	0.92	0.95	1.00
OS	Cover1	1.00	0.93	1.00	1.00	0.97	1.00	1.00	1.00	0.93	1.00	1.00	0.97	1.00	1.00
QALYs	Mean	72.47	82.65	75.80	77.74	79.06	67.90	72.59	105.94	96.45	84.41	83.72	94.34	68.64	74.75
QALYs	Bias	-0.08	10.10	3.25	5.20	6.52	-4.65	0.04	33.39	23.90	11.86	11.18	21.80	-3.90	2.21
QALYs	MC SE	0.15	0.29	0.19	0.19	0.28	0.15	0.17	2.04	1.01	0.80	0.54	1.11	0.16	0.22
QALYs	Rel Bias	0.00	0.14	0.04	0.07	0.09	-0.06	0.00	0.46	0.33	0.16	0.15	0.30	-0.05	0.03
QALYs	empSE	4.75	8.93	5.75	5.82	8.77	4.61	5.15	62.76	31.18	24.53	16.77	34.29	4.81	6.61
QALYs	MC SE	0.11	0.20	0.13	0.13	0.20	0.11	0.12	1.44	0.72	0.56	0.38	0.79	0.11	0.15
QALYs	MSE	22.50	181.62	43.63	60.90	119.27	42.86	26.50	5049.54	1542.59	741.71	405.93	1649.97	38.33	48.45
QALYs	MC SE	0.79	7.50	1.70	2.25	5.73	1.53	1.02	441.39	184.44	124.17	65.42	198.30	1.42	4.04
QALYs	Rel P	0.00%	-71.74%	-31.97%	-33.61%	-70.70%	5.86%	-15.11%	0.00%	305.06%	554.69%	1300.32%	234.91%	16921.11%	8927.23%
QALYs	ModelSE	72.54	39.37	29.63	22.90	43.47	6.18	14.47	72.54	39.37	29.63	22.90	43.47	6.18	14.47
QALYs	Cover2	1.00	0.89	1.00	1.00	0.95	0.84	1.00	1.00	0.86	1.00	1.00	0.95	0.87	1.00
QALYs	Cover1	1.00	0.94	1.00	1.00	0.98	0.99	1.00	1.00	0.94	1.00	1.00	0.98	0.99	1.00

Table 33 : Result for simulated scenario 11

	Scenario11	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	51.31	56.68	56.74	56.67	50.08	44.62	44.83	51.36	57.29	57.36	57.28	50.06	44.71	44.79
PFS	Bias	-4.65	0.72	0.78	0.70	-5.88	-11.34	-11.13	-4.60	1.33	1.39	1.32	-5.90	-11.25	-11.17
PFS	MC SE	0.11	0.17	0.17	0.17	0.13	0.09	0.09	0.11	0.17	0.18	0.17	0.13	0.09	0.09
PFS	Rel Bias	-8.3%	1.3%	1.4%	1.3%	-10.5%	-20.3%	-19.9%	-8.2%	2.4%	2.5%	2.4%	-10.5%	-20.1%	-20.0%
PFS	empSE	3.44	5.13	5.26	5.15	3.93	2.69	2.89	3.39	5.29	5.43	5.32	3.94	2.72	2.91
PFS	MC SE	0.08	0.12	0.12	0.12	0.09	0.06	0.07	0.08	0.12	0.12	0.12	0.09	0.06	0.07
PFS	MSE	33.39	26.78	28.28	27.03	50.02	135.84	132.16	32.69	29.78	31.44	30.05	50.35	133.96	133.28
PFS	MC SE	1.04	1.26	1.33	1.26	1.44	1.98	2.11	1.02	1.43	1.51	1.43	1.44	1.98	2.13
PFS	Rel P	0.00%	-55.12%	-57.39%	-55.57%	-23.49%	62.95%	41.44%	0.00%	-58.96%	-61.03%	-59.40%	-25.98%	55.58%	35.72%
PFS	ModelSE	3.44	5.23	5.25	5.23	4.00	2.80	2.92	3.44	5.23	5.25	5.23	4.00	2.80	2.92
PFS	Cover2	0.65	0.94	0.93	0.94	0.59	0.05	0.06	0.65	0.94	0.93	0.94	0.59	0.05	0.07
PFS	Cover1	0.93	1.00	1.00	1.00	0.93	0.46	0.55	0.93	1.00	1.00	1.00	0.93	0.46	0.55
OS	Mean	80.28	92.00	84.30	90.01	86.65	77.15	81.73	81.91	98.25	87.98	92.58	92.16	77.50	82.25
OS	Bias	-3.48	8.24	0.54	6.24	2.89	-6.61	-2.03	-1.85	14.49	4.22	8.81	8.40	-6.26	-1.51
OS	MC SE	0.18	0.19	0.21	0.24	0.18	0.13	0.19	0.53	0.89	0.69	0.38	0.89	0.13	0.20
OS	Rel Bias	- 0.04	0.10	0.01	0.07	0.03	- 0.08	- 0.02	- 0.02	0.17	0.05	0.11	0.10	- 0.07	- 0.02
OS	empSE	5.46	5.88	6.50	7.25	5.47	4.06	5.73	16.29	27.43	21.36	11.64	27.31	4.13	6.06
OS	MC SE	0.13	0.13	0.15	0.17	0.13	0.09	0.13	0.37	0.63	0.49	0.27	0.63	0.09	0.14
OS	MSE	41.88	102.34	42.54	91.52	38.23	60.20	36.97	268.40	961.64	473.62	213.11	815.60	56.22	38.95
OS	MC SE	0.13	0.13	0.15	0.17	0.13	0.09	0.13	125.48	209.26	159.50	45.23	219.48	1.73	1.38
OS	Rel P	0.00%	-13.75%	-29.58%	-43.38%	-0.48%	80.22%	-9.43%	0.00%	-64.75%	-41.86%	95.69%	-64.43%	1453.44%	622.62%
OS	ModelSE	21.52	38.26	28.50	18.09	37.31	4.87	7.53	21.52	38.26	28.50	18.09	37.31	4.87	7.53
OS	Cover2	0.92	0.87	0.95	1.00	0.99	0.69	0.95	0.93	0.85	0.96	0.97	0.99	0.70	0.96
OS	Cover1	0.99	0.97	1.00	1.00	1.00	0.97	1.00	0.99	0.97	1.00	1.00	1.00	0.97	1.00
QALYs	Mean	55.53	63.00	59.17	62.00	58.35	51.96	54.31	56.36	66.31	61.20	63.47	61.10	52.16	54.56
QALYs	Bias	-3.14	4.33	0.51	3.33	-0.32	-6.71	-4.35	-2.30	7.64	2.53	4.80	2.43	-6.50	-4.11
QALYs	MC SE	0.11	0.13	0.14	0.15	0.11	0.08	0.11	0.27	0.46	0.36	0.21	0.45	0.09	0.11
QALYs	Rel Bias	- 0.05	0.07	0.01	0.06	- 0.01	- 0.11	- 0.07	- 0.04	0.13	0.04	0.08	0.04	- 0.11	- 0.07
QALYs	empSE	3.43	4.11	4.28	4.69	3.50	2.59	3.28	8.44	14.04	11.05	6.58	13.87	2.63	3.46
QALYs	MC SE	0.08	0.09	0.10	0.11	0.08	0.06	0.08	0.19	0.32	0.25	0.15	0.32	0.06	0.08
QALYs	MSE	21.59	35.63	18.58	33.10	12.33	51.68	29.72	76.40	255.41	128.41	66.28	198.11	49.19	28.79
QALYs	MC SE	0.72	1.39	0.70	1.36	0.51	1.13	0.94	30.82	52.62	39.86	11.27	53.89	1.11	0.93
QALYs	Rel P	0.00%	-30.23%	-35.85%	-46.52%	-3.81%	75.67%	9.31%	0.00%	-63.92%	-41.74%	64.51%	-63.02%	932.04%	495.98%
QALYs	ModelSE	10.98	19.37	14.59	9.61	18.83	3.03	4.32	10.98	19.37	14.59	9.61	18.83	3.03	4.32
QALYs	Cover2	0.83	0.93	0.95	0.99	0.97	0.40	0.79	0.85	0.91	0.96	0.98	0.97	0.43	0.79
QALYs	Cover1	0.98	0.99	1.00	1.00	1.00	0.88	0.98	0.98	0.99	1.00	1.00	1.00	0.88	0.98

Table 34 : Result for simulated scenario 12

	Scenario12	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	52.22	55.23	55.07	54.99	49.26	44.83	45.99	51.96	55.83	55.67	55.59	49.20	44.91	46.04
PFS	Bias	-3.74	-0.73	-0.89	-0.97	-6.71	-11.14	-9.97	-4.01	-0.13	-0.29	-0.37	-6.76	-11.05	-9.93
PFS	MC SE	0.12	0.14	0.14	0.14	0.11	0.08	0.10	0.12	0.14	0.14	0.15	0.11	0.08	0.10
PFS	Rel Bias	-6.7%	-1.3%	-1.6%	-1.7%	-12.0%	-19.9%	-17.8%	-7.2%	-0.2%	-0.5%	-0.7%	-12.1%	-19.8%	-17.7%
PFS	empSE	3.83	4.20	4.27	4.37	3.51	2.59	3.14	3.72	4.40	4.47	4.56	3.52	2.61	3.18
PFS	MC SE	0.09	0.10	0.10	0.10	0.08	0.06	0.07	0.09	0.10	0.10	0.10	0.08	0.06	0.07
PFS	MSE	28.68	18.20	19.03	20.00	57.26	130.71	109.19	29.88	19.32	20.00	20.90	58.01	129.02	108.65
PFS	MC SE	0.99	0.66	0.70	0.76	1.49	1.88	2.03	1.02	0.68	0.71	0.77	1.49	1.88	2.05
PFS	Rel P	0.00%	-16.89%	-19.52%	-22.96%	19.35%	118.53%	48.83%	0.00%	-28.36%	-30.56%	-33.37%	11.93%	103.02%	36.57%
PFS	ModelSE	3.43	5.10	5.07	5.03	3.87	2.84	3.16	3.43	5.10	5.07	5.03	3.87	2.84	3.16
PFS	Cover2	0.67	0.92	0.91	0.90	0.53	0.05	0.17	0.65	0.92	0.91	0.91	0.52	0.05	0.17
PFS	Cover1	0.90	1.00	1.00	1.00	0.89	0.50	0.71	0.90	1.00	1.00	1.00	0.89	0.50	0.71
OS	Mean	68.43	73.87	69.30	74.17	67.90	62.08	70.12	68.82	74.80	70.43	75.58	68.52	62.21	70.63
OS	Bias	-2.77	2.67	-1.89	2.98	-3.30	-9.11	-1.07	-2.37	3.60	-0.77	4.38	-2.67	-8.98	-0.57
OS	MC SE	0.17	0.14	0.14	0.20	0.13	0.09	0.18	0.17	0.26	0.21	0.24	0.28	0.09	0.18
OS	Rel Bias	-0.04	0.04	-0.03	0.04	-0.05	-0.13	0.02	-0.03	0.05	-0.01	0.06	-0.04	-0.13	0.01
OS	empSE	5.14	4.22	4.20	6.08	4.03	2.82	5.49	5.30	7.97	6.34	7.51	8.71	2.84	5.66
OS	MC SE	0.12	0.10	0.10	0.14	0.09	0.06	0.13	0.12	0.18	0.15	0.17	0.20	0.07	0.13
OS	MSE	34.06	24.94	21.17	45.79	27.07	90.97	31.22	33.65	76.37	40.75	75.48	82.98	88.72	32.32
OS	MC SE	0.12	0.10	0.10	0.14	0.09	0.06	0.13	1.15	32.34	8.66	14.18	40.42	1.64	1.08
OS	Rel P	0.00%	48.34%	50.00%	-28.56%	62.84%	232.46%	-12.24%	0.00%	-55.79%	-30.23%	-50.23%	-63.05%	247.16%	-12.43%
OS	ModelSE	5.03	12.09	12.68	12.21	12.38	3.31	5.56	5.03	12.09	12.68	12.21	12.38	3.31	5.56
OS	Cover2	0.79	1.00	0.88	0.95	0.84	0.22	0.89	0.81	0.99	0.89	0.97	0.86	0.24	0.91
OS	Cover1	0.97	1.00	0.99	1.00	0.99	0.78	0.98	0.97	1.00	0.99	1.00	0.99	0.78	0.98
QALYs	Mean	49.88	53.50	51.17	53.58	48.73	44.49	48.86	50.00	54.15	51.91	54.46	49.02	44.58	49.12
QALYs	Bias	-2.51	1.12	-1.21	1.20	-3.66	-7.90	-3.53	-2.39	1.76	-0.47	2.08	-3.36	-7.81	-3.26
QALYs	MC SE	0.11	0.10	0.11	0.13	0.09	0.07	0.11	0.12	0.15	0.13	0.15	0.16	0.07	0.11
QALYs	Rel Bias	-0.05	0.02	-0.02	0.02	-0.07	-0.15	0.07	-0.05	0.03	-0.01	0.04	-0.06	-0.15	0.06
QALYs	empSE	3.51	3.23	3.29	4.13	2.92	2.10	3.42	3.56	4.74	4.11	4.73	4.88	2.12	3.52
QALYs	MC SE	0.08	0.07	0.08	0.09	0.07	0.05	0.08	0.08	0.11	0.09	0.11	0.11	0.05	0.08
QALYs	MSE	18.59	11.69	12.26	18.45	21.92	66.76	24.11	18.40	25.52	17.09	26.64	35.10	65.42	23.00
QALYs	MC SE	0.67	0.42	0.46	0.67	0.71	1.07	0.83	0.67	8.07	2.13	3.38	9.92	1.07	0.80
QALYs	Rel P	0.00%	17.80%	14.04%	-27.73%	44.27%	179.60%	5.35%	0.00%	-43.35%	-24.72%	-43.12%	-46.66%	184.00%	2.68%
QALYs	ModelSE	3.03	6.78	7.05	6.92	6.66	2.42	3.53	3.03	6.78	7.05	6.92	6.66	2.42	3.53
QALYs	Cover2	0.71	0.99	0.88	0.93	0.74	0.11	0.70	0.71	0.99	0.90	0.94	0.76	0.12	0.71
QALYs	Cover1	0.90	1.00	0.99	1.00	0.97	0.67	0.97	0.90	1.00	0.99	1.00	0.97	0.67	0.97

Table 35 : Result for simulated scenario 13

	Scenario13	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	65.98	66.21	66.35	66.35	63.39	58.09	56.75	66.50	66.87	67.01	67.02	63.10	58.21	57.16
PFS	Bias	-3.36	-3.13	-2.99	-2.99	-5.94	-11.25	-12.59	-2.83	-2.47	-2.32	-2.32	-6.24	-11.13	-12.18
PFS	MC SE	0.20	0.21	0.21	0.21	0.20	0.12	0.14	0.21	0.22	0.22	0.22	0.22	0.13	0.14
PFS	Rel Bias	-4.8%	-4.5%	-4.3%	-4.3%	-8.6%	-16.2%	-18.2%	-4.1%	-3.6%	-3.4%	-3.3%	-9.0%	-16.1%	-17.6%
PFS	empSE	6.30	6.42	6.55	6.55	6.14	3.84	4.25	6.49	6.71	6.84	6.83	6.84	3.86	4.29
PFS	MC SE	0.14	0.15	0.15	0.15	0.14	0.09	0.10	0.15	0.15	0.16	0.16	0.16	0.09	0.10
PFS	MSE	50.95	51.00	51.85	51.71	72.95	141.22	176.62	50.13	51.08	52.18	52.05	85.71	138.86	166.85
PFS	MC SE	2.02	2.04	2.13	2.10	2.45	2.88	3.48	2.04	2.14	2.24	2.22	3.24	2.87	3.41
PFS	Rel P	0.00%	-3.63%	-7.49%	-7.25%	5.50%	169.57%	119.90%	0.00%	-6.40%	-10.01%	-9.80%	-10.06%	182.20%	128.50%
PFS	ModelSE	5.87	6.14	6.18	6.19	7.32	3.81	3.88	5.87	6.14	6.18	6.19	7.32	3.81	3.88
PFS	Cover2	0.81	0.82	0.82	0.83	0.69	0.20	0.18	0.82	0.84	0.84	0.84	0.69	0.20	0.19
PFS	Cover1	0.97	0.98	0.98	0.98	0.93	0.70	0.66	0.97	0.98	0.98	0.98	0.93	0.70	0.66
OS	Mean	264.53	297.10	270.83	262.88	293.83	270.05	263.75	266.72	311.97	281.49	269.39	308.06	272.81	268.27
OS	Bias	5.67	38.23	11.97	4.02	34.97	11.18	4.89	7.86	53.10	22.63	10.53	49.19	13.94	9.41
OS	MC SE	0.60	0.98	0.72	0.72	0.94	0.79	0.58	0.62	1.69	1.36	1.02	1.67	0.81	0.59
OS	Rel Bias	0.02	0.15	0.05	0.02	0.14	0.04	0.02	0.03	0.21	0.09	0.04	0.19	0.05	0.04
OS	empSE	18.35	30.17	22.26	22.29	28.94	24.20	17.82	19.17	52.11	42.03	31.36	51.51	24.88	18.15
OS	MC SE	0.42	0.69	0.51	0.51	0.66	0.56	0.41	0.44	1.20	0.96	0.72	1.18	0.57	0.42
OS	MSE	368.52	2370.89	638.28	512.63	2059.51	710.30	341.05	428.66	5532.18	2276.83	1093.14	5070.42	813.00	417.66
OS	MC SE	0.42	0.69	0.51	0.51	0.66	0.56	0.41	26.09	601.80	351.56	115.07	552.62	34.35	15.63
OS	Rel P	0.00%	-63.00%	-32.06%	-32.25%	-59.80%	-42.53%	6.05%	0.00%	-86.47%	-79.21%	-62.65%	-86.16%	-40.68%	11.45%
OS	ModelSE	31.48	68.09	59.06	48.92	70.73	29.46	31.95	31.48	68.09	59.06	48.92	70.73	29.46	31.95
OS	Cover2	1.00	0.85	1.00	1.00	0.89	0.99	1.00	1.00	0.84	1.00	1.00	0.87	0.99	1.00
OS	Cover1	1.00	0.96	0.99	1.00	0.97	0.99	1.00	1.00	0.96	0.99	1.00	0.97	0.99	1.00
QALYs	Mean	152.06	168.41	155.32	151.34	165.93	152.45	148.90	153.31	176.04	160.85	154.80	172.96	153.86	151.28
QALYs	Bias	1.83	18.18	5.09	1.11	15.70	2.22	-1.33	3.08	25.81	10.62	4.57	22.73	3.63	1.05
QALYs	MC SE	0.31	0.50	0.37	0.37	0.48	0.40	0.30	0.32	0.85	0.69	0.52	0.85	0.41	0.30
QALYs	Rel Bias	0.01	0.12	0.03	0.01	0.10	0.01	- 0.01	0.02	0.17	0.07	0.03	0.15	0.02	0.01
QALYs	empSE	9.54	15.43	11.36	11.40	14.80	12.33	9.12	9.93	26.27	21.22	15.97	26.06	12.67	9.30
QALYs	MC SE	0.22	0.35	0.26	0.26	0.34	0.28	0.21	0.23	0.60	0.49	0.37	0.60	0.29	0.21
QALYs	MSE	94.30	568.22	154.70	131.02	465.26	156.67	84.83	108.08	1355.54	562.41	275.56	1194.71	173.53	87.51
QALYs	MC SE	3.57	22.44	5.68	4.42	19.30	6.29	2.96	6.25	149.25	87.27	28.71	134.87	7.15	3.05
QALYs	Rel P	0.00%	-61.75%	-29.40%	-29.92%	-58.42%	-40.07%	9.52%	0.00%	-85.70%	-78.08%	-61.29%	-85.46%	-38.51%	14.11%
QALYs	ModelSE	15.92	34.21	29.65	24.59	36.01	14.98	16.16	15.92	34.21	29.65	24.59	36.01	14.98	16.16
QALYs	Cover2	1.00	0.86	1.00	1.00	0.91	0.97	1.00	1.00	0.86	1.00	1.00	0.89	0.98	1.00
QALYs	Cover1	1.00	0.97	0.99	1.00	0.98	1.00	1.00	1.00	0.97	0.99	1.00	0.98	1.00	1.00

Table 36 : Result for simulated scenario 14

	Scenario14	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	64.98	65.61	65.77	65.65	62.06	57.33	54.49	65.36	66.22	66.39	66.27	62.07	57.43	54.69
PFS	Bias	-4.21	-3.57	-3.42	-3.53	-7.12	-11.85	-14.70	-3.82	-2.96	-2.79	-2.92	-7.11	-11.75	-14.49
PFS	MC SE	0.20	0.20	0.20	0.21	0.19	0.12	0.13	0.20	0.21	0.21	0.22	0.21	0.12	0.13
PFS	Rel Bias	-6.1%	-5.2%	-4.9%	-5.1%	-10.3%	-17.1%	-21.2%	-5.5%	-4.3%	-4.0%	-4.2%	-10.3%	-17.0%	-20.9%
PFS	empSE	6.01	6.20	6.26	6.38	5.92	3.66	3.87	6.08	6.47	6.55	6.66	6.47	3.68	3.91
PFS	MC SE	0.14	0.14	0.14	0.15	0.14	0.08	0.09	0.14	0.15	0.15	0.15	0.15	0.08	0.09
PFS	MSE	53.82	51.17	50.89	53.13	85.71	153.89	230.91	51.58	50.59	50.69	52.85	92.41	151.69	225.33
PFS	MC SE	2.03	2.10	2.07	2.18	2.72	2.89	3.70	1.97	2.17	2.16	2.28	3.34	2.89	3.68
PFS	Rel P	0.00%	-6.02%	-7.87%	-11.18%	3.33%	169.74%	141.71%	0.00%	-11.63%	-13.82%	-16.64%	-11.64%	172.67%	141.66%
PFS	ModelSE	5.38	5.96	6.05	6.00	6.91	3.68	3.43	5.38	5.96	6.05	6.00	6.91	3.68	3.43
PFS	Cover2	0.77	0.80	0.80	0.79	0.61	0.14	0.06	0.78	0.81	0.82	0.81	0.63	0.15	0.06
PFS	Cover1	0.97	0.97	0.97	0.97	0.90	0.65	0.42	0.97	0.97	0.97	0.97	0.90	0.65	0.42
OS	Mean	170.49	179.13	169.41	166.44	176.43	163.73	167.80	171.33	183.33	173.18	167.49	179.23	164.48	169.12
OS	Bias	4.05	12.68	2.96	-0.01	9.98	-2.72	1.35	4.88	16.88	6.74	1.04	12.78	-1.97	2.68
OS	MC SE	0.37	0.37	0.40	0.42	0.37	0.31	0.35	0.38	0.93	0.78	0.45	0.94	0.31	0.36
OS	Rel Bias	0.02	0.08	0.02	-0.00	0.06	-0.02	0.01	0.03	0.10	0.04	0.01	0.08	-0.01	0.02
OS	empSE	11.45	11.27	12.37	13.02	11.25	9.52	10.78	11.67	28.52	24.00	13.98	29.05	9.67	10.97
OS	MC SE	0.26	0.26	0.28	0.30	0.26	0.22	0.25	0.27	0.65	0.55	0.32	0.67	0.22	0.25
OS	MSE	147.43	287.75	161.62	169.45	226.10	97.86	117.92	160.00	1097.46	620.56	196.31	1006.15	97.24	127.48
OS	MC SE	0.26	0.26	0.28	0.30	0.26	0.22	0.25	5.86	409.25	147.23	8.41	405.31	3.57	4.45
OS	Rel P	0.00%	3.25%	-14.27%	-22.66%	3.65%	44.88%	12.89%	0.00%	-83.24%	-76.33%	-30.26%	-83.85%	45.81%	13.15%
OS	ModelSE	12.79	30.58	32.95	14.30	31.80	11.34	12.71	12.79	30.58	32.95	14.30	31.80	11.34	12.71
OS	Cover2	0.99	0.93	0.99	0.92	0.97	0.94	0.99	0.99	0.91	0.99	0.93	0.96	0.95	0.99
OS	Cover1	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00
QALYs	Mean	104.74	109.25	104.43	102.91	106.83	99.06	100.25	105.27	111.53	106.51	103.63	108.23	99.47	100.97
QALYs	Bias	0.76	5.27	0.45	-1.06	2.85	-4.92	-3.73	1.29	7.55	2.53	-0.35	4.25	-4.51	-3.01
QALYs	MC SE	0.21	0.21	0.22	0.23	0.21	0.17	0.19	0.21	0.48	0.40	0.25	0.49	0.17	0.19
QALYs	Rel Bias	0.01	0.05	0.00	0.01	0.03	0.05	0.04	0.01	0.07	0.02	0.00	0.04	0.04	0.03
QALYs	empSE	6.47	6.53	6.73	7.19	6.45	5.16	5.77	6.61	14.75	12.36	7.73	15.09	5.24	5.87
QALYs	MC SE	0.15	0.15	0.15	0.17	0.15	0.12	0.13	0.15	0.34	0.28	0.18	0.35	0.12	0.13
QALYs	MSE	42.41	70.34	45.52	52.77	49.74	50.81	47.23	45.34	274.52	158.98	59.80	245.57	47.78	43.47
QALYs	MC SE	1.50	2.97	1.68	2.00	2.09	1.87	1.74	1.63	101.45	36.26	2.46	98.25	1.79	1.59
QALYs	Rel P	0.00%	-1.74%	-7.69%	-18.99%	0.56%	56.97%	25.63%	0.00%	-79.92%	-71.38%	-26.83%	-80.80%	59.11%	26.87%
QALYs	ModelSE	6.67	15.68	16.80	7.84	16.83	6.11	6.68	6.67	15.68	16.80	7.84	16.83	6.11	6.68
QALYs	Cover2	0.96	0.96	0.97	0.90	0.99	0.84	0.86	0.97	0.95	0.98	0.91	0.99	0.85	0.88
QALYs	Cover1	0.99	0.99	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00	0.99	0.99

Table 37 : Result for simulated scenario 15

	Scenario15	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	64.51	65.67	65.82	65.76	61.43	57.59	53.65	64.63	66.28	66.45	66.39	61.73	57.69	53.87
PFS	Bias	-4.67	-3.51	-3.36	-3.42	-7.75	-11.59	-15.53	-4.55	-2.90	-2.73	-2.79	-7.45	-11.49	-15.31
PFS	MC SE	0.19	0.19	0.21	0.21	0.19	0.12	0.12	0.18	0.20	0.22	0.21	0.21	0.12	0.12
PFS	Rel Bias	-6.8%	-5.1%	-4.9%	-4.9%	-11.2%	-16.8%	-22.4%	-6.6%	-4.2%	-4.0%	-4.0%	-10.8%	-16.6%	-22.1%
PFS	empSE	5.72	5.96	6.43	6.33	5.82	3.63	3.76	5.61	6.22	6.71	6.61	6.37	3.64	3.79
PFS	MC SE	0.13	0.14	0.15	0.15	0.13	0.08	0.09	0.13	0.14	0.15	0.15	0.15	0.08	0.09
PFS	MSE	54.51	47.82	52.57	51.68	93.98	147.53	255.23	52.13	47.04	52.50	51.39	95.98	145.37	248.63
PFS	MC SE	1.95	1.88	2.11	2.07	2.84	2.79	3.81	1.89	1.91	2.20	2.14	3.43	2.78	3.80
PFS	Rel P	0.00%	-7.90%	-20.80%	-18.28%	-3.43%	148.89%	131.71%	0.00%	-18.66%	-30.23%	-27.91%	-22.38%	137.11%	119.34%
PFS	ModelSE	4.84	5.93	6.05	6.01	6.51	3.69	3.26	4.84	5.93	6.05	6.01	6.51	3.69	3.26
PFS	Cover2	0.71	0.81	0.80	0.81	0.56	0.15	0.03	0.72	0.83	0.81	0.82	0.59	0.15	0.03
PFS	Cover1	0.94	0.98	0.98	0.98	0.86	0.66	0.30	0.94	0.98	0.98	0.98	0.86	0.66	0.30
OS	Mean	130.45	136.69	128.66	128.52	132.36	124.80	128.78	130.94	137.17	129.42	129.30	132.56	125.09	129.50
OS	Bias	-1.89	4.34	-3.68	-3.82	0.01	-7.55	-3.57	-1.41	4.83	-2.92	-3.05	0.21	-7.25	-2.84
OS	MC SE	0.24	0.23	0.29	0.28	0.24	0.18	0.24	0.25	0.28	0.30	0.35	0.32	0.18	0.24
OS	Rel Bias	- 0.01	0.03 -	0.03 -	0.03	0.00 -	0.06 -	0.03	- 0.01	0.04 -	0.02 -	0.02	0.00 -	0.05 -	0.02
OS	empSE	7.52	7.13	8.91	8.62	7.52	5.61	7.37	7.67	8.76	9.23	10.86	9.94	5.65	7.52
OS	MC SE	0.17	0.16	0.20	0.20	0.17	0.13	0.17	0.18	0.20	0.21	0.25	0.23	0.13	0.17
OS	MSE	60.02	69.66	92.89	88.90	56.48	88.40	67.03	60.70	99.86	93.62	127.18	98.78	84.53	64.58
OS	MC SE	0.17	0.16	0.20	0.20	0.17	0.13	0.17	2.11	20.75	3.25	36.82	20.68	2.83	2.39
OS	Rel P	0.00%	11.07%	-28.86%	-24.03%	-0.09%	79.68%	3.92%	0.00%	-23.31%	-30.98%	-50.19%	-40.53%	84.09%	3.94%
OS	ModelSE	7.36	11.57	8.81	11.59	13.92	6.49	7.38	7.36	11.57	8.81	11.59	13.92	6.49	7.38
OS	Cover2	0.87	0.98	0.85	0.83	0.95	0.74	0.83	0.88	0.98	0.87	0.85	0.97	0.76	0.84
OS	Cover1	0.98	1.00	1.00	0.99	1.00	0.98	0.99	0.98	1.00	1.00	0.99	1.00	0.98	0.99
QALYs	Mean	84.58	88.04	84.08	83.99	84.61	79.67	80.48	84.86	88.47	84.64	84.56	84.80	79.85	80.91
QALYs	Bias	-2.35	1.12	-2.85	-2.94	-2.32	-7.25	-6.44	-2.07	1.54	-2.28	-2.36	-2.13	-7.08	-6.01
QALYs	MC SE	0.15	0.16	0.18	0.18	0.16	0.11	0.14	0.16	0.18	0.19	0.21	0.20	0.11	0.14
QALYs	Rel Bias	- 0.03	0.01 -	0.03 -	0.03 -	0.03 -	0.08 -	0.07	- 0.02	0.02 -	0.03 -	0.03 -	0.02 -	0.08 -	0.07
QALYs	empSE	4.78	4.85	5.57	5.53	4.98	3.41	4.17	4.84	5.60	5.80	6.62	6.28	3.43	4.25
QALYs	MC SE	0.11	0.11	0.13	0.13	0.11	0.08	0.10	0.11	0.13	0.13	0.15	0.14	0.08	0.10
QALYs	MSE	28.30	24.74	39.07	39.15	30.20	64.20	58.90	27.66	33.76	38.77	49.34	43.94	61.84	54.23
QALYs	MC SE	1.06	1.05	1.34	1.40	1.19	1.65	1.82	1.03	5.55	1.41	10.29	5.77	1.62	1.75
QALYs	Rel P	0.00%	-3.00%	-26.39%	-25.39%	-8.17%	96.47%	30.98%	0.00%	-25.49%	-30.33%	-46.56%	-40.66%	98.62%	29.43%
QALYs	ModelSE	4.07	6.78	5.62	6.76	8.35	3.92	4.19	4.07	6.78	5.62	6.76	8.35	3.92	4.19
QALYs	Cover2	0.77	0.98	0.84	0.81	0.88	0.53	0.60	0.79	0.98	0.86	0.83	0.90	0.54	0.63
QALYs	Cover1	0.93	1.00	0.99	0.97	0.99	0.92	0.92	0.93	1.00	0.99	0.97	0.99	0.92	0.92

Table 38 : Result for simulated scenario 16

	Scenario16	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.23	59.06	59.06	59.00	53.98	50.35	50.16	58.63	59.58	59.58	59.53	53.97	50.42	50.16
PFS	Bias	2.08	2.91	2.90	2.85	-2.17	-5.80	-6.00	2.47	3.43	3.42	3.37	-2.19	-5.73	-6.00
PFS	MC SE	0.16	0.17	0.17	0.17	0.14	0.10	0.11	0.16	0.17	0.17	0.17	0.15	0.10	0.11
PFS	Rel Bias	3.7%	5.2%	5.2%	5.1%	-3.9%	-10.3%	-10.7%	4.4%	6.1%	6.1%	6.0%	-3.9%	-10.2%	-10.7%
PFS	empSE	4.88	5.20	5.21	5.23	4.42	3.08	3.47	4.95	5.33	5.33	5.36	4.56	3.10	3.44
PFS	MC SE	0.11	0.12	0.12	0.12	0.10	0.07	0.08	0.11	0.12	0.12	0.12	0.10	0.07	0.08
PFS	MSE	28.09	35.48	35.48	35.44	24.23	43.14	47.95	30.59	40.10	40.11	40.03	25.54	42.45	47.85
PFS	MC SE	1.48	1.95	1.93	1.93	1.05	1.24	1.42	1.59	2.20	2.18	2.19	1.16	1.23	1.39
PFS	Rel P	0.00%	-12.05%	-12.15%	-13.02%	21.89%	150.93%	98.18%	0.00%	-13.69%	-13.82%	-14.66%	17.87%	155.25%	106.77%
PFS	ModelSE	4.61	4.81	4.82	4.82	4.16	3.10	3.12	4.61	4.81	4.82	4.82	4.16	3.10	3.12
PFS	Cover2	0.93	0.92	0.92	0.92	0.84	0.53	0.48	0.93	0.91	0.91	0.91	0.83	0.55	0.48
PFS	Cover1	0.99	0.99	0.99	0.99	0.98	0.91	0.89	0.99	0.99	0.99	0.99	0.98	0.91	0.89
OS	Mean	232.37	272.55	242.19	234.38	271.49	242.65	231.71	264.31	296.69	267.57	252.15	295.30	246.30	241.87
OS	Bias	4.05	44.23	13.87	6.06	43.17	14.33	3.39	35.99	68.37	39.25	23.83	66.98	17.98	13.55
OS	MC SE	0.73	1.23	0.84	0.76	1.20	0.84	0.85	2.67	1.83	2.11	1.49	1.87	0.87	0.82
OS	Rel Bias	0.02	0.19	0.06	0.03	0.19	0.06	0.01	0.16	0.30	0.17	0.10	0.29	0.08	0.06
OS	empSE	22.44	37.83	25.75	23.57	36.89	25.80	26.04	82.33	56.43	64.96	46.04	57.75	26.74	25.37
OS	MC SE	0.52	0.87	0.59	0.54	0.85	0.59	0.60	1.89	1.30	1.49	1.06	1.33	0.61	0.58
OS	MSE	519.40	3386.34	854.66	591.56	3223.35	870.31	688.78	8066.33	7855.67	5756.05	2685.69	7817.94	1037.30	826.72
OS	MC SE	0.52	0.87	0.59	0.54	0.85	0.59	0.60	1185.42	559.67	703.83	368.83	580.75	45.40	38.25
OS	Rel P	0.00%	-64.82%	-24.04%	-9.34%	-63.00%	-24.34%	-25.73%	0.00%	112.85%	60.62%	219.73%	103.21%	848.34%	952.81%
OS	ModelSE	86.28	81.51	76.60	67.27	82.16	32.10	55.12	86.28	81.51	76.60	67.27	82.16	32.10	55.12
OS	Cover2	1.00	0.84	1.00	1.00	0.86	0.99	1.00	0.99	0.82	0.99	1.00	0.85	1.00	1.00
OS	Cover1	0.99	0.95	0.99	1.00	0.96	1.00	0.99	0.99	0.95	0.99	1.00	0.96	1.00	0.99
QALYs	Mean	133.65	153.99	138.81	134.89	151.94	136.43	130.90	149.74	166.22	151.66	143.93	163.84	138.28	135.98
QALYs	Bias	2.65	22.99	7.81	3.88	20.93	5.43	-0.10	18.74	35.21	20.65	12.93	32.83	7.27	4.97
QALYs	MC SE	0.37	0.62	0.42	0.39	0.60	0.42	0.43	1.34	0.92	1.05	0.75	0.94	0.44	0.42
QALYs	Rel Bias	0.02	0.18	0.06	0.03	0.16	0.04 -	0.00	0.14	0.27	0.16	0.10	0.25	0.06	0.04
QALYs	empSE	11.50	19.10	13.03	11.88	18.61	13.06	13.16	41.29	28.31	32.48	23.04	28.98	13.53	12.82
QALYs	MC SE	0.26	0.44	0.30	0.27	0.43	0.30	0.30	0.95	0.65	0.75	0.53	0.67	0.31	0.29
QALYs	MSE	139.21	892.77	230.46	156.09	784.06	199.82	172.97	2054.12	2040.32	1480.12	697.16	1916.79	235.67	189.04
QALYs	MC SE	5.32	35.86	9.20	6.22	32.47	8.73	6.24	299.63	141.38	177.15	93.07	143.94	10.39	8.73
QALYs	Rel P	0.00%	-63.72%	-22.01%	-6.25%	-61.77%	-22.40%	-23.57%	0.00%	112.79%	61.64%	221.29%	103.04%	831.62%	936.60%
QALYs	ModelSE	43.35	40.79	38.31	33.63	41.17	16.22	27.66	43.35	40.79	38.31	33.63	41.17	16.22	27.66
QALYs	Cover2	1.00	0.83	1.00	1.00	0.87	0.98	1.00	0.99	0.82	0.99	1.00	0.86	0.99	1.00
QALYs	Cover1	0.99	0.95	0.99	0.99	0.97	1.00	0.99	0.99	0.95	0.99	0.99	0.97	1.00	0.99

Table 39 : Result for simulated scenario 17

	Scenario17	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	56.00	58.82	58.85	58.86	53.36	49.96	47.92	56.25	59.34	59.37	59.38	53.43	50.02	47.83
PFS	Bias	0.01	2.83	2.86	2.87	-2.62	-6.03	-8.06	0.26	3.35	3.38	3.39	-2.55	-5.96	-8.15
PFS	MC SE	0.13	0.16	0.16	0.16	0.14	0.10	0.10	0.13	0.17	0.17	0.17	0.14	0.10	0.10
PFS	Rel Bias	0.0%	5.1%	5.1%	5.1%	-4.7%	-10.8%	-14.4%	0.5%	6.0%	6.0%	6.1%	-4.6%	-10.7%	-14.6%
PFS	empSE	3.98	5.03	5.02	5.05	4.35	2.99	3.06	4.00	5.14	5.13	5.16	4.45	3.00	3.05
PFS	MC SE	0.09	0.12	0.12	0.12	0.10	0.07	0.07	0.09	0.12	0.12	0.12	0.10	0.07	0.07
PFS	MSE	15.86	33.31	33.41	33.74	25.82	45.29	74.33	16.04	37.59	37.75	38.07	26.26	44.59	75.74
PFS	MC SE	0.74	1.59	1.60	1.62	1.06	1.25	1.64	0.74	1.77	1.78	1.79	1.10	1.24	1.65
PFS	Rel P	0.00%	-37.30%	-37.09%	-37.80%	-16.24%	77.55%	69.79%	0.00%	-39.45%	-39.27%	-39.91%	-19.09%	77.15%	72.09%
PFS	ModelSE	3.83	4.77	4.78	4.78	4.03	3.05	2.75	3.83	4.77	4.78	4.78	4.03	3.05	2.75
PFS	Cover2	0.93	0.93	0.93	0.93	0.80	0.50	0.24	0.93	0.91	0.92	0.91	0.80	0.51	0.22
PFS	Cover1	1.00	1.00	1.00	1.00	0.98	0.90	0.73	1.00	1.00	1.00	1.00	0.98	0.90	0.73
OS	Mean	132.42	151.53	143.21	141.42	148.79	137.49	131.67	134.09	158.64	149.21	143.74	154.40	138.27	132.32
OS	Bias	-3.83	15.28	6.96	5.17	12.54	1.24	-4.58	-2.16	22.39	12.96	7.49	18.15	2.02	-3.93
OS	MC SE	0.23	0.34	0.35	0.36	0.34	0.28	0.23	0.35	0.93	0.86	0.39	0.90	0.28	0.23
OS	Rel Bias	- 0.03	0.11	0.05	0.04	0.09	0.01	- 0.03	- 0.02	0.16	0.10	0.06	0.13	0.01	- 0.03
OS	empSE	6.98	10.55	10.82	11.15	10.41	8.56	7.11	10.83	28.66	26.39	11.88	27.60	8.70	7.24
OS	MC SE	0.16	0.24	0.25	0.26	0.24	0.20	0.16	0.25	0.66	0.61	0.27	0.63	0.20	0.17
OS	MSE	63.38	344.63	165.29	150.91	265.36	74.81	71.46	121.86	1321.49	863.42	197.04	1090.52	79.74	67.77
OS	MC SE	0.16	0.24	0.25	0.26	0.24	0.20	0.16	30.34	256.04	240.68	9.43	254.34	3.34	2.38
OS	Rel P	0.00%	-56.23%	-58.36%	-60.79%	-55.01%	-33.57%	-3.55%	0.00%	-85.71%	-83.15%	-16.82%	-84.61%	54.88%	123.90%
OS	ModelSE	18.15	36.09	32.22	14.49	34.27	10.35	10.61	18.15	36.09	32.22	14.49	34.27	10.35	10.61
OS	Cover2	0.97	0.83	1.00	0.98	0.92	0.99	0.96	0.98	0.82	1.00	0.98	0.92	0.99	0.97
OS	Cover1	1.00	0.97	1.00	1.00	0.98	1.00	1.00	1.00	0.97	1.00	1.00	0.98	1.00	1.00
QALYs	Mean	83.01	93.41	89.26	88.37	90.40	83.73	80.21	83.92	97.12	92.42	89.69	93.23	84.14	80.51
QALYs	Bias	-1.91	8.49	4.34	3.45	5.48	-1.19	-4.71	-1.00	12.20	7.49	4.76	8.31	-0.78	-4.41
QALYs	MC SE	0.13	0.19	0.19	0.19	0.18	0.15	0.13	0.19	0.48	0.43	0.21	0.46	0.15	0.13
QALYs	Rel Bias	- 0.02	0.10	0.05	0.04	0.06	- 0.01	- 0.06	- 0.01	0.14	0.09	0.06	0.10	- 0.01	- 0.05
QALYs	empSE	4.03	5.85	5.82	5.97	5.65	4.60	3.87	5.80	14.68	13.34	6.36	14.09	4.67	3.94
QALYs	MC SE	0.09	0.13	0.13	0.14	0.13	0.11	0.09	0.13	0.34	0.31	0.15	0.32	0.11	0.09
QALYs	MSE	19.90	106.27	52.59	47.47	61.96	22.57	37.15	34.65	364.18	234.01	63.10	267.23	22.42	34.94
QALYs	MC SE	0.74	3.67	2.19	2.06	2.69	0.85	1.23	7.75	66.46	60.40	2.81	64.11	0.86	1.18
QALYs	Rel P	0.00%	-52.56%	-51.94%	-54.37%	-49.13%	-23.24%	8.41%	0.00%	-84.38%	-81.08%	-16.73%	-83.03%	54.28%	117.19%
QALYs	ModelSE	9.31	18.21	16.28	7.57	17.33	5.51	5.56	9.31	18.21	16.28	7.57	17.33	5.51	5.56
QALYs	Cover2	0.96	0.82	0.99	0.96	0.96	0.95	0.87	0.97	0.79	0.98	0.95	0.96	0.96	0.87
QALYs	Cover1	1.00	0.96	0.99	0.99	0.99	1.00	1.00	1.00	0.96	0.99	0.99	0.99	1.00	1.00

Table 40 : Result for simulated scenario 18

	Scenario18	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	53.77	58.87	58.86	58.85	53.19	50.11	46.80	53.91	59.39	59.38	59.36	53.23	50.16	46.72
PFS	Bias	-2.19	2.91	2.90	2.89	-2.77	-5.86	-9.17	-2.06	3.42	3.41	3.40	-2.73	-5.80	-9.24
PFS	MC SE	0.11	0.16	0.16	0.16	0.13	0.09	0.09	0.11	0.16	0.17	0.17	0.14	0.10	0.09
PFS	Rel Bias	-3.9%	5.2%	5.2%	5.2%	-4.9%	-10.5%	-16.4%	-3.7%	6.1%	6.1%	6.1%	-4.9%	-10.4%	-16.5%
PFS	empSE	3.30	4.84	5.02	5.03	4.14	2.92	2.88	3.29	4.94	5.13	5.14	4.25	2.93	2.89
PFS	MC SE	0.08	0.11	0.12	0.12	0.10	0.07	0.07	0.08	0.11	0.12	0.12	0.10	0.07	0.07
PFS	MSE	15.69	31.83	33.55	33.60	24.79	42.81	92.32	15.06	36.15	37.92	37.97	25.56	42.22	93.69
PFS	MC SE	0.68	1.45	1.53	1.58	1.00	1.20	1.76	0.66	1.62	1.71	1.75	1.05	1.20	1.78
PFS	Rel P	0.00%	-53.47%	-56.76%	-56.95%	-36.46%	27.90%	31.15%	0.00%	-55.62%	-58.76%	-58.95%	-40.07%	25.96%	30.01%
PFS	ModelSE	3.22	4.76	4.77	4.76	3.95	3.05	2.55	3.22	4.76	4.77	4.76	3.95	3.05	2.55
PFS	Cover2	0.86	0.94	0.93	0.92	0.79	0.52	0.10	0.86	0.92	0.91	0.91	0.79	0.52	0.10
PFS	Cover1	0.98	0.99	0.99	0.99	0.97	0.90	0.57	0.98	0.99	0.99	0.99	0.97	0.90	0.57
OS	Mean	99.01	114.03	107.14	107.04	109.52	102.75	98.24	99.35	115.14	108.71	108.43	109.99	103.02	98.41
OS	Bias	-6.65	8.37	1.48	1.38	3.86	-2.91	-7.42	-6.31	9.49	3.05	2.77	4.33	-2.64	-7.24
OS	MC SE	0.15	0.18	0.22	0.21	0.18	0.15	0.15	0.15	0.27	0.34	0.31	0.20	0.16	0.15
OS	Rel Bias	- 0.06	0.08	0.01	0.01	0.04	- 0.03	- 0.07	- 0.06	0.09	0.03	0.03	0.04	- 0.02	- 0.07
OS	empSE	4.57	5.65	6.84	6.55	5.63	4.72	4.55	4.63	8.45	10.45	9.69	6.22	4.78	4.61
OS	MC SE	0.10	0.13	0.16	0.15	0.13	0.11	0.10	0.11	0.19	0.24	0.22	0.14	0.11	0.11
OS	MSE	65.07	101.98	48.98	44.73	46.57	30.77	75.73	61.24	161.37	118.38	101.42	57.48	29.78	73.72
OS	MC SE	0.10	0.13	0.16	0.15	0.13	0.11	0.10	1.73	38.96	40.23	41.83	3.01	1.10	1.97
OS	Rel P	0.00%	-34.63%	-55.36%	-51.24%	-34.02%	-6.37%	0.90%	0.00%	-70.00%	-80.38%	-77.17%	-44.68%	-6.17%	0.67%
OS	ModelSE	5.05	11.58	13.90	11.73	8.50	5.53	5.02	5.05	11.58	13.90	11.73	8.50	5.53	5.02
OS	Cover2	0.69	0.84	0.97	0.95	0.98	0.91	0.63	0.71	0.82	0.96	0.94	0.97	0.91	0.65
OS	Cover1	0.99	0.98	1.00	1.00	1.00	1.00	0.98	0.99	0.98	1.00	1.00	1.00	1.00	0.98
QALYs	Mean	65.64	74.67	71.23	71.17	70.72	66.40	63.16	65.84	75.39	72.17	72.02	70.96	66.56	63.22
QALYs	Bias	-3.98	5.06	1.61	1.56	1.10	-3.21	-6.46	-3.77	5.77	2.55	2.41	1.35	-3.06	-6.39
QALYs	MC SE	0.09	0.12	0.14	0.13	0.12	0.09	0.09	0.09	0.16	0.19	0.18	0.13	0.09	0.09
QALYs	Rel Bias	- 0.06	0.07	0.02	0.02	0.02	- 0.05	- 0.09	- 0.05	0.08	0.04	0.03	0.02	- 0.04	- 0.09
QALYs	empSE	2.86	3.79	4.25	4.11	3.56	2.83	2.72	2.89	4.93	5.89	5.52	3.87	2.85	2.75
QALYs	MC SE	0.07	0.09	0.10	0.09	0.08	0.06	0.06	0.07	0.11	0.14	0.13	0.09	0.07	0.06
QALYs	MSE	24.00	39.89	20.60	19.31	13.89	18.30	49.09	22.55	57.59	41.12	36.25	16.80	17.49	48.40
QALYs	MC SE	0.71	1.42	0.88	0.90	0.62	0.64	1.09	0.68	9.75	10.73	10.91	0.90	0.62	1.09
QALYs	Rel P	0.00%	-43.08%	-54.73%	-51.73%	-35.69%	2.16%	10.66%	0.00%	-65.72%	-75.96%	-72.68%	-44.45%	2.28%	10.52%
QALYs	ModelSE	2.85	6.39	7.43	6.40	4.93	3.32	2.91	2.85	6.39	7.43	6.40	4.93	3.32	2.91
QALYs	Cover2	0.64	0.86	0.96	0.94	0.99	0.81	0.36	0.67	0.83	0.95	0.91	0.99	0.82	0.37
QALYs	Cover1	0.97	0.98	1.00	0.99	1.00	0.99	0.90	0.97	0.98	1.00	0.99	1.00	0.99	0.90

Table 41 : Result for simulated scenario 19

	Scenario19	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.94	60.21	60.30	60.14	56.36	47.29	54.45	59.17	61.01	61.11	60.95	55.83	47.48	54.70
PFS	Bias	-10.14	-8.87	-8.79	-8.94	-12.72	-21.79	-14.63	-9.92	-8.07	-7.97	-8.13	-13.26	-21.60	-14.38
PFS	MC SE	0.21	0.23	0.24	0.24	0.20	0.12	0.17	0.20	0.24	0.25	0.25	0.21	0.12	0.18
PFS	Rel Bias	-14.7%	-12.8%	-12.7%	-12.9%	-18.4%	-31.5%	-21.2%	-14.4%	-11.7%	-11.5%	-11.8%	-19.2%	-31.3%	-20.8%
PFS	empSE	6.39	7.17	7.45	7.45	6.17	3.70	5.37	6.26	7.43	7.76	7.76	6.61	3.75	5.43
PFS	MC SE	0.15	0.16	0.17	0.17	0.14	0.08	0.12	0.14	0.17	0.18	0.18	0.15	0.09	0.12
PFS	MSE	143.60	130.10	132.69	135.38	199.79	488.55	242.93	137.45	120.18	123.65	126.27	219.40	480.61	236.34
PFS	MC SE	3.93	3.92	4.20	4.19	4.64	5.05	4.84	3.80	3.80	4.38	4.38	5.47	5.07	4.80
PFS	Rel P	0.00%	-20.68%	-26.55%	-26.44%	7.24%	198.02%	41.69%	0.00%	-28.96%	-34.86%	-34.89%	-10.32%	178.34%	32.65%
PFS	ModelSE	6.13	7.07	7.19	7.14	7.81	3.76	5.85	6.13	7.07	7.19	7.14	7.81	3.76	5.85
PFS	Cover2	0.52	0.63	0.63	0.61	0.44	0.01	0.28	0.54	0.66	0.66	0.65	0.45	0.01	0.30
PFS	Cover1	0.89	0.92	0.92	0.92	0.84	0.23	0.84	0.89	0.92	0.92	0.92	0.84	0.23	0.84
OS	Mean	118.62	115.54	119.52	131.85	113.34	83.13	126.12	122.54	117.15	132.78	157.85	114.58	83.79	131.42
OS	Bias	-2.99	-6.07	-2.09	10.24	-8.26	-38.48	4.51	0.93	-4.46	11.17	36.24	-7.03	-37.82	9.81
OS	MC SE	0.35	0.32	0.41	0.85	0.32	0.18	0.40	0.73	0.44	1.56	1.36	0.66	0.19	0.46
OS	Rel Bias	-0.02	-0.05	0.02	0.08	-0.07	0.32	0.04	0.01	-0.04	0.09	0.30	-0.06	0.31	0.08
OS	empSE	10.89	9.91	12.74	26.12	9.81	5.60	12.40	22.63	13.58	48.23	42.02	20.38	5.76	14.11
OS	MC SE	0.25	0.23	0.29	0.60	0.23	0.13	0.28	0.52	0.31	1.11	0.96	0.47	0.13	0.32
OS	MSE	127.41	134.94	166.57	786.33	164.40	1512.27	173.85	512.50	204.20	2448.80	3077.30	464.46	1463.63	295.02
OS	MC SE	0.25	0.23	0.29	0.60	0.23	0.13	0.28	268.43	62.16	560.92	355.03	180.47	14.07	12.58
OS	Rel P	0.00%	20.73%	-26.95%	-82.61%	23.30%	278.52%	-22.82%	0.00%	177.57%	-77.99%	-70.99%	23.26%	1443.28%	157.27%
OS	ModelSE	29.80	19.06	54.00	76.55	26.02	6.93	29.63	29.80	19.06	54.00	76.55	26.02	6.93	29.63
OS	Cover2	1.00	0.95	1.00	1.00	0.93	0.00	1.00	1.00	0.96	1.00	1.00	0.94	0.00	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.31	1.00	1.00	1.00	1.00	1.00	1.00	0.31	1.00
QALYs	Mean	76.99	75.84	77.85	83.97	73.58	55.75	79.39	79.02	76.88	84.72	97.21	74.04	56.14	82.12
QALYs	Bias	-4.54	-5.69	-3.68	2.44	-7.95	-25.78	-2.14	-2.51	-4.65	3.19	15.68	-7.49	-25.39	0.59
QALYs	MC SE	0.20	0.19	0.22	0.43	0.18	0.12	0.23	0.38	0.25	0.79	0.69	0.35	0.12	0.25
QALYs	Rel Bias	-0.06	-0.07	0.05	0.03	-0.10	0.32	0.03	-0.03	-0.06	0.04	0.19	-0.09	0.31	0.01
QALYs	empSE	6.16	5.95	6.93	13.27	5.63	3.59	6.93	11.82	7.75	24.23	21.29	10.82	3.68	7.81
QALYs	MC SE	0.14	0.14	0.16	0.30	0.13	0.08	0.16	0.27	0.18	0.56	0.49	0.25	0.08	0.18
QALYs	MSE	58.46	67.80	61.56	181.88	94.80	677.44	52.56	145.88	81.56	596.64	698.69	173.05	658.20	61.35
QALYs	MC SE	2.26	2.30	2.25	11.93	2.85	5.97	2.12	67.35	17.05	135.21	86.42	44.98	6.02	2.61
QALYs	Rel P	0.00%	7.13%	-21.08%	-78.46%	19.81%	194.36%	-21.06%	0.00%	132.84%	-76.20%	-69.17%	19.36%	933.03%	128.82%
QALYs	ModelSE	15.28	10.42	27.35	38.60	14.15	4.26	15.73	15.28	10.42	27.35	38.60	14.15	4.26	15.73
QALYs	Cover2	0.98	0.89	0.99	0.99	0.84	0.00	1.00	0.98	0.90	0.99	1.00	0.85	0.00	1.00
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.22	1.00	1.00	1.00	1.00	1.00	1.00	0.22	1.00

Table 42 : Result for simulated scenario 20

	Scenario20	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.89	60.09	60.13	60.10	55.77	48.37	53.64	58.82	60.89	60.93	60.90	55.41	48.55	53.85
PFS	Bias	-10.18	-8.97	-8.93	-8.97	-13.30	-20.70	-15.43	-10.24	-8.17	-8.13	-8.16	-13.66	-20.52	-15.22
PFS	MC SE	0.19	0.22	0.22	0.23	0.19	0.12	0.16	0.19	0.23	0.23	0.24	0.20	0.12	0.17
PFS	Rel Bias	-14.7%	-13.0%	-12.9%	-13.0%	-19.3%	-30.0%	-22.3%	-14.8%	-11.8%	-11.8%	-11.8%	-19.8%	-29.7%	-22.0%
PFS	empSE	6.01	6.73	6.80	7.03	5.80	3.59	5.07	5.73	7.01	7.08	7.32	6.05	3.63	5.13
PFS	MC SE	0.14	0.15	0.16	0.16	0.13	0.08	0.12	0.13	0.16	0.16	0.17	0.14	0.08	0.12
PFS	MSE	139.63	125.77	126.08	129.78	210.49	441.23	263.59	137.68	115.81	116.24	120.08	223.20	434.10	257.76
PFS	MC SE	3.85	3.83	3.88	3.90	4.62	4.72	4.98	3.74	3.71	3.77	3.79	5.08	4.73	4.96
PFS	Rel P	0.00%	-20.47%	-22.12%	-27.03%	7.01%	180.25%	40.40%	0.00%	-33.12%	-34.52%	-38.70%	-10.45%	149.06%	24.90%
PFS	ModelSE	5.73	7.01	7.02	7.01	7.19	3.88	5.46	5.73	7.01	7.02	7.01	7.19	3.88	5.46
PFS	Cover2	0.51	0.63	0.63	0.63	0.39	0.01	0.23	0.51	0.66	0.66	0.65	0.39	0.01	0.23
PFS	Cover1	0.87	0.92	0.92	0.92	0.81	0.30	0.79	0.87	0.92	0.92	0.92	0.81	0.30	0.79
OS	Mean	90.27	88.72	89.55	96.19	85.93	68.94	93.29	91.85	93.17	95.54	109.71	88.51	69.24	95.04
OS	Bias	-5.04	-6.59	-5.76	0.87	-9.38	-26.37	-2.02	-3.47	-2.14	0.23	14.39	-6.80	-26.07	-0.27
OS	MC SE	0.29	0.25	0.27	0.46	0.24	0.13	0.30	0.32	0.94	0.91	0.94	0.72	0.14	0.32
OS	Rel Bias	-0.05	-0.07	0.06	0.01	-0.10	-0.28	0.02	-0.04	0.02	0.00	0.15	-0.07	-0.27	0.00
OS	empSE	8.91	7.79	8.24	14.23	7.32	4.12	9.38	9.91	29.10	28.12	29.07	22.07	4.18	9.89
OS	MC SE	0.20	0.18	0.19	0.33	0.17	0.09	0.22	0.23	0.67	0.65	0.67	0.51	0.10	0.23
OS	MSE	104.73	104.14	101.11	203.11	141.48	712.43	92.04	110.14	850.35	789.94	1051.34	532.96	697.04	97.74
OS	MC SE	0.20	0.18	0.19	0.33	0.17	0.09	0.22	10.19	218.16	201.60	255.63	153.63	7.00	3.81
OS	Rel P	0.00%	30.67%	16.79%	-60.81%	48.25%	368.34%	-9.85%	0.00%	-88.40%	-87.58%	-88.38%	-79.84%	461.98%	0.47%
OS	ModelSE	14.46	39.32	40.13	47.37	30.55	4.80	14.39	14.46	39.32	40.13	47.37	30.55	4.80	14.39
OS	Cover2	0.93	0.81	0.90	1.00	0.74	0.00	0.98	0.95	0.82	0.92	1.00	0.75	0.00	0.99
OS	Cover1	1.00	1.00	1.00	1.00	0.99	0.26	1.00	1.00	1.00	1.00	1.00	0.99	0.26	1.00
QALYs	Mean	62.80	62.39	62.81	66.12	59.70	48.98	62.74	63.57	64.85	66.05	73.13	60.88	49.19	63.67
QALYs	Bias	-5.57	-5.99	-5.56	-2.25	-8.68	-19.40	-5.64	-4.81	-3.52	-2.33	4.75	-7.50	-19.19	-4.70
QALYs	MC SE	0.18	0.17	0.17	0.25	0.14	0.10	0.18	0.19	0.49	0.47	0.49	0.37	0.10	0.19
QALYs	Rel Bias	-0.08	-0.09	-0.08	-0.03	-0.13	-0.28	-0.08	-0.07	-0.05	-0.03	-0.07	-0.11	-0.28	-0.07
QALYs	empSE	5.50	5.10	5.26	7.82	4.44	3.00	5.64	5.94	15.06	14.55	15.03	11.45	3.05	5.90
QALYs	MC SE	0.13	0.12	0.12	0.18	0.10	0.07	0.13	0.14	0.35	0.33	0.35	0.26	0.07	0.14
QALYs	MSE	61.25	61.80	58.58	66.16	95.06	385.18	63.53	58.34	239.04	217.01	248.36	187.28	377.50	56.92
QALYs	MC SE	2.08	1.87	1.84	2.38	2.26	3.73	2.19	2.83	53.40	49.37	61.83	36.95	3.74	2.05
QALYs	Rel P	0.00%	16.23%	9.11%	-50.61%	52.91%	235.07%	-4.99%	0.00%	-84.45%	-83.35%	-84.39%	-73.11%	280.45%	1.22%
QALYs	ModelSE	7.70	20.12	20.52	24.16	16.01	3.44	8.32	7.70	20.12	20.52	24.16	16.01	3.44	8.32
QALYs	Cover2	0.81	0.77	0.85	0.99	0.67	0.00	0.88	0.82	0.79	0.88	0.99	0.67	0.00	0.89
QALYs	Cover1	0.98	0.99	1.00	1.00	0.98	0.25	1.00	0.98	0.99	1.00	1.00	0.98	0.25	1.00

Table 43 : Result for simulated scenario 21

	Scenario21	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.65	59.88	59.91	59.90	54.66	49.06	53.36	57.90	60.76	60.78	60.78	54.39	49.22	53.62
PFS	Bias	-10.42	-9.18	-9.16	-9.16	-14.40	-20.00	-15.70	-11.16	-8.31	-8.28	-8.28	-14.67	-19.85	-15.44
PFS	MC SE	0.19	0.20	0.20	0.22	0.16	0.12	0.17	0.17	0.21	0.21	0.23	0.17	0.12	0.18
PFS	Rel Bias	-15.1%	-13.3%	-13.3%	-13.3%	-20.9%	-29.0%	-22.7%	-16.2%	-12.0%	-12.0%	-12.0%	-21.2%	-28.7%	-22.4%
PFS	empSE	5.73	6.30	6.22	6.63	5.08	3.65	5.36	5.34	6.58	6.50	6.95	5.09	3.69	5.47
PFS	MC SE	0.13	0.14	0.14	0.15	0.12	0.08	0.12	0.12	0.15	0.15	0.16	0.12	0.08	0.13
PFS	MSE	141.27	124.04	122.49	127.91	233.24	413.45	275.29	153.04	112.26	110.76	116.87	241.16	407.51	268.32
PFS	MC SE	3.92	3.85	3.77	3.98	4.59	4.66	5.88	3.91	3.68	3.60	3.99	4.66	4.68	5.86
PFS	Rel P	0.00%	-17.48%	-15.27%	-25.43%	27.06%	146.47%	13.95%	0.00%	-34.14%	-32.48%	-41.00%	10.09%	109.27%	-4.50%
PFS	ModelSE	5.37	7.23	7.22	7.32	6.21	4.08	5.48	5.37	7.23	7.22	7.32	6.21	4.08	5.48
PFS	Cover2	0.47	0.63	0.64	0.63	0.29	0.01	0.23	0.43	0.66	0.66	0.66	0.27	0.01	0.24
PFS	Cover1	0.82	0.92	0.92	0.92	0.79	0.39	0.79	0.82	0.92	0.92	0.92	0.79	0.39	0.79
OS	Mean	72.48	73.27	73.56	77.55	69.27	59.43	74.72	73.46	74.86	76.88	84.36	70.01	59.62	75.70
OS	Bias	-9.45	-8.66	-8.37	-4.38	-12.66	-22.50	-7.21	-8.48	-7.07	-5.05	2.42	-11.92	-22.31	-6.23
OS	MC SE	0.22	0.19	0.19	0.32	0.15	0.12	0.23	0.23	0.40	0.64	0.44	0.32	0.12	0.24
OS	Rel Bias	-0.12	-0.11	-0.10	-0.05	-0.15	-0.27	-0.09	-0.10	-0.09	0.06	0.03	-0.15	-0.27	-0.08
OS	empSE	6.92	5.79	5.92	9.98	4.67	3.78	6.97	7.19	12.41	19.68	13.56	9.92	3.83	7.30
OS	MC SE	0.16	0.13	0.14	0.23	0.11	0.09	0.16	0.17	0.28	0.45	0.31	0.23	0.09	0.17
OS	MSE	137.13	108.55	105.12	118.67	182.02	520.38	100.56	123.50	203.89	412.22	189.62	240.37	512.49	91.99
OS	MC SE	0.16	0.13	0.14	0.23	0.11	0.09	0.16	3.96	68.34	138.14	17.30	52.12	5.49	2.82
OS	Rel P	0.00%	42.85%	36.70%	-51.91%	119.26%	234.28%	-1.48%	0.00%	-66.42%	-86.64%	-71.88%	-47.43%	252.32%	-2.85%
OS	ModelSE	8.96	17.22	27.69	24.55	14.71	4.30	9.41	8.96	17.22	27.69	24.55	14.71	4.30	9.41
OS	Cover2	0.72	0.69	0.75	0.95	0.49	0.00	0.82	0.75	0.72	0.80	0.97	0.50	0.00	0.84
OS	Cover1	0.97	0.99	0.99	1.00	0.95	0.31	0.99	0.97	0.99	0.99	1.00	0.95	0.31	0.99
QALYs	Mean	53.83	54.60	54.75	56.74	51.03	44.44	53.37	54.10	55.66	56.68	60.41	51.32	44.57	53.94
QALYs	Bias	-7.85	-7.09	-6.93	-4.94	-10.65	-17.25	-8.32	-7.59	-6.03	-5.01	-1.27	-10.36	-17.11	-7.75
QALYs	MC SE	0.16	0.15	0.15	0.20	0.11	0.10	0.15	0.16	0.23	0.34	0.26	0.18	0.10	0.16
QALYs	Rel Bias	-0.13	-0.11	-0.11	-0.08	-0.17	-0.28	-0.13	-0.12	-0.10	-0.08	-0.02	-0.17	-0.28	-0.13
QALYs	empSE	4.83	4.51	4.51	6.25	3.35	2.95	4.77	4.91	7.20	10.45	7.95	5.48	2.98	4.96
QALYs	MC SE	0.11	0.10	0.10	0.14	0.08	0.07	0.11	0.11	0.17	0.24	0.18	0.13	0.07	0.11
QALYs	MSE	84.91	70.51	68.37	63.44	124.62	306.21	91.86	81.62	88.06	134.20	64.77	137.39	301.64	84.56
QALYs	MC SE	2.49	1.93	1.87	2.05	2.17	3.26	2.52	2.44	16.65	33.26	4.13	12.08	3.27	2.43
QALYs	Rel P	0.00%	14.82%	14.79%	-40.33%	107.77%	168.56%	2.58%	0.00%	-53.45%	-77.93%	-61.87%	-19.82%	170.86%	-1.86%
QALYs	ModelSE	5.22	9.68	14.54	13.21	8.35	3.33	6.08	5.22	9.68	14.54	13.21	8.35	3.33	6.08
QALYs	Cover2	0.57	0.66	0.70	0.91	0.39	0.01	0.63	0.58	0.69	0.74	0.94	0.38	0.01	0.65
QALYs	Cover1	0.89	0.99	0.99	1.00	0.93	0.34	0.98	0.89	0.99	0.99	1.00	0.93	0.34	0.98

Table 44 : Result for simulated scenario 22

	Scenario22	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	55.14	56.92	56.91	56.91	50.06	42.91	49.01	55.50	57.78	57.77	57.77	49.71	43.09	49.17
PFS	Bias	-0.84	0.94	0.93	0.93	-5.92	-13.07	-6.97	-0.48	1.80	1.79	1.79	-6.28	-12.89	-6.81
PFS	MC SE	0.18	0.21	0.22	0.22	0.17	0.11	0.16	0.18	0.22	0.23	0.23	0.17	0.11	0.16
PFS	Rel Bias	-1.5%	1.7%	1.7%	1.7%	-10.6%	-23.3%	-12.5%	-0.9%	3.2%	3.2%	3.2%	-11.2%	-23.0%	-12.2%
PFS	empSE	5.47	6.53	6.67	6.67	5.10	3.27	4.79	5.48	6.78	6.96	6.96	5.19	3.33	4.89
PFS	MC SE	0.13	0.15	0.15	0.15	0.12	0.08	0.11	0.13	0.16	0.16	0.16	0.12	0.08	0.11
PFS	MSE	30.60	43.49	45.24	45.36	60.96	181.44	71.57	30.25	49.19	51.51	51.66	66.29	177.19	70.24
PFS	MC SE	1.43	2.19	3.16	3.16	1.85	2.69	2.23	1.44	2.54	3.95	3.95	2.09	2.70	2.16
PFS	Rel P	0.00%	-29.82%	-32.63%	-32.80%	15.28%	179.54%	30.42%	0.00%	-34.69%	-37.88%	-38.05%	11.57%	170.39%	25.61%
PFS	ModelSE	6.00	6.72	6.81	6.81	5.91	3.35	5.32	6.00	6.72	6.81	6.81	5.91	3.35	5.32
PFS	Cover2	0.93	0.94	0.94	0.94	0.66	0.08	0.64	0.94	0.95	0.95	0.94	0.66	0.09	0.64
PFS	Cover1	1.00	1.00	1.00	1.00	0.95	0.57	0.95	1.00	1.00	1.00	1.00	0.95	0.57	0.95
OS	Mean	131.19	130.26	132.92	147.41	127.68	86.02	141.89	158.14	138.14	153.20	179.51	133.42	87.50	157.99
OS	Bias	2.17	1.24	3.90	18.39	-1.34	-42.99	12.88	29.12	9.12	24.19	50.49	4.40	-41.51	28.97
OS	MC SE	0.45	0.42	0.58	1.27	0.44	0.25	0.52	2.50	1.29	1.67	1.81	1.19	0.26	1.04
OS	Rel Bias	0.02	0.01	0.03	0.14	-0.01	-0.33	0.10	0.23	0.07	0.19	0.39	0.03	-0.32	0.22
OS	empSE	14.00	12.99	17.89	39.05	13.47	7.56	16.05	77.07	39.74	51.48	55.80	36.66	8.11	31.67
OS	MC SE	0.32	0.30	0.41	0.90	0.31	0.17	0.37	1.77	0.91	1.18	1.28	0.84	0.19	0.73
OS	MSE	200.59	170.17	334.94	1861.12	182.98	1905.64	423.08	6781.14	1660.67	3232.65	5659.80	1361.95	1789.00	1841.56
OS	MC SE	0.32	0.30	0.41	0.90	0.31	0.17	0.37	890.19	460.35	530.50	535.66	409.38	21.52	137.13
OS	Rel P	0.00%	16.15%	-38.74%	-87.14%	8.11%	242.68%	-23.87%	0.00%	276.10%	124.08%	90.75%	341.90%	8930.17%	491.97%
OS	ModelSE	91.01	40.66	65.48	88.60	40.61	10.16	69.72	91.01	40.66	65.48	88.60	40.61	10.16	69.72
OS	Cover2	1.00	0.99	1.00	1.00	1.00	0.03	1.00	1.00	1.00	1.00	1.00	1.00	0.05	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.64	1.00	1.00	1.00	1.00	1.00	1.00	0.64	1.00
QALYs	Mean	82.14	82.20	83.53	90.78	78.86	55.89	85.65	95.72	86.40	93.93	107.09	81.62	56.68	93.75
QALYs	Bias	0.83	0.90	2.23	9.47	-2.45	-25.42	4.35	14.42	5.10	12.63	25.78	0.32	-24.62	12.44
QALYs	MC SE	0.24	0.23	0.30	0.64	0.23	0.14	0.28	1.26	0.65	0.84	0.91	0.60	0.15	0.53
QALYs	Rel Bias	0.01	0.01	0.03	0.12	-0.03	-0.31	0.05	0.18	0.06	0.16	0.32	0.00	-0.30	0.15
QALYs	empSE	7.47	7.13	9.28	19.65	7.21	4.30	8.49	38.70	20.18	25.83	28.01	18.62	4.57	16.11
QALYs	MC SE	0.17	0.16	0.21	0.45	0.17	0.10	0.20	0.89	0.46	0.59	0.64	0.43	0.10	0.37
QALYs	MSE	56.50	51.55	91.03	475.39	57.93	664.49	90.88	1704.03	432.77	825.99	1448.65	346.48	627.12	414.22
QALYs	MC SE	2.22	1.92	4.72	37.82	1.93	7.05	4.03	222.89	116.73	132.76	134.49	101.76	7.24	32.46
QALYs	Rel P	0.00%	9.98%	-35.16%	-85.53%	7.43%	202.33%	-22.49%	0.00%	267.81%	124.48%	90.87%	331.95%	7074.54%	476.82%
QALYs	ModelSE	45.90	20.63	32.95	44.48	20.76	5.59	35.42	45.90	20.63	32.95	44.48	20.76	5.59	35.42
QALYs	Cover2	1.00	1.00	1.00	1.00	0.99	0.02	1.00	1.00	1.00	1.00	1.00	0.99	0.03	1.00
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.59	1.00	1.00	1.00	1.00	1.00	1.00	0.59	1.00

Table 45 : Result for simulated scenario 23

	Scenario23	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	52.55	56.81	56.77	56.83	49.62	44.03	47.71	52.56	57.67	57.64	57.70	49.41	44.21	47.91
PFS	Bias	-3.38	0.88	0.84	0.90	-6.31	-11.90	-8.22	-3.37	1.74	1.70	1.77	-6.52	-11.72	-8.03
PFS	MC SE	0.14	0.20	0.20	0.21	0.16	0.11	0.13	0.14	0.21	0.21	0.22	0.16	0.11	0.13
PFS	Rel Bias	-6.0%	1.6%	1.5%	1.6%	-11.3%	-21.3%	-14.7%	-6.0%	3.1%	3.0%	3.2%	-11.7%	-21.0%	-14.3%
PFS	empSE	4.27	6.27	6.31	6.60	5.01	3.43	4.04	4.17	6.52	6.56	6.90	5.08	3.49	4.09
PFS	MC SE	0.10	0.14	0.14	0.15	0.12	0.08	0.09	0.10	0.15	0.15	0.16	0.12	0.08	0.09
PFS	MSE	29.68	40.05	40.55	44.38	64.92	153.43	83.86	28.71	45.47	45.94	50.70	68.25	149.60	81.08
PFS	MC SE	1.24	1.87	1.92	3.03	1.94	2.57	2.16	1.19	2.17	2.23	3.84	2.07	2.57	2.14
PFS	Rel P	0.00%	-53.55%	-54.19%	-58.12%	-27.36%	55.00%	12.16%	0.00%	-59.08%	-59.63%	-63.47%	-32.51%	43.13%	4.17%
PFS	ModelSE	4.81	6.80	6.80	6.91	5.36	3.64	4.78	4.81	6.80	6.80	6.91	5.36	3.64	4.78
PFS	Cover2	0.86	0.94	0.94	0.94	0.64	0.15	0.54	0.86	0.95	0.95	0.95	0.63	0.16	0.55
PFS	Cover1	0.99	1.00	1.00	1.00	0.94	0.71	0.94	0.99	1.00	1.00	1.00	0.94	0.71	0.94
OS	Mean	79.46	82.89	83.83	89.32	77.98	63.44	83.02	81.58	86.82	91.95	105.16	79.51	63.80	85.15
OS	Bias	-2.47	0.96	1.90	7.40	-3.95	-18.49	1.09	-0.35	4.89	10.02	23.23	-2.42	-18.12	3.22
OS	MC SE	0.23	0.22	0.24	0.46	0.21	0.13	0.24	0.53	0.85	1.10	1.12	0.48	0.14	0.28
OS	Rel Bias	- 0.03	0.01	0.02	0.09	- 0.05	- 0.23	0.01	- 0.00	0.06	0.12	0.28	- 0.03	- 0.22	0.04
OS	empSE	7.01	6.88	7.44	14.16	6.56	4.08	7.43	16.41	26.33	33.77	34.66	14.85	4.17	8.49
OS	MC SE	0.16	0.16	0.17	0.32	0.15	0.09	0.17	0.38	0.60	0.78	0.80	0.34	0.10	0.19
OS	MSE	55.13	48.18	58.90	254.94	58.65	358.51	56.41	269.10	716.59	1239.34	1739.63	226.07	345.83	82.40
OS	MC SE	0.16	0.16	0.17	0.32	0.15	0.09	0.17	128.40	230.44	290.91	271.91	95.36	4.85	6.37
OS	Rel P	0.00%	3.75%	-11.31%	-75.51%	13.91%	195.12%	-11.19%	0.00%	-61.17%	-76.38%	-77.58%	22.13%	1451.84%	273.41%
OS	ModelSE	22.44	30.14	41.03	49.73	19.77	4.82	15.94	22.44	30.14	41.03	49.73	19.77	4.82	15.94
OS	Cover2	0.98	0.97	1.00	1.00	0.90	0.06	1.00	0.98	0.98	1.00	1.00	0.91	0.07	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	1.00	1.00	0.60	1.00
QALYs	Mean	55.49	58.48	58.95	61.71	53.87	44.93	55.82	56.56	60.71	63.26	69.89	54.58	45.16	56.94
QALYs	Bias	-2.25	0.74	1.20	3.97	-3.87	-12.82	-1.92	-1.18	2.97	5.52	12.15	-3.16	-12.58	-0.80
QALYs	MC SE	0.13	0.16	0.16	0.25	0.13	0.09	0.14	0.28	0.44	0.56	0.58	0.25	0.10	0.16
QALYs	Rel Bias	- 0.04	0.01	0.02	0.07	- 0.07	- 0.22	- 0.03	- 0.02	0.05	0.10	0.21	- 0.05	- 0.22	- 0.01
QALYs	empSE	4.14	4.84	4.89	7.79	4.09	2.90	4.34	8.55	13.65	17.27	17.78	7.83	2.96	4.81
QALYs	MC SE	0.09	0.11	0.11	0.18	0.09	0.07	0.10	0.20	0.31	0.40	0.41	0.18	0.07	0.11
QALYs	MSE	22.15	23.94	25.32	76.39	31.66	172.67	22.47	74.44	194.82	328.43	463.41	71.32	167.00	23.77
QALYs	MC SE	0.79	0.95	1.04	5.07	1.04	2.38	0.83	31.60	58.09	73.88	69.39	23.31	2.38	1.35
QALYs	Rel P	0.00%	-26.90%	-28.40%	-71.81%	2.35%	102.83%	-8.98%	0.00%	-60.73%	-75.48%	-76.87%	19.14%	734.00%	215.72%
QALYs	ModelSE	11.60	15.59	20.91	25.25	10.47	3.34	8.83	11.60	15.59	20.91	25.25	10.47	3.34	8.83
QALYs	Cover2	0.96	0.97	1.00	1.00	0.86	0.07	0.98	0.97	0.98	1.00	1.00	0.86	0.07	0.98
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	1.00	1.00	0.60	1.00

Table 46 : Result for simulated scenario 24

	Scenario24	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	52.15	55.66	55.55	55.50	48.80	44.95	47.83	51.78	56.55	56.45	56.39	48.68	45.09	48.26
PFS	Bias	-3.78	-0.26	-0.37	-0.43	-7.13	-10.98	-8.10	-4.15	0.63	0.52	0.46	-7.24	-10.83	-7.66
PFS	MC SE	0.14	0.18	0.17	0.18	0.14	0.11	0.12	0.13	0.18	0.18	0.19	0.14	0.11	0.12
PFS	Rel Bias	-6.8%	-0.5%	-0.7%	-0.8%	-12.7%	-19.6%	-14.5%	-7.4%	1.1%	0.9%	0.8%	-12.9%	-19.4%	-13.7%
PFS	empSE	4.36	5.41	5.38	5.55	4.39	3.33	3.58	4.09	5.69	5.66	5.83	4.38	3.37	3.66
PFS	MC SE	0.10	0.12	0.12	0.13	0.10	0.08	0.08	0.09	0.13	0.13	0.13	0.10	0.08	0.08
PFS	MSE	33.24	29.27	29.08	30.96	70.00	131.57	78.41	33.95	32.76	32.24	34.19	71.63	128.62	72.06
PFS	MC SE	1.11	1.03	1.12	1.20	1.98	2.31	1.93	1.13	1.18	1.24	1.32	2.01	2.31	1.87
PFS	Rel P	0.00%	-35.01%	-34.43%	-38.33%	-1.26%	71.60%	47.94%	0.00%	-48.25%	-47.60%	-50.71%	-12.80%	47.96%	25.42%
PFS	ModelSE	4.52	6.76	6.74	6.75	4.87	3.73	4.66	4.52	6.76	6.74	6.75	4.87	3.73	4.66
PFS	Cover2	0.79	0.94	0.94	0.93	0.58	0.21	0.54	0.77	0.95	0.95	0.94	0.57	0.22	0.59
PFS	Cover1	0.99	1.00	1.00	1.00	0.94	0.76	0.96	0.99	1.00	1.00	1.00	0.94	0.76	0.96
OS	Mean	64.78	66.84	67.10	69.88	61.14	54.21	66.49	65.45	68.17	70.19	76.19	62.04	54.39	67.69
OS	Bias	-2.82	-0.77	-0.51	2.27	-6.46	-13.40	-1.11	-2.15	0.56	2.58	8.58	-5.56	-13.22	0.08
OS	MC SE	0.20	0.18	0.18	0.26	0.14	0.11	0.21	0.21	0.31	0.66	0.43	0.28	0.12	0.21
OS	Rel Bias	-0.04	-0.01	0.01	0.03	-0.10	-0.20	0.02	-0.03	0.01	0.04	0.13	-0.08	0.20	0.00
OS	empSE	6.04	5.45	5.49	8.14	4.20	3.50	6.36	6.56	9.56	20.25	13.14	8.56	3.55	6.58
OS	MC SE	0.14	0.13	0.13	0.19	0.10	0.08	0.15	0.15	0.22	0.46	0.30	0.20	0.08	0.15
OS	MSE	44.39	30.27	30.35	71.34	59.37	191.72	41.64	47.66	91.62	416.36	246.29	104.06	187.26	43.28
OS	MC SE	0.14	0.13	0.13	0.19	0.10	0.08	0.15	2.70	44.11	168.52	34.39	40.81	3.01	1.56
OS	Rel P	0.00%	22.76%	21.09%	-44.94%	107.04%	197.40%	-9.82%	0.00%	-52.88%	-89.50%	-75.07%	-41.17%	242.25%	-0.59%
OS	ModelSE	8.27	13.89	24.40	24.03	12.95	3.97	7.98	8.27	13.89	24.40	24.03	12.95	3.97	7.98
OS	Cover2	0.89	0.92	0.95	0.99	0.73	0.12	0.96	0.90	0.93	0.97	1.00	0.73	0.14	0.98
OS	Cover1	1.00	1.00	1.00	1.00	0.99	0.68	1.00	1.00	1.00	1.00	1.00	0.99	0.68	1.00
QALYs	Mean	48.04	50.12	50.21	51.59	45.21	40.59	47.59	48.26	51.05	52.03	55.01	45.63	40.72	48.32
QALYs	Bias	-2.54	-0.46	-0.37	1.01	-5.37	-9.99	-2.99	-2.32	0.47	1.45	4.43	-4.95	-9.86	-2.26
QALYs	MC SE	0.13	0.14	0.13	0.17	0.10	0.09	0.13	0.14	0.19	0.34	0.24	0.16	0.09	0.13
QALYs	Rel Bias	-0.05	-0.01	0.01	0.02	-0.11	-0.20	0.06	-0.05	0.01	0.03	0.09	-0.10	-0.19	0.04
QALYs	empSE	4.05	4.22	4.14	5.18	3.11	2.71	3.88	4.26	5.82	10.62	7.40	4.79	2.74	4.00
QALYs	MC SE	0.09	0.10	0.10	0.12	0.07	0.06	0.09	0.10	0.13	0.24	0.17	0.11	0.06	0.09
QALYs	MSE	22.84	18.01	17.25	27.81	38.52	107.15	23.94	23.47	34.03	114.73	74.27	47.49	104.68	21.11
QALYs	MC SE	0.72	0.60	0.59	1.13	1.00	1.73	0.80	0.86	10.77	42.15	8.69	9.84	1.73	0.72
QALYs	Rel P	0.00%	-7.97%	-4.32%	-38.89%	68.99%	123.88%	9.02%	0.00%	-46.49%	-83.94%	-66.89%	-21.21%	141.17%	13.00%
QALYs	ModelSE	4.69	8.06	12.87	12.77	7.16	3.07	5.16	4.69	8.06	12.87	12.77	7.16	3.07	5.16
QALYs	Cover2	0.82	0.93	0.96	0.99	0.69	0.15	0.89	0.83	0.95	0.97	1.00	0.69	0.16	0.91
QALYs	Cover1	0.99	1.00	1.00	1.00	1.00	0.71	1.00	0.99	1.00	1.00	1.00	1.00	0.71	1.00

Table 47 : Result for simulated scenario 25

	Scenario25	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	60.41	61.24	61.21	61.10	57.48	49.41	55.49	60.75	61.85	61.82	61.71	56.97	49.55	55.71
PFS	Bias	-8.69	-7.86	-7.89	-8.00	-11.62	-19.69	-13.61	-8.35	-7.25	-7.28	-7.39	-12.13	-19.55	-13.39
PFS	MC SE	0.17	0.20	0.20	0.20	0.17	0.11	0.14	0.17	0.21	0.20	0.21	0.20	0.11	0.14
PFS	Rel Bias	-12.6%	-11.4%	-11.4%	-11.6%	-16.8%	-28.5%	-19.7%	-12.1%	-10.5%	-10.5%	-10.7%	-17.6%	-28.3%	-19.4%
PFS	empSE	5.35	6.08	6.06	6.09	5.31	3.24	4.32	5.36	6.33	6.31	6.35	6.03	3.27	4.37
PFS	MC SE	0.12	0.14	0.14	0.14	0.12	0.07	0.10	0.12	0.15	0.14	0.15	0.14	0.08	0.10
PFS	MSE	104.15	98.71	98.91	101.04	163.26	398.04	203.90	98.40	92.68	92.83	94.95	183.46	392.75	198.38
PFS	MC SE	2.89	3.01	3.03	3.07	3.59	4.06	3.68	2.80	3.01	3.03	3.07	4.64	4.07	3.65
PFS	Rel P	0.00%	-22.59%	-22.07%	-22.93%	1.60%	171.98%	53.34%	0.00%	-28.34%	-27.86%	-28.71%	-20.88%	168.24%	50.48%
PFS	ModelSE	5.26	5.86	5.85	5.84	7.16	3.22	4.62	5.26	5.86	5.85	5.84	7.16	3.22	4.62
PFS	Cover2	0.54	0.60	0.60	0.59	0.39	0.00	0.19	0.56	0.63	0.63	0.62	0.40	0.00	0.20
PFS	Cover1	0.90	0.92	0.92	0.92	0.80	0.14	0.75	0.90	0.92	0.92	0.92	0.80	0.14	0.75
OS	Mean	148.31	143.20	146.11	155.08	140.88	106.55	153.00	150.79	144.37	154.67	172.54	140.65	107.29	157.27
OS	Bias	0.36	-4.75	-1.84	7.13	-7.07	-41.40	5.05	2.84	-3.58	6.72	24.59	-7.30	-40.66	9.32
OS	MC SE	0.43	0.39	0.45	0.76	0.39	0.22	0.44	0.47	0.46	1.30	1.37	0.50	0.23	0.46
OS	Rel Bias	0.00	-0.03	0.01	0.05	-0.05	0.28	0.03	0.02	-0.02	0.05	0.17	-0.05	0.27	0.06
OS	empSE	13.35	12.00	13.82	23.37	12.03	6.81	13.47	14.55	14.13	40.18	42.22	15.37	6.95	14.06
OS	MC SE	0.31	0.28	0.32	0.54	0.28	0.16	0.31	0.33	0.32	0.92	0.97	0.35	0.16	0.32
OS	MSE	178.22	166.35	194.08	596.37	194.62	1760.32	206.75	219.58	212.31	1657.67	2385.34	289.19	1701.82	284.36
OS	MC SE	0.31	0.28	0.32	0.54	0.28	0.16	0.31	13.48	17.49	425.91	366.91	18.19	18.30	11.30
OS	Rel P	0.00%	23.81%	-6.61%	-67.36%	23.10%	284.65%	-1.77%	0.00%	6.03%	-86.88%	-88.12%	-10.36%	338.22%	7.12%
OS	ModelSE	24.93	19.65	43.88	61.36	23.17	8.29	28.08	24.93	19.65	43.88	61.36	23.17	8.29	28.08
OS	Cover2	1.00	0.92	0.99	1.00	0.92	0.00	1.00	1.00	0.93	1.00	1.00	0.93	0.00	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.39	1.00	1.00	1.00	1.00	1.00	1.00	0.39	1.00
QALYs	Mean	92.28	89.97	91.42	95.87	87.68	68.10	93.14	93.62	90.74	95.88	104.78	87.42	68.51	95.35
QALYs	Bias	-2.43	-4.73	-3.29	1.16	-7.02	-26.61	-1.56	-1.08	-3.97	1.17	10.07	-7.29	-26.20	0.64
QALYs	MC SE	0.23	0.21	0.23	0.39	0.21	0.13	0.23	0.25	0.25	0.66	0.69	0.27	0.13	0.24
QALYs	Rel Bias	-0.03	-0.05	0.03	0.01	-0.07	-0.28	-0.02	-0.01	-0.04	0.01	0.11	-0.08	0.28	0.01
QALYs	empSE	6.95	6.60	7.23	11.91	6.47	3.94	7.02	7.56	7.69	20.33	21.30	8.36	4.02	7.33
QALYs	MC SE	0.16	0.15	0.17	0.27	0.15	0.09	0.16	0.17	0.18	0.47	0.49	0.19	0.09	0.17
QALYs	MSE	54.15	65.96	62.97	143.00	91.15	723.42	51.64	58.26	74.76	414.44	554.71	122.93	702.34	54.11
QALYs	MC SE	1.87	2.39	2.23	7.44	3.03	6.80	1.85	3.22	4.60	105.86	88.96	6.15	6.82	2.00
QALYs	Rel P	0.00%	10.72%	-7.51%	-65.94%	15.29%	210.61%	-1.94%	0.00%	-3.28%	-86.18%	-87.40%	-18.24%	254.12%	6.31%
QALYs	ModelSE	12.71	10.40	22.22	30.92	12.86	4.69	14.68	12.71	10.40	22.22	30.92	12.86	4.69	14.68
QALYs	Cover2	0.99	0.88	0.97	1.00	0.84	0.00	1.00	1.00	0.90	0.99	1.00	0.85	0.00	1.00
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.26	1.00	1.00	1.00	1.00	1.00	1.00	0.26	1.00

Table 48 : Result for simulated scenario 26

	Scenario26	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	60.14	61.45	61.44	61.26	57.30	50.15	54.59	60.39	62.06	62.05	61.87	56.90	50.27	54.79
PFS	Bias	-8.93	-7.63	-7.63	-7.81	-11.78	-18.93	-14.49	-8.69	-7.02	-7.02	-7.21	-12.17	-18.81	-14.29
PFS	MC SE	0.17	0.19	0.19	0.19	0.17	0.10	0.13	0.17	0.20	0.20	0.20	0.18	0.10	0.13
PFS	Rel Bias	-12.9%	-11.0%	-11.0%	-11.3%	-17.0%	-27.4%	-21.0%	-12.6%	-10.2%	-10.2%	-10.4%	-17.6%	-27.2%	-20.7%
PFS	empSE	5.19	5.91	5.91	5.90	5.16	3.10	4.06	5.13	6.14	6.14	6.12	5.47	3.13	4.09
PFS	MC SE	0.12	0.14	0.14	0.14	0.12	0.07	0.09	0.12	0.14	0.14	0.14	0.13	0.07	0.09
PFS	MSE	106.70	93.12	93.20	95.83	165.29	367.79	226.44	101.77	86.95	87.00	89.43	178.07	363.52	220.83
PFS	MC SE	2.89	2.83	2.84	2.91	3.54	3.77	3.76	2.80	2.75	2.77	2.83	4.08	3.77	3.75
PFS	Rel P	0.00%	-23.03%	-23.08%	-22.69%	0.88%	180.24%	62.88%	0.00%	-30.29%	-30.31%	-29.89%	-12.00%	169.13%	57.02%
PFS	ModelSE	4.94	5.82	5.82	5.77	6.97	3.28	4.30	4.94	5.82	5.82	5.77	6.97	3.28	4.30
PFS	Cover2	0.50	0.63	0.62	0.61	0.37	0.00	0.12	0.51	0.65	0.65	0.63	0.37	0.00	0.12
PFS	Cover1	0.87	0.93	0.93	0.93	0.78	0.15	0.68	0.87	0.93	0.93	0.93	0.78	0.15	0.68
OS	Mean	110.08	105.16	106.91	112.60	102.40	85.38	111.86	111.21	106.96	110.70	120.93	103.53	85.70	113.47
OS	Bias	-2.02	-6.94	-5.19	0.49	-9.71	-26.73	-0.24	-0.90	-5.15	-1.40	8.82	-8.58	-26.40	1.37
OS	MC SE	0.32	0.25	0.29	0.45	0.25	0.14	0.31	0.33	0.62	0.82	0.68	0.63	0.14	0.32
OS	Rel Bias	-0.02	-0.06	0.05	0.00	-0.09	-0.24	0.00	-0.01	-0.05	0.01	0.08	0.08	0.24	0.01
OS	empSE	9.82	7.73	8.95	13.89	7.86	4.17	9.70	10.24	19.11	25.14	20.86	19.55	4.22	9.92
OS	MC SE	0.23	0.18	0.21	0.32	0.18	0.10	0.22	0.24	0.44	0.58	0.48	0.45	0.10	0.23
OS	MSE	100.39	107.88	107.08	193.06	155.90	731.68	94.11	105.61	391.32	633.40	512.57	455.28	714.71	100.12
OS	MC SE	0.23	0.18	0.21	0.32	0.18	0.10	0.22	3.46	126.10	193.16	85.90	120.08	7.19	3.76
OS	Rel P	0.00%	61.34%	20.24%	-50.05%	56.20%	453.38%	2.40%	0.00%	-71.27%	-83.40%	-75.89%	-72.54%	488.87%	6.67%
OS	ModelSE	12.69	25.58	31.42	34.81	27.49	5.12	13.39	12.69	25.58	31.42	34.81	27.49	5.12	13.39
OS	Cover2	0.95	0.80	0.87	1.00	0.70	0.00	0.98	0.95	0.80	0.89	1.00	0.71	0.00	0.99
OS	Cover1	1.00	0.99	1.00	1.00	0.98	0.25	1.00	1.00	0.99	1.00	1.00	0.98	0.25	1.00
QALYs	Mean	73.08	71.01	71.89	74.68	68.39	57.73	72.31	73.72	72.10	73.97	79.02	68.83	57.93	73.17
QALYs	Bias	-3.69	-5.76	-4.89	-2.10	-8.39	-19.04	-4.47	-3.05	-4.68	-2.81	2.25	-7.94	-18.84	-3.60
QALYs	MC SE	0.18	0.16	0.17	0.24	0.15	0.09	0.18	0.19	0.33	0.42	0.35	0.33	0.09	0.18
QALYs	Rel Bias	-0.05	-0.08	-0.06	-0.03	-0.11	-0.25	-0.06	-0.04	-0.06	0.04	0.03	-0.10	-0.25	-0.05
QALYs	empSE	5.57	4.94	5.28	7.34	4.64	2.77	5.41	5.80	10.18	13.00	10.84	10.23	2.80	5.53
QALYs	MC SE	0.13	0.11	0.12	0.17	0.11	0.06	0.12	0.13	0.23	0.30	0.25	0.23	0.06	0.13
QALYs	MSE	44.66	57.54	51.75	58.15	91.81	370.22	49.20	42.96	125.40	176.68	122.46	167.51	362.85	43.48
QALYs	MC SE	1.59	1.81	1.74	2.05	2.31	3.40	1.74	1.51	31.34	47.42	20.63	28.92	3.40	1.56
QALYs	Rel P	0.00%	27.40%	11.38%	-42.26%	44.53%	304.92%	6.11%	0.00%	-67.51%	-80.07%	-71.35%	-67.80%	329.86%	10.26%
QALYs	ModelSE	6.65	13.26	16.10	17.81	14.61	3.32	7.50	6.65	13.26	16.10	17.81	14.61	3.32	7.50
QALYs	Cover2	0.87	0.75	0.82	0.99	0.58	0.00	0.89	0.88	0.76	0.86	1.00	0.60	0.00	0.91
QALYs	Cover1	0.99	0.98	1.00	1.00	0.95	0.16	1.00	0.99	0.98	1.00	1.00	0.95	0.16	1.00

Table 49 : Result for simulated scenario 27

	Scenario27	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	60.19	61.01	61.03	61.13	56.27	50.38	54.51	60.04	61.60	61.62	61.73	56.04	50.47	54.56
PFS	Bias	-8.87	-8.05	-8.03	-7.93	-12.79	-18.68	-14.56	-9.03	-7.47	-7.44	-7.34	-13.03	-18.59	-14.51
PFS	MC SE	0.17	0.18	0.18	0.19	0.16	0.10	0.13	0.16	0.19	0.19	0.20	0.16	0.10	0.13
PFS	Rel Bias	-12.8%	-11.7%	-11.6%	-11.5%	-18.5%	-27.0%	-21.1%	-13.1%	-10.8%	-10.8%	-10.6%	-18.9%	-26.9%	-21.0%
PFS	empSE	5.11	5.60	5.63	5.84	4.78	3.03	4.14	4.90	5.78	5.83	6.04	4.93	3.05	4.15
PFS	MC SE	0.12	0.13	0.13	0.13	0.11	0.07	0.10	0.11	0.13	0.13	0.14	0.11	0.07	0.10
PFS	MSE	104.77	96.14	96.25	96.98	186.55	358.15	229.08	105.53	89.09	89.35	90.25	193.96	354.89	227.73
PFS	MC SE	2.93	2.98	2.95	2.98	3.73	3.66	3.88	2.89	2.88	2.86	2.90	3.94	3.66	3.87
PFS	Rel P	0.00%	-16.73%	-17.80%	-23.41%	14.10%	184.79%	52.13%	0.00%	-28.04%	-29.30%	-34.16%	-1.25%	158.57%	39.19%
PFS	ModelSE	4.65	5.59	5.63	5.67	5.54	3.35	4.29	4.65	5.59	5.63	5.67	5.54	3.35	4.29
PFS	Cover2	0.48	0.61	0.61	0.61	0.27	0.00	0.13	0.47	0.64	0.64	0.64	0.25	0.00	0.13
PFS	Cover1	0.85	0.92	0.92	0.92	0.71	0.19	0.67	0.85	0.92	0.92	0.92	0.71	0.19	0.67
OS	Mean	83.50	82.03	82.31	85.06	78.28	68.41	85.38	84.21	82.56	83.57	89.30	78.70	68.55	86.02
OS	Bias	-6.65	-8.12	-7.83	-5.08	-11.87	-21.74	-4.76	-5.94	-7.59	-6.57	-0.84	-11.45	-21.60	-4.12
OS	MC SE	0.22	0.18	0.19	0.28	0.17	0.11	0.22	0.22	0.19	0.26	0.37	0.20	0.11	0.23
OS	Rel Bias	- 0.07	- 0.09	- 0.09	- 0.06	- 0.13	- 0.24	- 0.05	- 0.07	- 0.08	- 0.07	- 0.01	- 0.13	- 0.24	- 0.05
OS	empSE	6.65	5.65	5.92	8.69	5.19	3.31	6.82	6.85	5.99	8.06	11.33	6.15	3.33	7.00
OS	MC SE	0.15	0.13	0.14	0.20	0.12	0.08	0.16	0.16	0.14	0.19	0.26	0.14	0.08	0.16
OS	MSE	88.37	97.87	96.39	101.24	167.81	483.36	69.17	82.13	93.45	108.15	128.88	168.94	477.49	65.95
OS	MC SE	0.15	0.13	0.14	0.20	0.12	0.08	0.16	2.82	2.82	12.00	17.21	4.30	4.64	2.26
OS	Rel P	0.00%	38.13%	25.97%	-41.48%	63.98%	304.01%	-5.08%	0.00%	30.58%	-27.82%	-63.43%	23.96%	323.02%	-4.24%
OS	ModelSE	7.78	6.67	13.78	18.42	8.28	3.81	8.13	7.78	6.67	13.78	18.42	8.28	3.81	8.13
OS	Cover2	0.77	0.63	0.68	0.90	0.43	0.00	0.86	0.80	0.64	0.72	0.94	0.44	0.00	0.87
OS	Cover1	0.99	0.96	0.99	1.00	0.87	0.16	0.99	0.99	0.96	0.99	1.00	0.87	0.16	0.99
QALYs	Mean	59.81	59.32	59.47	60.87	56.02	49.32	59.04	60.12	59.76	60.27	63.17	56.16	49.42	59.38
QALYs	Bias	-5.99	-6.48	-6.33	-4.92	-9.77	-16.47	-6.75	-5.68	-6.03	-5.52	-2.62	-9.63	-16.38	-6.41
QALYs	MC SE	0.14	0.14	0.14	0.18	0.11	0.08	0.14	0.15	0.14	0.17	0.22	0.12	0.08	0.14
QALYs	Rel Bias	- 0.09	- 0.10	- 0.10	- 0.07	- 0.15	- 0.25	- 0.10	- 0.09	- 0.09	- 0.08	- 0.04	- 0.15	- 0.25	- 0.10
QALYs	empSE	4.43	4.20	4.31	5.42	3.47	2.48	4.34	4.51	4.37	5.15	6.64	3.84	2.50	4.43
QALYs	MC SE	0.10	0.10	0.10	0.12	0.08	0.06	0.10	0.10	0.10	0.12	0.15	0.09	0.06	0.10
QALYs	MSE	55.47	59.54	58.55	53.54	107.55	277.47	64.40	52.57	55.49	56.93	50.90	107.53	274.38	60.79
QALYs	MC SE	1.78	1.71	1.73	1.60	2.03	2.64	1.91	1.72	1.66	3.00	4.25	2.38	2.64	1.86
QALYs	Rel P	0.00%	11.41%	5.93%	-33.08%	63.16%	219.92%	4.13%	0.00%	6.53%	-23.14%	-53.80%	38.07%	226.98%	3.51%
QALYs	ModelSE	4.35	4.78	7.71	9.92	5.48	2.84	5.09	4.35	4.78	7.71	9.92	5.48	2.84	5.09
QALYs	Cover2	0.61	0.61	0.66	0.84	0.38	0.00	0.65	0.63	0.64	0.69	0.89	0.38	0.00	0.66
QALYs	Cover1	0.88	0.96	0.98	1.00	0.85	0.16	0.97	0.88	0.96	0.98	1.00	0.85	0.16	0.97

Table 50 : Result for simulated scenario 28

	Scenario28	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	55.79	57.20	57.12	57.08	50.83	44.68	49.44	56.11	57.83	57.74	57.70	50.61	44.79	49.49
PFS	Bias	-0.23	1.18	1.10	1.06	-5.19	-11.34	-6.58	0.09	1.81	1.72	1.68	-5.41	-11.23	-6.53
PFS	MC SE	0.16	0.18	0.18	0.18	0.14	0.09	0.12	0.16	0.18	0.19	0.19	0.14	0.09	0.12
PFS	Rel Bias	-0.4%	2.1%	2.0%	1.9%	-9.3%	-20.2%	-11.7%	0.2%	3.2%	3.1%	3.0%	-9.7%	-20.1%	-11.7%
PFS	empSE	4.80	5.53	5.58	5.62	4.31	2.85	3.83	4.84	5.70	5.75	5.79	4.29	2.87	3.83
PFS	MC SE	0.11	0.13	0.13	0.13	0.10	0.07	0.09	0.11	0.13	0.13	0.13	0.10	0.07	0.09
PFS	MSE	23.11	31.99	32.35	32.66	45.49	136.80	57.98	23.40	35.73	36.03	36.30	47.67	134.40	57.30
PFS	MC SE	0.98	1.53	1.52	1.53	1.42	2.07	1.67	1.01	1.73	1.72	1.73	1.50	2.06	1.65
PFS	Rel P	0.00%	-24.62%	-25.99%	-26.93%	24.48%	184.21%	56.98%	0.00%	-27.91%	-29.26%	-30.16%	27.50%	185.08%	59.39%
PFS	ModelSE	4.78	5.30	5.28	5.28	4.71	2.86	3.94	4.78	5.30	5.28	5.28	4.71	2.86	3.94
PFS	Cover2	0.92	0.93	0.93	0.93	0.63	0.06	0.55	0.92	0.92	0.92	0.92	0.62	0.06	0.56
PFS	Cover1	1.00	1.00	1.00	1.00	0.94	0.51	0.94	1.00	1.00	1.00	1.00	0.94	0.51	0.94
OS	Mean	147.44	148.53	150.43	158.44	146.22	105.70	153.07	158.72	155.27	164.97	181.97	152.22	106.87	161.78
OS	Bias	-0.05	1.05	2.95	10.96	-1.26	-41.78	5.59	11.24	7.79	17.48	34.48	4.73	-40.61	14.30
OS	MC SE	0.49	0.44	0.53	0.88	0.45	0.27	0.52	1.50	1.25	1.58	1.81	1.32	0.28	0.72
OS	Rel Bias	- 0.00	0.01	0.02	0.07	- 0.01	- 0.28	0.04	0.08	0.05	0.12	0.23	0.03	- 0.28	0.10
OS	empSE	15.03	13.51	16.45	27.02	13.85	8.24	15.99	46.26	38.66	48.73	55.85	40.67	8.56	22.01
OS	MC SE	0.35	0.31	0.38	0.62	0.32	0.19	0.37	1.06	0.89	1.12	1.28	0.93	0.20	0.51
OS	MSE	225.76	183.34	279.08	849.70	193.18	1813.30	286.57	2263.88	1553.30	2677.61	4304.50	1674.51	1722.73	688.24
OS	MC SE	0.35	0.31	0.38	0.62	0.32	0.19	0.37	499.18	411.71	491.94	558.47	423.32	22.26	50.11
OS	Rel P	0.00%	23.88%	-16.50%	-69.05%	17.84%	232.72%	-11.59%	0.00%	43.21%	-9.88%	-31.39%	29.39%	2820.61%	341.84%
OS	ModelSE	56.88	37.95	53.43	70.91	41.04	10.46	46.24	56.88	37.95	53.43	70.91	41.04	10.46	46.24
OS	Cover2	1.00	0.99	1.00	1.00	0.99	0.05	1.00	1.00	0.99	1.00	1.00	0.99	0.06	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.64	1.00	1.00	1.00	1.00	1.00	1.00	0.64	1.00
QALYs	Mean	90.46	91.43	92.35	96.35	88.36	66.26	91.37	96.19	94.98	99.81	108.29	91.29	66.87	95.74
QALYs	Bias	-0.09	0.88	1.80	5.80	-2.19	-24.29	0.82	5.65	4.44	9.26	17.74	0.74	-23.68	5.19
QALYs	MC SE	0.25	0.23	0.28	0.45	0.23	0.15	0.27	0.76	0.63	0.79	0.91	0.66	0.15	0.37
QALYs	Rel Bias	- 0.00	0.01	0.02	0.06	- 0.02	- 0.27	0.01	0.06	0.05	0.10	0.20	0.01	- 0.26	0.06
QALYs	empSE	7.80	7.23	8.55	13.75	7.23	4.49	8.31	23.35	19.50	24.45	28.08	20.46	4.65	11.24
QALYs	MC SE	0.18	0.17	0.20	0.32	0.17	0.10	0.19	0.54	0.45	0.56	0.64	0.47	0.11	0.26
QALYs	MSE	60.77	53.06	76.30	222.58	57.07	610.28	69.69	576.67	399.50	682.83	1102.50	418.76	582.18	153.14
QALYs	MC SE	2.10	1.97	3.24	14.52	1.90	7.03	2.46	126.12	102.95	123.20	140.51	103.79	7.08	11.58
QALYs	Rel P	0.00%	16.20%	-16.82%	-67.85%	16.22%	201.47%	-11.97%	0.00%	43.43%	-8.76%	-30.83%	30.27%	2426.91%	331.70%
QALYs	ModelSE	28.74	19.16	26.86	35.58	20.82	5.62	23.53	28.74	19.16	26.86	35.58	20.82	5.62	23.53
QALYs	Cover2	1.00	0.99	1.00	1.00	0.98	0.02	1.00	1.00	1.00	1.00	1.00	0.98	0.03	1.00
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.56	1.00	1.00	1.00	1.00	1.00	1.00	0.56	1.00

Table 51 : Result for simulated scenario 29

	Scenario29	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	53.55	57.11	57.11	57.05	50.48	45.26	48.12	53.69	57.73	57.73	57.67	50.36	45.36	48.20
PFS	Bias	-2.39	1.17	1.17	1.11	-5.46	-10.68	-7.82	-2.26	1.79	1.79	1.73	-5.58	-10.58	-7.74
PFS	MC SE	0.12	0.18	0.18	0.18	0.14	0.09	0.11	0.12	0.19	0.19	0.19	0.14	0.09	0.11
PFS	Rel Bias	-4.3%	2.1%	2.1%	2.0%	-9.8%	-19.1%	-14.0%	-4.0%	3.2%	3.2%	3.1%	-10.0%	-18.9%	-13.8%
PFS	empSE	3.82	5.57	5.57	5.62	4.29	2.83	3.35	3.81	5.74	5.75	5.79	4.28	2.86	3.37
PFS	MC SE	0.09	0.13	0.13	0.13	0.10	0.06	0.08	0.09	0.13	0.13	0.13	0.10	0.07	0.08
PFS	MSE	20.27	32.37	32.38	32.79	48.19	122.17	72.43	19.58	36.15	36.17	36.53	49.45	120.07	71.26
PFS	MC SE	0.84	1.53	1.53	1.55	1.46	1.94	1.69	0.82	1.74	1.74	1.76	1.50	1.94	1.68
PFS	Rel P	0.00%	-53.09%	-53.11%	-53.90%	-20.82%	81.75%	29.38%	0.00%	-56.03%	-56.07%	-56.80%	-20.81%	77.28%	27.74%
PFS	ModelSE	3.93	5.29	5.29	5.28	4.28	2.96	3.59	3.93	5.29	5.29	5.28	4.28	2.96	3.59
PFS	Cover2	0.86	0.93	0.93	0.92	0.60	0.08	0.40	0.86	0.92	0.92	0.91	0.59	0.09	0.41
PFS	Cover1	0.99	1.00	1.00	1.00	0.93	0.58	0.91	0.99	1.00	1.00	1.00	0.93	0.58	0.91
OS	Mean	91.34	95.80	96.20	99.69	91.37	76.23	92.87	92.36	98.64	101.28	108.51	92.57	76.57	94.32
OS	Bias	-3.55	0.91	1.30	4.79	-3.52	-18.66	-2.02	-2.53	3.75	6.39	13.62	-2.32	-18.32	-0.57
OS	MC SE	0.24	0.24	0.27	0.38	0.23	0.13	0.24	0.28	0.69	0.86	0.75	0.57	0.13	0.26
OS	Rel Bias	- 0.04	0.01	0.01	0.05	- 0.04	- 0.20	- 0.02	- 0.03	0.04	0.07	0.14	- 0.02	- 0.19	- 0.01
OS	empSE	7.25	7.29	8.35	11.76	7.12	4.10	7.40	8.69	21.39	26.36	23.15	17.53	4.15	8.09
OS	MC SE	0.17	0.17	0.19	0.27	0.16	0.09	0.17	0.20	0.49	0.61	0.53	0.40	0.10	0.19
OS	MSE	65.12	53.87	71.42	161.23	62.98	364.88	58.79	81.89	471.08	734.89	720.75	312.47	352.79	65.77
OS	MC SE	0.17	0.17	0.19	0.27	0.16	0.09	0.17	10.80	163.72	183.60	141.45	148.14	4.91	4.04
OS	Rel P	0.00%	-1.04%	-24.71%	-62.03%	3.75%	213.28%	-4.03%	0.00%	-83.48%	-89.12%	-85.90%	-75.42%	339.24%	15.34%
OS	ModelSE	13.21	26.00	32.66	33.78	19.02	4.94	12.58	13.21	26.00	32.66	33.78	19.02	4.94	12.58
OS	Cover2	0.96	0.97	0.97	1.00	0.88	0.05	0.99	0.97	0.97	0.98	1.00	0.88	0.06	0.99
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.59	1.00	1.00	1.00	1.00	1.00	1.00	0.59	1.00
QALYs	Mean	61.73	65.03	65.23	66.96	60.83	51.69	60.87	62.29	66.64	67.96	71.56	61.40	51.90	61.62
QALYs	Bias	-2.49	0.80	1.00	2.73	-3.40	-12.53	-3.36	-1.94	2.41	3.73	7.33	-2.83	-12.33	-2.61
QALYs	MC SE	0.13	0.15	0.17	0.21	0.14	0.09	0.14	0.16	0.36	0.44	0.39	0.30	0.09	0.15
QALYs	Rel Bias	- 0.04	0.01	0.02	0.04	- 0.05	- 0.20	- 0.05	- 0.03	0.04	0.06	0.11	- 0.04	- 0.19	- 0.04
QALYs	empSE	4.13	4.72	5.14	6.51	4.24	2.66	4.19	4.82	11.19	13.63	11.96	9.11	2.69	4.52
QALYs	MC SE	0.09	0.11	0.12	0.15	0.10	0.06	0.10	0.11	0.26	0.31	0.27	0.21	0.06	0.10
QALYs	MSE	23.27	22.86	27.35	49.75	29.51	164.20	28.83	26.97	130.90	199.46	196.69	90.95	159.36	27.22
QALYs	MC SE	0.78	0.90	1.01	2.53	1.01	2.16	0.98	2.61	41.80	47.06	35.54	36.64	2.14	1.15
QALYs	Rel P	0.00%	-23.26%	-35.31%	-59.71%	-5.13%	140.38%	-2.88%	0.00%	-81.46%	-87.50%	-83.78%	-72.03%	219.72%	13.64%
QALYs	ModelSE	6.96	13.35	16.61	17.20	9.92	3.12	6.92	6.96	13.35	16.61	17.20	9.92	3.12	6.92
QALYs	Cover2	0.94	0.97	0.98	1.00	0.82	0.03	0.94	0.94	0.97	0.98	1.00	0.81	0.03	0.95
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.51	1.00	1.00	1.00	1.00	1.00	1.00	0.51	1.00

Table 52 : Result for simulated scenario 30

	Scenario30	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	52.16	57.07	57.08	56.94	50.13	45.70	47.53	52.16	57.68	57.70	57.55	50.03	45.80	47.76
PFS	Bias	-3.77	1.14	1.15	1.01	-5.80	-10.22	-8.40	-3.76	1.76	1.77	1.62	-5.90	-10.13	-8.17
PFS	MC SE	0.12	0.17	0.17	0.18	0.14	0.09	0.10	0.12	0.18	0.18	0.18	0.13	0.09	0.10
PFS	Rel Bias	-6.7%	2.0%	2.1%	1.8%	-10.4%	-18.3%	-15.0%	-6.7%	3.1%	3.2%	2.9%	-10.5%	-18.1%	-14.6%
PFS	empSE	3.67	5.36	5.33	5.43	4.17	2.71	3.02	3.63	5.53	5.50	5.61	4.14	2.74	3.06
PFS	MC SE	0.08	0.12	0.12	0.12	0.10	0.06	0.07	0.08	0.13	0.13	0.13	0.10	0.06	0.07
PFS	MSE	27.65	30.02	29.71	30.52	50.99	111.88	79.63	27.32	33.64	33.33	34.04	51.89	110.00	76.10
PFS	MC SE	0.92	1.33	1.32	1.35	1.52	1.80	1.63	0.90	1.52	1.51	1.54	1.54	1.80	1.61
PFS	Rel P	0.00%	-53.27%	-52.72%	-54.50%	-22.72%	82.77%	47.06%	0.00%	-56.96%	-56.45%	-58.11%	-23.26%	75.79%	40.21%
PFS	ModelSE	3.54	5.24	5.25	5.23	4.05	3.01	3.43	3.54	5.24	5.25	5.23	4.05	3.01	3.43
PFS	Cover2	0.72	0.94	0.94	0.93	0.57	0.08	0.31	0.72	0.93	0.93	0.92	0.58	0.09	0.32
PFS	Cover1	0.95	1.00	1.00	1.00	0.92	0.63	0.89	0.95	1.00	1.00	1.00	0.92	0.63	0.89
OS	Mean	71.18	75.77	76.09	77.53	69.67	62.21	72.17	71.54	77.12	78.57	81.71	70.77	62.36	72.93
OS	Bias	-4.22	0.37	0.70	2.14	-5.72	-13.19	-3.23	-3.86	1.73	3.17	6.31	-4.63	-13.04	-2.47
OS	MC SE	0.21	0.19	0.19	0.24	0.16	0.10	0.22	0.22	0.36	0.48	0.43	0.35	0.10	0.22
OS	Rel Bias	- 0.06	0.00	0.01	0.03	- 0.08	- 0.17	- 0.04	- 0.05	0.02	0.04	0.08	- 0.06	- 0.17	- 0.03
OS	empSE	6.57	5.76	5.95	7.38	4.85	3.03	6.64	6.81	11.06	14.84	13.36	10.67	3.06	6.85
OS	MC SE	0.15	0.13	0.14	0.17	0.11	0.07	0.15	0.16	0.25	0.34	0.31	0.24	0.07	0.16
OS	MSE	60.88	33.28	35.91	58.89	56.21	183.08	54.49	61.20	125.08	230.07	218.11	135.08	179.30	52.99
OS	MC SE	0.15	0.13	0.14	0.17	0.11	0.07	0.15	1.71	42.27	69.11	58.45	41.52	2.59	1.60
OS	Rel P	0.00%	30.00%	21.63%	-20.71%	83.48%	369.37%	-2.26%	0.00%	-62.07%	-78.95%	-74.02%	-59.25%	395.71%	-1.28%
OS	ModelSE	5.89	17.41	23.71	21.05	17.18	3.53	6.28	5.89	17.41	23.71	21.05	17.18	3.53	6.28
OS	Cover2	0.73	0.93	0.96	0.99	0.67	0.05	0.80	0.74	0.93	0.95	0.99	0.68	0.06	0.83
OS	Cover1	0.98	0.99	1.00	1.00	0.98	0.55	1.00	0.98	0.99	1.00	1.00	0.98	0.55	1.00
QALYs	Mean	51.23	55.00	55.17	55.85	49.88	44.81	50.34	51.42	55.87	56.59	58.12	50.39	44.92	50.79
QALYs	Bias	-3.24	0.53	0.69	1.37	-4.60	-9.66	-4.13	-3.06	1.39	2.12	3.64	-4.08	-9.56	-3.68
QALYs	MC SE	0.13	0.14	0.14	0.16	0.11	0.07	0.13	0.14	0.21	0.26	0.24	0.19	0.07	0.13
QALYs	Rel Bias	- 0.06	0.01	0.01	0.03	- 0.08	- 0.18	- 0.08	- 0.06	0.03	0.04	0.07	- 0.07	- 0.18	- 0.07
QALYs	empSE	4.07	4.30	4.33	4.83	3.38	2.24	3.86	4.21	6.45	8.04	7.43	5.77	2.26	3.97
QALYs	MC SE	0.09	0.10	0.10	0.11	0.08	0.05	0.09	0.10	0.15	0.18	0.17	0.13	0.05	0.09
QALYs	MSE	27.05	18.74	19.18	25.16	32.56	98.34	31.95	27.04	43.55	69.05	68.47	49.92	96.42	29.31
QALYs	MC SE	0.81	0.76	0.77	1.07	0.95	1.40	0.98	0.80	10.87	17.24	14.60	9.70	1.40	0.91
QALYs	Rel P	0.00%	-10.39%	-11.53%	-28.94%	45.14%	230.56%	11.22%	0.00%	-57.51%	-72.61%	-67.96%	-46.81%	246.83%	12.31%
QALYs	ModelSE	3.36	9.25	12.26	11.01	8.96	2.59	3.96	3.36	9.25	12.26	11.01	8.96	2.59	3.96
QALYs	Cover2	0.67	0.93	0.95	0.98	0.64	0.05	0.69	0.67	0.93	0.95	0.99	0.65	0.06	0.71
QALYs	Cover1	0.91	1.00	1.00	1.00	0.96	0.56	1.00	0.91	1.00	1.00	1.00	0.96	0.56	1.00

Table 53 : Result for simulated scenario 31

	Scenario31	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	66.45	66.47	66.36	66.37	63.99	58.29	60.76	67.01	67.12	67.01	67.02	63.41	58.41	61.54
PFS	Bias	-2.87	-2.85	-2.96	-2.95	-5.33	-11.03	-8.56	-2.31	-2.20	-2.31	-2.30	-5.91	-10.91	-7.78
PFS	MC SE	0.21	0.21	0.21	0.21	0.20	0.13	0.17	0.21	0.22	0.22	0.22	0.24	0.13	0.17
PFS	Rel Bias	-4.1%	-4.1%	-4.3%	-4.3%	-7.7%	-15.9%	-12.3%	-3.3%	-3.2%	-3.3%	-3.3%	-8.5%	-15.7%	-11.2%
PFS	empSE	6.44	6.45	6.45	6.45	6.27	3.87	5.23	6.61	6.71	6.70	6.69	7.31	3.89	5.39
PFS	MC SE	0.15	0.15	0.15	0.15	0.14	0.09	0.12	0.15	0.15	0.15	0.15	0.17	0.09	0.12
PFS	MSE	49.74	49.75	50.31	50.22	67.62	136.53	100.58	49.01	49.77	50.17	50.04	88.34	134.08	89.50
PFS	MC SE	1.98	2.00	2.02	2.02	2.32	2.85	2.87	2.01	2.09	2.11	2.11	3.87	2.84	2.70
PFS	Rel P	0.00%	-0.31%	-0.18%	-0.08%	5.76%	177.41%	51.56%	0.00%	-2.80%	-2.61%	-2.47%	-18.32%	188.26%	50.54%
PFS	ModelSE	6.06	6.14	6.12	6.13	7.94	3.88	4.88	6.06	6.14	6.12	6.13	7.94	3.88	4.88
PFS	Cover2	0.83	0.83	0.82	0.83	0.72	0.23	0.50	0.84	0.84	0.83	0.84	0.72	0.24	0.54
PFS	Cover1	0.98	0.98	0.98	0.98	0.94	0.73	0.90	0.98	0.98	0.98	0.98	0.94	0.73	0.90
OS	Mean	322.77	327.17	327.74	328.98	323.12	271.67	320.66	325.01	325.73	333.83	335.38	323.10	273.96	328.26
OS	Bias	-4.72	-0.31	0.25	1.50	-4.36	-55.81	-6.83	-2.47	-1.76	6.35	7.90	-4.38	-53.52	0.78
OS	MC SE	0.86	0.88	0.91	1.06	0.87	0.77	0.81	0.87	0.94	1.47	1.46	1.03	0.79	0.79
OS	Rel Bias	-0.01	0.00	0.00	0.00	-0.01	-0.17	0.02	-0.01	0.01	0.02	0.02	-0.01	0.16	0.00
OS	empSE	26.42	27.03	28.18	32.74	26.73	23.74	24.91	26.76	28.89	45.43	44.95	31.66	24.23	24.43
OS	MC SE	0.61	0.62	0.65	0.75	0.61	0.54	0.57	0.61	0.66	1.04	1.03	0.73	0.56	0.56
OS	MSE	719.42	729.81	793.62	1073.11	732.59	3677.87	666.40	721.55	836.90	2101.93	2080.66	1020.35	3451.06	596.98
OS	MC SE	0.61	0.62	0.65	0.75	0.61	0.54	0.57	23.32	41.92	406.72	290.69	55.99	83.57	21.85
OS	Rel P	0.00%	-4.46%	-12.15%	-34.90%	-2.29%	23.87%	12.48%	0.00%	-14.20%	-65.30%	-64.55%	-28.54%	21.97%	19.96%
OS	ModelSE	43.51	40.72	50.12	58.49	47.10	28.24	45.46	43.51	40.72	50.12	58.49	47.10	28.24	45.46
OS	Cover2	1.00	0.98	0.98	0.98	0.97	0.47	1.00	1.00	0.98	0.99	0.99	0.98	0.49	1.00
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00
QALYs	Mean	181.32	183.53	183.78	184.40	180.76	153.32	178.56	182.61	183.00	187.02	187.80	180.57	154.51	182.59
QALYs	Bias	-3.22	-1.01	-0.76	-0.14	-3.78	-31.21	-5.98	-1.93	-1.54	2.48	3.26	-3.96	-30.03	-1.94
QALYs	MC SE	0.43	0.45	0.47	0.54	0.45	0.40	0.41	0.44	0.48	0.75	0.74	0.54	0.41	0.41
QALYs	Rel Bias	-0.02	-0.01	-0.00	-0.00	-0.02	-0.17	0.03	-0.01	0.01	0.01	0.02	-0.02	-0.16	0.01
QALYs	empSE	13.39	13.84	14.41	16.67	13.75	12.23	12.72	13.59	14.83	22.99	22.79	16.64	12.48	12.55
QALYs	MC SE	0.31	0.32	0.33	0.38	0.32	0.28	0.29	0.31	0.34	0.53	0.52	0.38	0.29	0.29
QALYs	MSE	189.59	192.26	208.10	277.72	203.04	1123.84	197.39	188.30	222.00	534.15	529.41	292.24	1057.62	161.02
QALYs	MC SE	5.98	6.67	7.62	10.39	7.54	24.63	7.31	6.24	11.35	101.59	72.60	17.54	24.23	6.03
QALYs	Rel P	0.00%	-6.28%	-13.64%	-35.46%	-5.04%	19.87%	10.88%	0.00%	-15.96%	-65.04%	-64.42%	-33.25%	18.56%	17.39%
QALYs	ModelSE	21.88	20.63	25.28	29.46	24.84	14.49	23.11	21.88	20.63	25.28	29.46	24.84	14.49	23.11
QALYs	Cover2	1.00	0.98	0.97	0.98	0.96	0.41	1.00	1.00	0.98	0.98	0.98	0.97	0.44	1.00
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00

Table 54 : Result for simulated scenario 32

	Scenario32	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	65.39	66.15	66.17	66.02	63.16	57.87	59.01	65.84	66.79	66.81	66.65	63.17	57.97	59.30
PFS	Bias	-3.74	-2.99	-2.97	-3.11	-5.98	-11.27	-10.13	-3.29	-2.35	-2.33	-2.48	-5.97	-11.16	-9.83
PFS	MC SE	0.20	0.21	0.21	0.20	0.20	0.12	0.15	0.20	0.21	0.21	0.21	0.22	0.12	0.15
PFS	Rel Bias	-5.4%	-4.3%	-4.3%	-4.5%	-8.6%	-16.3%	-14.7%	-4.8%	-3.4%	-3.4%	-3.6%	-8.6%	-16.1%	-14.2%
PFS	empSE	6.02	6.33	6.36	6.32	6.22	3.81	4.67	6.14	6.57	6.61	6.56	6.79	3.83	4.72
PFS	MC SE	0.14	0.15	0.15	0.14	0.14	0.09	0.11	0.14	0.15	0.15	0.15	0.16	0.09	0.11
PFS	MSE	50.23	48.89	49.19	49.54	74.36	141.44	124.36	48.43	48.66	49.07	49.18	81.61	139.25	119.00
PFS	MC SE	1.92	1.95	1.97	1.96	2.53	2.84	3.07	1.91	2.03	2.06	2.06	3.09	2.83	3.02
PFS	Rel P	0.00%	-9.41%	-10.33%	-9.10%	-6.29%	150.06%	66.30%	0.00%	-12.85%	-13.86%	-12.59%	-18.27%	156.19%	68.66%
PFS	ModelSE	5.56	6.07	6.09	6.05	7.48	3.81	4.37	5.56	6.07	6.09	6.05	7.48	3.81	4.37
PFS	Cover2	0.80	0.83	0.83	0.82	0.68	0.20	0.37	0.81	0.84	0.84	0.83	0.70	0.20	0.38
PFS	Cover1	0.98	0.98	0.98	0.98	0.94	0.70	0.83	0.98	0.98	0.98	0.98	0.94	0.70	0.83
OS	Mean	195.28	188.75	190.07	193.56	187.21	165.72	192.88	196.23	188.93	192.60	199.92	186.33	166.35	194.92
OS	Bias	4.41	-2.12	-0.80	2.69	-3.66	-25.15	2.01	5.36	-1.93	1.73	9.05	-4.54	-24.52	4.05
OS	MC SE	0.60	0.43	0.47	0.59	0.43	0.31	0.55	0.62	0.47	0.73	1.02	0.55	0.32	0.56
OS	Rel Bias	0.02	-0.01	-0.00	0.01	-0.02	0.13	0.01	0.03	-0.01	0.01	0.05	-0.02	0.13	0.02
OS	empSE	18.56	13.31	14.43	18.18	13.35	9.66	17.03	18.98	14.46	22.55	31.34	17.08	9.81	17.15
OS	MC SE	0.43	0.31	0.33	0.42	0.31	0.22	0.39	0.44	0.33	0.52	0.72	0.39	0.23	0.39
OS	MSE	363.59	181.53	208.62	337.53	191.51	725.65	293.91	388.54	212.60	510.80	1063.03	312.19	697.17	310.18
OS	MC SE	0.43	0.31	0.33	0.42	0.31	0.22	0.39	13.64	15.67	129.99	198.58	42.39	15.69	10.93
OS	Rel P	0.00%	94.37%	65.45%	4.17%	93.17%	268.96%	18.72%	0.00%	72.26%	-29.15%	-63.33%	23.40%	274.55%	22.46%
OS	ModelSE	17.95	17.00	26.00	38.70	21.60	11.50	18.15	17.95	17.00	26.00	38.70	21.60	11.50	18.15
OS	Cover2	0.98	0.95	0.96	0.97	0.94	0.40	0.98	0.98	0.95	0.96	0.98	0.94	0.43	0.99
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00
QALYs	Mean	117.26	114.22	114.89	116.58	112.55	100.22	114.14	117.87	114.50	116.34	119.96	112.11	100.57	115.25
QALYs	Bias	1.08	-1.96	-1.29	0.41	-3.62	-15.95	-2.03	1.69	-1.67	0.17	3.78	-4.06	-15.61	-0.93
QALYs	MC SE	0.31	0.24	0.26	0.31	0.24	0.17	0.29	0.32	0.26	0.39	0.52	0.31	0.18	0.29
QALYs	Rel Bias	0.01	-0.02	-0.01	0.00	-0.03	-0.14	-0.02	0.01	-0.01	0.00	0.03	-0.03	-0.13	-0.01
QALYs	empSE	9.61	7.50	7.94	9.63	7.48	5.35	8.83	9.86	8.06	11.88	16.06	9.53	5.42	8.91
QALYs	MC SE	0.22	0.17	0.18	0.22	0.17	0.12	0.20	0.23	0.19	0.27	0.37	0.22	0.12	0.20
QALYs	MSE	93.50	60.02	64.71	92.90	68.97	283.12	81.94	99.94	67.72	141.01	271.89	107.19	272.97	80.15
QALYs	MC SE	3.16	2.35	2.41	3.41	2.64	5.57	2.60	3.49	4.17	33.18	49.12	11.73	5.52	2.57
QALYs	Rel P	0.00%	64.30%	46.44%	-0.44%	65.30%	222.87%	18.64%	0.00%	49.52%	-31.15%	-62.31%	7.03%	230.35%	22.43%
QALYs	ModelSE	9.21	9.19	13.49	19.70	12.32	6.30	9.65	9.21	9.19	13.49	19.70	12.32	6.30	9.65
QALYs	Cover2	0.95	0.93	0.94	0.96	0.91	0.30	0.93	0.96	0.93	0.95	0.97	0.91	0.31	0.95
QALYs	Cover1	1.00	1.00	1.00	1.00	0.99	0.85	1.00	1.00	1.00	1.00	1.00	0.99	0.85	1.00

Table 55 : Result for simulated scenario 33

	Scenario33	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	64.34	65.96	65.96	65.84	62.50	57.71	57.49	64.66	66.58	66.57	66.46	62.71	57.81	57.76
PFS	Bias	-4.75	-3.13	-3.13	-3.24	-6.59	-11.38	-11.60	-4.43	-2.51	-2.51	-2.63	-6.38	-11.28	-11.33
PFS	MC SE	0.18	0.21	0.21	0.21	0.20	0.12	0.14	0.18	0.21	0.21	0.22	0.22	0.12	0.14
PFS	Rel Bias	-6.9%	-4.5%	-4.5%	-4.7%	-9.5%	-16.5%	-16.8%	-6.4%	-3.6%	-3.6%	-3.8%	-9.2%	-16.3%	-16.4%
PFS	empSE	5.60	6.34	6.34	6.38	6.16	3.69	4.18	5.62	6.59	6.59	6.63	6.76	3.72	4.22
PFS	MC SE	0.13	0.15	0.15	0.15	0.14	0.08	0.10	0.13	0.15	0.15	0.15	0.16	0.09	0.10
PFS	MSE	53.88	49.99	49.98	51.15	81.30	143.18	151.99	51.13	49.70	49.72	50.79	86.29	141.10	146.02
PFS	MC SE	1.93	1.95	1.98	2.04	2.63	2.77	3.16	1.87	2.02	2.08	2.14	3.43	2.77	3.12
PFS	Rel P	0.00%	-21.97%	-21.92%	-22.81%	-17.29%	129.95%	80.08%	0.00%	-27.29%	-27.30%	-28.12%	-30.85%	128.33%	77.75%
PFS	ModelSE	5.09	6.01	6.01	5.99	7.23	3.78	4.04	5.09	6.01	6.01	5.99	7.23	3.78	4.04
PFS	Cover2	0.74	0.82	0.82	0.81	0.64	0.18	0.22	0.75	0.83	0.83	0.82	0.67	0.19	0.23
PFS	Cover1	0.96	0.98	0.98	0.98	0.92	0.69	0.73	0.96	0.98	0.98	0.98	0.92	0.69	0.73
OS	Mean	134.15	129.28	129.85	131.60	126.24	116.14	133.03	134.81	130.44	131.09	133.98	126.73	116.39	133.96
OS	Bias	1.24	-3.62	-3.05	-1.30	-6.67	-16.76	0.12	1.91	-2.46	-1.82	1.07	-6.18	-16.52	1.05
OS	MC SE	0.40	0.28	0.29	0.36	0.28	0.18	0.38	0.41	0.53	0.44	0.40	0.52	0.18	0.39
OS	Rel Bias	0.01	-0.03	-0.02	-0.01	-0.05	0.13	0.00	0.01	-0.02	0.01	0.01	-0.05	0.12	0.01
OS	empSE	12.20	8.78	9.08	10.98	8.61	5.40	11.60	12.58	16.34	13.47	12.35	16.10	5.44	11.87
OS	MC SE	0.28	0.20	0.21	0.25	0.20	0.12	0.27	0.29	0.38	0.31	0.28	0.37	0.12	0.27
OS	MSE	150.29	90.20	91.61	122.18	118.46	310.05	134.43	161.74	272.74	184.63	153.56	296.94	302.31	141.81
OS	MC SE	0.28	0.20	0.21	0.25	0.20	0.12	0.27	5.16	91.30	61.66	13.05	83.14	5.82	4.63
OS	Rel P	0.00%	93.01%	80.77%	23.47%	101.05%	411.07%	10.66%	0.00%	-40.71%	-12.80%	3.74%	-38.91%	435.59%	12.38%
OS	ModelSE	10.05	20.36	16.95	15.91	20.46	6.25	10.10	10.05	20.36	16.95	15.91	20.46	6.25	10.10
OS	Cover2	0.84	0.86	0.89	0.92	0.79	0.26	0.85	0.84	0.87	0.90	0.94	0.81	0.27	0.86
OS	Cover1	0.99	0.99	0.99	1.00	0.99	0.78	1.00	0.99	0.99	0.99	1.00	0.99	0.78	1.00
QALYs	Mean	86.38	84.43	84.71	85.55	81.87	75.38	83.76	86.81	85.19	85.51	86.93	82.18	75.54	84.31
QALYs	Bias	-0.80	-2.75	-2.47	-1.63	-5.31	-11.80	-3.42	-0.37	-1.98	-1.66	-0.25	-5.00	-11.64	-2.87
QALYs	MC SE	0.22	0.19	0.19	0.21	0.18	0.11	0.20	0.23	0.30	0.25	0.24	0.30	0.11	0.21
QALYs	Rel Bias	-0.01	-0.03	-0.03	-0.02	-0.06	-0.14	-0.04	-0.00	-0.02	-0.02	-0.00	-0.06	-0.13	-0.03
QALYs	empSE	6.82	5.78	5.84	6.54	5.61	3.45	6.25	7.04	9.14	7.83	7.25	9.14	3.48	6.41
QALYs	MC SE	0.16	0.13	0.13	0.15	0.13	0.08	0.14	0.16	0.21	0.18	0.17	0.21	0.08	0.15
QALYs	MSE	47.06	40.94	40.20	45.40	59.63	151.02	50.72	49.66	87.34	63.99	52.53	108.54	147.62	49.24
QALYs	MC SE	1.53	1.56	1.50	1.67	2.04	2.64	1.83	1.60	23.35	16.36	3.56	21.15	2.63	1.74
QALYs	Rel P	0.00%	39.09%	36.06%	8.57%	47.68%	290.35%	18.90%	0.00%	-40.62%	-19.11%	-5.60%	-40.71%	310.32%	20.80%
QALYs	ModelSE	5.33	10.79	9.21	8.77	11.45	3.94	5.80	5.33	10.79	9.21	8.77	11.45	3.94	5.80
QALYs	Cover2	0.79	0.84	0.86	0.91	0.74	0.18	0.77	0.79	0.84	0.87	0.93	0.76	0.19	0.79
QALYs	Cover1	0.94	0.99	0.99	1.00	0.97	0.72	0.99	0.94	0.99	0.99	1.00	0.97	0.72	0.99

Table 56 : Result for simulated scenario 34

	Scenario34	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.99	59.07	59.00	58.98	54.19	50.05	53.11	59.48	59.60	59.52	59.51	54.01	50.13	53.34
PFS	Bias	2.87	2.95	2.88	2.86	-1.93	-6.07	-3.02	3.36	3.47	3.40	3.38	-2.12	-5.99	-2.79
PFS	MC SE	0.17	0.17	0.17	0.17	0.15	0.10	0.13	0.17	0.17	0.17	0.17	0.15	0.10	0.13
PFS	Rel Bias	5.1%	5.2%	5.1%	5.1%	-3.4%	-10.8%	-5.4%	6.0%	6.2%	6.1%	6.0%	-3.8%	-10.7%	-5.0%
PFS	empSE	5.17	5.16	5.18	5.21	4.68	3.09	4.05	5.27	5.28	5.29	5.32	4.73	3.11	4.08
PFS	MC SE	0.12	0.12	0.12	0.12	0.11	0.07	0.09	0.12	0.12	0.12	0.12	0.11	0.07	0.09
PFS	MSE	34.93	35.33	35.08	35.23	25.58	46.39	25.49	38.97	39.87	39.56	39.71	26.86	45.55	24.39
PFS	MC SE	1.69	1.71	1.72	1.73	1.05	1.27	0.98	1.86	1.89	1.91	1.92	1.11	1.26	0.96
PFS	Rel P	0.00%	0.27%	-0.30%	-1.28%	22.30%	180.22%	63.20%	0.00%	-0.47%	-1.09%	-2.07%	23.70%	186.94%	66.70%
PFS	ModelSE	4.82	4.83	4.82	4.82	4.39	3.09	3.85	4.82	4.83	4.82	4.82	4.39	3.09	3.85
PFS	Cover2	0.91	0.91	0.92	0.91	0.84	0.49	0.79	0.90	0.90	0.90	0.90	0.83	0.50	0.81
PFS	Cover1	0.99	0.99	0.99	0.99	0.98	0.90	0.99	0.99	0.99	0.99	0.99	0.98	0.90	0.99
OS	Mean	302.03	297.76	298.55	299.79	296.60	229.84	296.06	304.98	302.84	310.51	314.28	302.78	232.81	301.44
OS	Bias	3.88	-0.40	0.40	1.64	-1.55	-68.31	-2.10	6.82	4.68	12.35	16.13	4.63	-65.35	3.29
OS	MC SE	0.88	0.90	0.99	1.13	0.91	0.77	0.86	0.99	1.36	1.92	1.83	1.40	0.80	0.87
OS	Rel Bias	0.01 -	0.00	0.00	0.01 -	0.01 -	0.23 -	0.01	0.02	0.02	0.04	0.05	0.02 -	0.22	0.01
OS	empSE	27.26	27.83	30.54	34.92	28.14	23.80	26.58	30.45	41.86	59.09	56.40	43.29	24.52	26.91
OS	MC SE	0.63	0.64	0.70	0.80	0.65	0.55	0.61	0.70	0.96	1.36	1.29	0.99	0.56	0.62
OS	MSE	757.10	774.12	931.90	1220.54	793.49	5232.17	710.14	972.95	1772.55	3640.09	3437.25	1893.85	4871.22	734.23
OS	MC SE	0.63	0.64	0.70	0.80	0.65	0.55	0.61	105.39	330.06	637.20	467.19	344.45	100.49	29.73
OS	Rel P	0.00%	-4.12%	-20.36%	-39.07%	-6.19%	31.17%	5.15%	0.00%	-47.08%	-73.44%	-70.84%	-50.52%	54.26%	28.06%
OS	ModelSE	43.60	48.12	57.29	65.15	48.49	28.63	40.74	43.60	48.12	57.29	65.15	48.49	28.63	40.74
OS	Cover2	1.00	0.99	0.98	0.99	0.99	0.34	0.99	1.00	0.99	0.98	0.99	0.98	0.37	0.99
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	0.89	1.00
QALYs	Mean	168.71	166.60	166.98	167.59	164.56	129.94	163.96	170.33	169.30	173.11	174.99	167.59	131.44	166.72
QALYs	Bias	2.80	0.69	1.06	1.68	-1.36	-35.98	-1.95	4.42	3.38	7.20	9.08	1.68	-34.47	0.81
QALYs	MC SE	0.45	0.46	0.50	0.57	0.46	0.39	0.44	0.50	0.68	0.96	0.92	0.71	0.40	0.45
QALYs	Rel Bias	0.02	0.00	0.01	0.01 -	0.01 -	0.22 -	0.01	0.03	0.02	0.04	0.05	0.01 -	0.21	0.00
QALYs	empSE	13.87	14.07	15.49	17.67	14.30	12.12	13.57	15.48	20.99	29.62	28.30	21.75	12.48	13.73
QALYs	MC SE	0.32	0.32	0.36	0.41	0.33	0.28	0.31	0.36	0.48	0.68	0.65	0.50	0.29	0.32
QALYs	MSE	199.93	198.15	240.66	314.77	206.00	1441.00	187.67	258.93	451.73	928.42	882.67	475.42	1343.93	189.07
QALYs	MC SE	7.15	7.08	8.57	12.58	7.33	27.57	6.45	27.78	82.55	160.53	117.95	84.67	27.09	7.39
QALYs	Rel P	0.00%	-2.82%	-19.80%	-38.42%	-5.91%	30.99%	4.49%	0.00%	-45.62%	-72.69%	-70.08%	-49.34%	53.90%	27.06%
QALYs	ModelSE	21.89	24.12	28.69	32.63	24.45	14.56	20.71	21.89	24.12	28.69	32.63	24.45	14.56	20.71
QALYs	Cover2	1.00	0.99	0.99	0.99	0.99	0.31	0.98	1.00	1.00	0.99	0.99	0.99	0.35	0.99
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.88	1.00	1.00	1.00	1.00	1.00	1.00	0.88	1.00

Table 57 : Result for simulated scenario 35

	Scenario35	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	57.85	59.03	59.07	58.96	53.88	50.12	51.38	58.18	59.56	59.60	59.49	53.85	50.19	51.41
PFS	Bias	1.81	2.99	3.03	2.92	-2.16	-5.93	-4.66	2.14	3.52	3.56	3.45	-2.19	-5.85	-4.63
PFS	MC SE	0.15	0.17	0.17	0.17	0.15	0.10	0.12	0.15	0.17	0.17	0.17	0.15	0.10	0.12
PFS	Rel Bias	3.2%	5.3%	5.4%	5.2%	-3.9%	-10.6%	-8.3%	3.8%	6.3%	6.4%	6.1%	-3.9%	-10.4%	-8.3%
PFS	empSE	4.54	5.24	5.24	5.20	4.59	3.07	3.62	4.57	5.35	5.35	5.31	4.66	3.09	3.65
PFS	MC SE	0.10	0.12	0.12	0.12	0.11	0.07	0.08	0.11	0.12	0.12	0.12	0.11	0.07	0.08
PFS	MSE	23.89	36.39	36.67	35.57	25.68	44.56	34.81	25.47	40.95	41.27	40.03	26.50	43.76	34.77
PFS	MC SE	1.16	1.79	1.80	1.75	1.03	1.25	1.16	1.24	1.97	1.98	1.93	1.08	1.24	1.15
PFS	Rel P	0.00%	-24.93%	-24.99%	-23.76%	-1.96%	118.21%	57.50%	0.00%	-26.86%	-26.90%	-25.74%	-3.71%	119.29%	57.16%
PFS	ModelSE	4.30	4.83	4.83	4.81	4.19	3.11	3.44	4.30	4.83	4.83	4.81	4.19	3.11	3.44
PFS	Cover2	0.93	0.90	0.91	0.91	0.82	0.51	0.65	0.93	0.89	0.89	0.89	0.82	0.52	0.66
PFS	Cover1	1.00	0.99	0.99	0.99	0.97	0.90	0.96	1.00	0.99	0.99	0.99	0.97	0.90	0.96
OS	Mean	152.42	159.09	159.92	162.47	157.12	135.72	151.51	153.57	160.84	164.13	170.20	157.63	136.37	153.22
OS	Bias	-4.10	2.57	3.40	5.95	0.60	-20.80	-5.01	-2.95	4.32	7.61	13.68	1.11	-20.15	-3.30
OS	MC SE	0.39	0.38	0.40	0.50	0.40	0.27	0.39	0.41	0.54	0.80	1.10	0.53	0.27	0.41
OS	Rel Bias	- 0.03	0.02	0.02	0.04	0.00	- 0.13	- 0.03	- 0.02	0.03	0.05	0.09	0.01	- 0.13	- 0.02
OS	empSE	12.11	11.62	12.43	15.32	12.19	8.22	11.96	12.55	16.57	24.64	33.90	16.35	8.29	12.53
OS	MC SE	0.28	0.27	0.29	0.35	0.28	0.19	0.27	0.29	0.38	0.57	0.78	0.38	0.19	0.29
OS	MSE	163.31	141.47	165.86	269.75	148.76	499.98	168.04	165.94	292.81	664.12	1334.91	268.31	474.76	167.76
OS	MC SE	0.28	0.27	0.29	0.35	0.28	0.19	0.27	6.27	89.60	183.60	321.51	73.96	10.93	5.71
OS	Rel P	0.00%	8.60%	-5.05%	-37.49%	-1.30%	116.93%	2.44%	0.00%	-42.65%	-74.07%	-86.30%	-41.13%	128.74%	0.21%
OS	ModelSE	15.96	18.14	26.22	35.15	18.33	10.16	16.22	15.96	18.14	26.22	35.15	18.33	10.16	16.22
OS	Cover2	0.97	0.98	0.98	0.99	0.97	0.45	0.95	0.97	0.98	0.99	0.99	0.96	0.47	0.96
OS	Cover1	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00
QALYs	Mean	93.56	97.25	97.68	98.93	94.72	82.90	91.17	94.24	98.29	99.94	102.94	94.97	83.24	92.03
QALYs	Bias	-1.51	2.18	2.61	3.85	-0.35	-12.18	-3.90	-0.84	3.21	4.87	7.87	-0.10	-11.83	-3.04
QALYs	MC SE	0.21	0.21	0.22	0.26	0.21	0.15	0.21	0.22	0.28	0.41	0.56	0.28	0.15	0.22
QALYs	Rel Bias	- 0.02	0.02	0.03	0.04	- 0.00	- 0.13	- 0.04	- 0.01	0.03	0.05	0.08	- 0.00	- 0.12	- 0.03
QALYs	empSE	6.52	6.35	6.73	8.00	6.54	4.49	6.38	6.76	8.75	12.65	17.11	8.54	4.53	6.68
QALYs	MC SE	0.15	0.15	0.15	0.18	0.15	0.10	0.15	0.16	0.20	0.29	0.39	0.20	0.10	0.15
QALYs	MSE	44.73	45.01	52.07	78.75	42.91	168.37	55.84	46.31	86.81	183.53	354.45	72.81	160.46	53.77
QALYs	MC SE	1.52	1.85	2.19	3.31	1.67	3.58	1.90	1.84	23.67	47.32	80.92	18.46	3.51	1.85
QALYs	Rel P	0.00%	5.49%	-6.17%	-33.57%	-0.78%	110.91%	4.53%	0.00%	-40.36%	-71.46%	-84.41%	-37.35%	122.43%	2.40%
QALYs	ModelSE	8.25	9.40	13.35	17.77	9.57	5.50	8.54	8.25	9.40	13.35	17.77	9.57	5.50	8.54
QALYs	Cover2	0.96	0.97	0.99	0.99	0.95	0.40	0.92	0.97	0.97	0.99	0.99	0.95	0.41	0.93
QALYs	Cover1	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00

Table 58 : Result for simulated scenario 36

	Scenario36	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	55.46	58.99	58.92	58.90	53.63	49.96	49.81	55.63	59.52	59.45	59.42	53.62	50.03	49.86
PFS	Bias	-0.49	3.04	2.97	2.94	-2.32	-5.99	-6.14	-0.32	3.56	3.49	3.47	-2.33	-5.92	-6.09
PFS	MC SE	0.13	0.17	0.17	0.17	0.15	0.10	0.11	0.13	0.17	0.17	0.17	0.15	0.10	0.11
PFS	Rel Bias	-0.9%	5.4%	5.3%	5.3%	-4.1%	-10.7%	-11.0%	-0.6%	6.4%	6.2%	6.2%	-4.2%	-10.6%	-10.9%
PFS	empSE	3.90	5.15	5.11	5.18	4.52	2.97	3.39	3.91	5.27	5.22	5.30	4.66	2.98	3.40
PFS	MC SE	0.09	0.12	0.12	0.12	0.10	0.07	0.08	0.09	0.12	0.12	0.12	0.11	0.07	0.08
PFS	MSE	15.44	35.78	34.89	35.48	25.78	44.72	49.19	15.37	40.43	39.44	40.03	27.11	43.92	48.69
PFS	MC SE	0.68	1.66	1.61	1.66	1.00	1.23	1.39	0.68	1.85	1.79	1.85	1.07	1.22	1.39
PFS	Rel P	0.00%	-42.70%	-41.67%	-43.33%	-25.50%	72.60%	32.44%	0.00%	-44.94%	-43.94%	-45.53%	-29.52%	71.74%	32.05%
PFS	ModelSE	3.62	4.83	4.82	4.82	4.07	3.10	3.15	3.62	4.83	4.82	4.82	4.07	3.10	3.15
PFS	Cover2	0.91	0.91	0.91	0.91	0.81	0.51	0.47	0.91	0.90	0.90	0.89	0.80	0.52	0.47
PFS	Cover1	0.99	1.00	1.00	1.00	0.98	0.89	0.90	0.99	1.00	1.00	1.00	0.98	0.89	0.90
OS	Mean	100.38	106.14	106.46	107.75	102.14	93.61	99.92	100.75	108.66	109.64	111.32	103.05	93.82	100.40
OS	Bias	-3.93	1.82	2.14	3.44	-2.17	-10.71	-4.40	-3.56	4.35	5.32	7.00	-1.27	-10.49	-3.92
OS	MC SE	0.28	0.22	0.24	0.28	0.22	0.14	0.27	0.28	0.59	0.76	0.57	0.46	0.15	0.27
OS	Rel Bias	- 0.04	0.02	0.02	0.03	- 0.02	- 0.10	- 0.04	- 0.03	0.04	0.05	0.07	- 0.01	- 0.10	- 0.04
OS	empSE	8.48	6.91	7.32	8.62	6.92	4.47	8.19	8.65	18.25	23.47	17.57	14.09	4.49	8.43
OS	MC SE	0.19	0.16	0.17	0.20	0.16	0.10	0.19	0.20	0.42	0.54	0.40	0.32	0.10	0.19
OS	MSE	87.39	51.05	58.12	86.04	52.52	134.55	86.42	87.49	351.60	578.37	357.34	199.82	130.31	86.40
OS	MC SE	0.19	0.16	0.17	0.20	0.16	0.10	0.19	3.06	103.29	229.78	100.76	71.23	3.08	2.78
OS	Rel P	0.00%	50.64%	34.29%	-3.14%	50.45%	261.00%	7.22%	0.00%	-77.52%	-86.40%	-75.75%	-62.27%	270.71%	5.26%
OS	ModelSE	6.78	26.41	26.38	23.78	19.39	5.16	6.94	6.78	26.41	26.38	23.78	19.39	5.16	6.94
OS	Cover2	0.79	0.93	0.95	0.97	0.90	0.44	0.79	0.81	0.92	0.95	0.97	0.90	0.45	0.81
OS	Cover1	1.00	1.00	1.00	1.00	0.99	0.91	1.00	1.00	1.00	1.00	1.00	0.99	0.91	1.00
QALYs	Mean	66.83	70.77	70.91	71.55	67.16	61.79	64.90	67.07	72.19	72.65	73.48	67.61	61.92	65.16
QALYs	Bias	-2.11	1.82	1.96	2.60	-1.78	-7.15	-4.04	-1.88	3.24	3.71	4.54	-1.33	-7.02	-3.79
QALYs	MC SE	0.16	0.15	0.15	0.17	0.14	0.09	0.15	0.16	0.31	0.39	0.30	0.25	0.09	0.15
QALYs	Rel Bias	- 0.03	0.03	0.03	0.04	- 0.03	- 0.10	- 0.06	- 0.03	0.05	0.05	0.07	- 0.02	- 0.10	- 0.05
QALYs	empSE	4.85	4.56	4.68	5.19	4.37	2.83	4.63	4.95	9.65	12.16	9.31	7.67	2.85	4.75
QALYs	MC SE	0.11	0.10	0.11	0.12	0.10	0.07	0.11	0.11	0.22	0.28	0.21	0.18	0.07	0.11
QALYs	MSE	27.99	24.09	25.69	33.68	22.26	59.16	37.75	28.00	103.51	161.43	107.11	60.50	57.44	36.87
QALYs	MC SE	0.95	1.06	1.05	1.46	0.84	1.33	1.14	1.00	26.49	58.10	25.83	18.19	1.31	1.11
QALYs	Rel P	0.00%	13.21%	7.64%	-12.64%	23.22%	193.03%	9.83%	0.00%	-73.68%	-83.43%	-71.71%	-58.32%	201.56%	8.65%
QALYs	ModelSE	3.71	13.49	13.47	12.22	10.03	3.24	4.10	3.71	13.49	13.47	12.22	10.03	3.24	4.10
QALYs	Cover2	0.78	0.92	0.94	0.97	0.88	0.39	0.69	0.78	0.90	0.92	0.97	0.87	0.41	0.71
QALYs	Cover1	0.97	0.99	1.00	1.00	0.99	0.88	0.99	0.97	0.99	1.00	1.00	0.99	0.88	0.99

Table 59 : Result for simulated scenario 37

	Scenario37	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	60.50	61.45	61.17	61.09	56.02	47.19	57.35	61.01	62.34	62.06	61.98	54.98	47.41	57.84
PFS	Bias	-8.65	-7.71	-7.98	-8.06	-13.13	-21.96	-11.80	-8.14	-6.81	-7.09	-7.18	-14.17	-21.74	-11.31
PFS	MC SE	0.22	0.25	0.24	0.24	0.20	0.12	0.22	0.22	0.26	0.25	0.25	0.22	0.12	0.23
PFS	Rel Bias	-12.5%	-11.1%	-11.5%	-11.7%	-19.0%	-31.8%	-17.1%	-11.8%	-9.8%	-10.2%	-10.4%	-20.5%	-31.4%	-16.4%
PFS	empSE	6.86	7.59	7.45	7.41	6.16	3.58	6.89	6.89	7.89	7.75	7.71	6.87	3.63	7.03
PFS	MC SE	0.16	0.17	0.17	0.17	0.14	0.08	0.16	0.16	0.18	0.18	0.18	0.16	0.08	0.16
PFS	MSE	121.84	116.96	119.05	119.92	210.39	495.05	186.73	113.65	108.57	110.17	110.93	248.07	485.98	177.37
PFS	MC SE	3.75	3.74	3.78	3.78	4.66	4.93	5.03	3.62	3.66	3.70	3.70	6.11	4.95	4.96
PFS	Rel P	0.00%	-18.36%	-15.12%	-14.35%	24.12%	267.61%	-0.99%	0.00%	-23.87%	-20.98%	-20.31%	0.45%	260.21%	-4.14%
PFS	ModelSE	6.76	7.35	7.28	7.25	8.84	3.79	6.94	6.76	7.35	7.28	7.25	8.84	3.79	6.94
PFS	Cover2	0.62	0.67	0.66	0.66	0.42	0.01	0.49	0.65	0.70	0.69	0.69	0.42	0.01	0.51
PFS	Cover1	0.94	0.94	0.94	0.94	0.82	0.27	0.90	0.94	0.94	0.94	0.94	0.82	0.27	0.90
OS	Mean	165.41	158.18	179.96	296.32	153.73	81.24	173.77	176.21	164.38	201.18	311.84	160.57	81.85	183.90
OS	Bias	2.50	-4.74	17.04	133.40	-9.19	-81.68	10.86	13.29	1.47	38.26	148.92	-2.35	-81.06	20.98
OS	MC SE	0.64	0.78	1.11	3.38	0.76	0.18	0.65	1.45	0.88	1.83	2.79	0.95	0.18	0.78
OS	Rel Bias	0.02	-0.03	0.10	0.82	-0.06	0.50	0.07	0.08	0.01	0.23	0.91	-0.01	0.50	0.13
OS	empSE	19.68	23.98	34.15	104.24	23.44	5.41	20.07	44.78	27.10	56.38	86.09	29.32	5.56	24.07
OS	MC SE	0.45	0.55	0.78	2.39	0.54	0.12	0.46	1.03	0.62	1.29	1.98	0.67	0.13	0.55
OS	MSE	393.14	596.69	1455.29	28651.83	633.32	6700.77	520.12	2179.65	735.69	4639.77	29580.83	864.19	6602.43	1019.07
OS	MC SE	0.45	0.55	0.78	2.39	0.54	0.12	0.46	515.12	33.62	497.13	915.40	38.71	29.17	70.58
OS	Rel P	0.00%	-32.63%	-66.78%	-96.44%	-29.52%	1222.78%	-3.82%	0.00%	173.05%	-36.93%	-72.94%	133.26%	6382.95%	246.08%
OS	ModelSE	58.76	46.73	86.70	121.97	50.65	6.41	52.34	58.76	46.73	86.70	121.97	50.65	6.41	52.34
OS	Cover2	1.00	0.97	1.00	0.68	0.97	0.00	1.00	1.00	0.98	1.00	0.71	0.97	0.00	1.00
OS	Cover1	1.00	1.00	1.00	0.96	1.00	0.01	1.00	1.00	1.00	1.00	0.96	1.00	0.01	1.00
QALYs	Mean	100.86	97.52	108.33	166.49	93.67	54.78	104.09	106.41	100.89	119.21	174.51	96.78	55.15	109.30
QALYs	Bias	-1.35	-4.68	6.13	64.28	-8.53	-47.43	1.89	4.20	-1.31	17.00	72.31	-5.43	-47.06	7.10
QALYs	MC SE	0.33	0.40	0.57	1.70	0.39	0.11	0.34	0.73	0.45	0.93	1.41	0.49	0.12	0.40
QALYs	Rel Bias	-0.01	-0.05	0.06	0.63	-0.08	0.46	0.02	0.04	-0.01	0.17	0.71	-0.05	0.46	0.07
QALYs	empSE	10.07	12.34	17.44	52.49	12.00	3.48	10.58	22.59	13.91	28.60	43.47	15.25	3.57	12.48
QALYs	MC SE	0.23	0.28	0.40	1.20	0.28	0.08	0.24	0.52	0.32	0.66	1.00	0.35	0.08	0.29
QALYs	MSE	103.07	174.02	341.44	6884.17	216.74	2261.53	115.48	527.53	195.13	1106.06	7116.11	261.79	2226.98	205.92
QALYs	MC SE	3.26	5.53	19.79	249.85	6.97	10.68	4.50	127.35	8.05	122.30	225.54	10.55	10.84	15.32
QALYs	Rel P	0.00%	-33.44%	-66.68%	-96.32%	-29.64%	735.00%	-9.53%	0.00%	163.61%	-37.59%	-72.99%	119.43%	3908.90%	227.81%
QALYs	ModelSE	29.69	23.77	43.69	61.34	26.36	4.05	26.92	29.69	23.77	43.69	61.34	26.36	4.05	26.92
QALYs	Cover2	1.00	0.96	1.00	0.69	0.95	0.00	1.00	1.00	0.97	1.00	0.72	0.96	0.00	1.00
QALYs	Cover1	1.00	0.99	1.00	0.96	1.00	0.01	1.00	1.00	0.99	1.00	0.96	1.00	0.01	1.00

Table 60 : Result for simulated scenario 38

	Scenario38	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	60.02	62.40	62.10	61.95	54.98	48.79	56.09	60.41	63.36	63.04	62.89	54.38	49.01	56.51
PFS	Bias	-9.04	-6.66	-6.96	-7.11	-14.08	-20.27	-12.97	-8.65	-5.70	-6.02	-6.17	-14.69	-20.06	-12.55
PFS	MC SE	0.19	0.25	0.25	0.25	0.20	0.12	0.19	0.19	0.26	0.26	0.26	0.22	0.12	0.20
PFS	Rel Bias	-13.1%	-9.6%	-10.1%	-10.3%	-20.4%	-29.4%	-18.8%	-12.5%	-8.3%	-8.7%	-8.9%	-21.3%	-29.0%	-18.2%
PFS	empSE	5.89	7.75	7.70	7.67	6.27	3.69	5.93	5.85	8.07	8.02	7.99	6.93	3.74	6.04
PFS	MC SE	0.14	0.18	0.18	0.18	0.14	0.08	0.14	0.13	0.19	0.18	0.18	0.16	0.09	0.14
PFS	MSE	116.38	104.39	107.63	109.36	237.55	424.48	203.47	109.05	97.66	100.43	101.89	263.71	416.20	194.00
PFS	MC SE	3.42	3.44	3.50	3.51	4.86	4.75	4.88	3.29	3.39	3.44	3.45	6.29	4.76	4.83
PFS	Rel P	0.00%	-42.34%	-41.48%	-41.15%	-11.78%	154.81%	-1.44%	0.00%	-47.55%	-46.78%	-46.49%	-28.87%	144.70%	-6.38%
PFS	ModelSE	6.20	7.55	7.48	7.45	8.38	4.09	6.22	6.20	7.55	7.48	7.45	8.38	4.09	6.22
PFS	Cover2	0.62	0.73	0.71	0.71	0.36	0.01	0.42	0.64	0.75	0.74	0.73	0.35	0.01	0.44
PFS	Cover1	0.93	0.95	0.95	0.95	0.79	0.39	0.89	0.93	0.95	0.95	0.95	0.79	0.39	0.89
OS	Mean	105.10	97.81	108.55	188.35	95.56	65.77	107.68	107.22	101.62	117.27	200.54	98.25	66.07	110.24
OS	Bias	-0.92	-8.21	2.53	82.33	-10.46	-40.26	1.65	1.20	-4.40	11.25	94.52	-7.78	-39.95	4.21
OS	MC SE	0.42	0.32	0.45	1.92	0.38	0.13	0.43	0.48	0.66	0.98	1.78	0.49	0.13	0.44
OS	Rel Bias	-0.01	-0.08	0.02	0.78	-0.10	0.38	0.02	-0.01	-0.04	0.11	0.89	-0.07	0.38	0.04
OS	empSE	13.09	9.73	13.97	59.09	11.64	3.94	13.10	14.65	20.29	30.19	54.94	15.21	4.01	13.66
OS	MC SE	0.30	0.22	0.32	1.36	0.27	0.09	0.30	0.34	0.47	0.69	1.26	0.35	0.09	0.31
OS	MSE	171.96	162.08	201.42	10265.83	244.79	1636.03	174.24	215.78	430.53	1037.36	11948.63	291.56	1612.21	204.25
OS	MC SE	0.30	0.22	0.32	1.36	0.27	0.09	0.30	17.39	129.27	216.59	393.34	13.49	10.31	10.36
OS	Rel P	0.00%	80.87%	-12.26%	-95.09%	26.41%	1002.45%	-0.23%	0.00%	-47.87%	-76.47%	-92.89%	-7.25%	1233.90%	14.94%
OS	ModelSE	20.32	28.76	44.64	79.46	22.45	4.66	19.06	20.32	28.76	44.64	79.46	22.45	4.66	19.06
OS	Cover2	0.99	0.86	0.99	0.75	0.84	0.00	1.00	0.99	0.89	0.99	0.78	0.87	0.00	1.00
OS	Cover1	1.00	1.00	1.00	0.96	0.99	0.04	1.00	1.00	1.00	1.00	0.96	0.99	0.04	1.00
QALYs	Mean	70.56	67.63	72.91	112.76	64.28	47.52	70.66	71.74	69.82	77.55	119.14	65.44	47.74	72.07
QALYs	Bias	-3.17	-6.11	-0.83	39.03	-9.45	-26.21	-3.07	-2.00	-3.91	3.82	45.41	-8.29	-25.99	-1.66
QALYs	MC SE	0.23	0.17	0.26	0.98	0.21	0.10	0.23	0.25	0.34	0.51	0.92	0.27	0.10	0.24
QALYs	Rel Bias	-0.04	-0.08	0.01	0.53	-0.13	0.36	0.04	-0.03	0.05	0.05	0.62	-0.11	0.35	0.02
QALYs	empSE	7.03	5.36	8.11	30.15	6.38	2.99	7.18	7.84	10.51	15.73	28.23	8.29	3.04	7.53
QALYs	MC SE	0.16	0.12	0.19	0.69	0.15	0.07	0.16	0.18	0.24	0.36	0.65	0.19	0.07	0.17
QALYs	MSE	59.49	65.94	66.46	2431.66	130.04	695.83	60.97	65.43	125.55	261.82	2857.99	137.45	684.84	59.37
QALYs	MC SE	1.90	2.25	2.18	93.79	3.85	5.02	2.11	4.10	32.76	52.70	97.73	5.45	5.07	2.92
QALYs	Rel P	0.00%	72.44%	-24.86%	-94.56%	21.60%	453.89%	-4.15%	0.00%	-44.26%	-75.14%	-92.28%	-10.50%	566.49%	8.52%
QALYs	ModelSE	10.56	15.10	22.90	40.29	12.68	3.47	10.74	10.56	15.10	22.90	40.29	12.68	3.47	10.74
QALYs	Cover2	0.93	0.91	0.97	0.79	0.76	0.00	0.94	0.94	0.92	0.98	0.82	0.80	0.00	0.97
QALYs	Cover1	1.00	1.00	1.00	0.97	0.98	0.09	1.00	1.00	1.00	1.00	0.97	0.98	0.09	1.00

Table 61 : Result for simulated scenario 39

	Scenario39	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	60.66	62.90	62.92	62.78	54.26	49.75	55.96	60.89	63.88	63.91	63.76	54.03	49.93	56.50
PFS	Bias	-8.38	-6.14	-6.12	-6.26	-14.78	-19.29	-13.08	-8.15	-5.16	-5.14	-5.28	-15.01	-19.11	-12.54
PFS	MC SE	0.19	0.24	0.25	0.25	0.18	0.12	0.17	0.19	0.25	0.26	0.26	0.19	0.12	0.18
PFS	Rel Bias	-12.1%	-8.9%	-8.9%	-9.1%	-21.4%	-27.9%	-18.9%	-11.8%	-7.5%	-7.4%	-7.7%	-21.7%	-27.7%	-18.2%
PFS	empSE	5.86	7.32	7.68	7.76	5.66	3.80	5.30	5.77	7.60	7.99	8.09	5.85	3.85	5.44
PFS	MC SE	0.13	0.17	0.18	0.18	0.13	0.09	0.12	0.13	0.17	0.18	0.19	0.13	0.09	0.12
PFS	MSE	104.57	91.26	96.31	99.47	250.44	386.71	199.13	99.70	84.30	90.19	93.22	259.58	379.93	186.98
PFS	MC SE	3.16	3.11	3.26	3.34	4.99	4.66	4.49	3.06	2.96	3.20	3.29	5.25	4.66	4.42
PFS	Rel P	0.00%	-35.95%	-41.70%	-43.04%	7.07%	137.64%	22.12%	0.00%	-42.38%	-47.95%	-49.13%	-2.72%	124.67%	12.20%
PFS	ModelSE	6.26	7.58	7.68	7.65	6.84	4.28	6.07	6.26	7.58	7.68	7.65	6.84	4.28	6.07
PFS	Cover2	0.65	0.77	0.76	0.75	0.29	0.03	0.40	0.66	0.80	0.79	0.78	0.29	0.03	0.44
PFS	Cover1	0.95	0.96	0.96	0.96	0.75	0.47	0.90	0.95	0.96	0.96	0.96	0.75	0.47	0.90
OS	Mean	84.29	75.83	83.51	132.30	69.84	58.32	85.35	85.49	77.20	86.83	139.42	70.74	58.53	86.80
OS	Bias	-2.34	-10.80	-3.12	45.67	-16.79	-28.31	-1.28	-1.14	-9.43	0.20	52.78	-15.89	-28.10	0.17
OS	MC SE	0.32	0.20	0.32	1.21	0.18	0.13	0.32	0.34	0.22	0.54	1.15	0.23	0.13	0.33
OS	Rel Bias	- 0.03	- 0.12	- 0.04	0.53	- 0.19	- 0.33	- 0.01	- 0.01	- 0.11	0.00	0.61	- 0.18	- 0.32	0.00
OS	empSE	9.91	6.18	9.89	37.31	5.69	3.86	9.96	10.37	6.78	16.74	35.47	7.03	3.91	10.17
OS	MC SE	0.23	0.14	0.23	0.86	0.13	0.09	0.23	0.24	0.16	0.38	0.81	0.16	0.09	0.23
OS	MSE	103.68	154.79	107.47	3476.50	314.21	816.27	100.72	108.81	134.85	279.98	4042.73	302.03	804.81	103.43
OS	MC SE	0.23	0.14	0.23	0.86	0.13	0.09	0.23	3.23	4.23	109.07	148.44	6.91	7.01	3.47
OS	Rel P	0.00%	157.54%	0.48%	-92.94%	203.32%	559.73%	-0.91%	0.00%	133.79%	-61.60%	-91.44%	117.62%	603.49%	3.96%
OS	ModelSE	12.15	10.01	22.01	48.90	10.53	4.41	11.96	12.15	10.01	22.01	48.90	10.53	4.41	11.96
OS	Cover2	0.91	0.68	0.90	0.87	0.41	0.00	0.93	0.92	0.71	0.92	0.89	0.45	0.00	0.95
OS	Cover1	1.00	0.97	1.00	0.97	0.84	0.18	1.00	1.00	0.97	1.00	0.97	0.84	0.18	1.00
QALYs	Mean	60.34	56.79	60.63	84.98	51.20	44.09	59.47	61.01	57.77	62.59	88.84	51.58	44.25	60.35
QALYs	Bias	-3.69	-7.24	-3.40	20.96	-12.83	-19.94	-4.56	-3.02	-6.26	-1.44	24.81	-12.45	-19.78	-3.68
QALYs	MC SE	0.20	0.16	0.22	0.63	0.13	0.10	0.19	0.21	0.17	0.31	0.61	0.15	0.10	0.20
QALYs	Rel Bias	- 0.06	- 0.11	- 0.05	0.33	- 0.20	- 0.31	- 0.07	- 0.05	- 0.10	- 0.02	0.39	- 0.19	- 0.31	- 0.06
QALYs	empSE	6.06	4.94	6.70	19.52	3.96	3.04	5.87	6.35	5.18	9.60	18.77	4.50	3.08	6.02
QALYs	MC SE	0.14	0.11	0.15	0.45	0.09	0.07	0.13	0.15	0.12	0.22	0.43	0.10	0.07	0.14
QALYs	MSE	50.33	76.81	56.43	819.97	180.23	406.97	55.23	49.32	66.01	94.24	967.17	175.22	400.82	49.80
QALYs	MC SE	1.69	2.14	1.86	37.17	3.24	3.86	1.97	1.61	2.02	26.28	37.72	3.56	3.88	1.78
QALYs	Rel P	0.00%	50.84%	-18.14%	-90.35%	134.31%	296.95%	6.78%	0.00%	50.08%	-56.36%	-88.57%	99.24%	323.45%	10.92%
QALYs	ModelSE	6.62	6.85	12.09	25.28	6.83	3.46	7.47	6.62	6.85	12.09	25.28	6.83	3.46	7.47
QALYs	Cover2	0.81	0.72	0.87	0.89	0.33	0.00	0.86	0.82	0.77	0.90	0.91	0.35	0.00	0.87
QALYs	Cover1	0.99	0.97	1.00	0.98	0.81	0.26	1.00	0.99	0.97	1.00	0.98	0.81	0.26	1.00

Table 62 : Result for simulated scenario 40

	Scenario40	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	57.09	57.30	57.19	57.23	49.95	42.87	51.78	57.80	58.19	58.08	58.12	49.47	43.06	52.13
PFS	Bias	1.00	1.21	1.10	1.14	-6.14	-13.22	-4.31	1.71	2.10	1.99	2.03	-6.62	-13.03	-3.96
PFS	MC SE	0.21	0.21	0.21	0.21	0.17	0.10	0.18	0.21	0.22	0.22	0.22	0.17	0.11	0.18
PFS	Rel Bias	1.8%	2.2%	2.0%	2.0%	-10.9%	-23.6%	-7.7%	3.0%	3.7%	3.5%	3.6%	-11.8%	-23.2%	-7.1%
PFS	empSE	6.36	6.49	6.45	6.45	5.14	3.22	5.49	6.51	6.76	6.73	6.72	5.25	3.27	5.56
PFS	MC SE	0.15	0.15	0.15	0.15	0.12	0.07	0.13	0.15	0.16	0.15	0.15	0.12	0.08	0.13
PFS	MSE	41.40	43.48	42.80	42.85	64.05	185.08	48.72	45.29	50.08	49.16	49.28	71.33	180.47	46.63
PFS	MC SE	1.80	1.91	1.89	1.88	1.89	2.67	1.80	2.02	2.29	2.25	2.24	2.30	2.66	1.79
PFS	Rel P	0.00%	-3.85%	-2.85%	-2.79%	53.34%	288.97%	34.15%	0.00%	-7.20%	-6.23%	-6.16%	53.84%	296.16%	37.07%
PFS	ModelSE	6.57	6.80	6.77	6.78	6.20	3.37	5.66	6.57	6.80	6.77	6.78	6.20	3.37	5.66
PFS	Cover2	0.94	0.95	0.94	0.94	0.64	0.07	0.79	0.94	0.95	0.95	0.95	0.63	0.08	0.80
PFS	Cover1	1.00	1.00	1.00	1.00	0.94	0.56	0.98	1.00	1.00	1.00	1.00	0.94	0.56	0.98
OS	Mean	199.90	182.84	216.20	308.38	180.05	83.83	204.17	206.77	193.33	241.28	323.85	192.79	85.06	212.21
OS	Bias	9.53	-7.53	25.83	118.01	-10.32	-106.55	13.79	16.40	2.96	50.91	133.48	2.42	-105.31	21.84
OS	MC SE	0.70	0.95	1.79	3.73	0.93	0.23	0.75	1.21	1.18	2.18	2.99	1.24	0.24	0.91
OS	Rel Bias	0.05	-0.04	0.14	0.62	-0.05	-0.56	0.07	0.09	0.02	0.27	0.70	0.01	-0.55	0.11
OS	empSE	21.54	29.42	55.09	115.06	28.71	7.02	23.26	37.31	36.37	67.05	92.10	38.17	7.44	27.91
OS	MC SE	0.49	0.68	1.26	2.64	0.66	0.16	0.53	0.86	0.83	1.54	2.11	0.88	0.17	0.64
OS	MSE	554.27	921.27	3698.19	27151.56	930.00	11401.12	730.67	1659.49	1330.29	7082.27	26290.82	1461.29	11145.42	1255.28
OS	MC SE	0.49	0.68	1.26	2.64	0.66	0.16	0.53	402.78	288.97	555.68	885.48	296.74	50.43	77.77
OS	Rel P	0.00%	-46.40%	-84.71%	-96.50%	-43.73%	842.09%	-14.25%	0.00%	5.22%	-69.03%	-83.59%	-4.46%	2416.47%	78.67%
OS	ModelSE	51.16	65.43	104.08	129.96	70.20	9.03	51.60	51.16	65.43	104.08	129.96	70.20	9.03	51.60
OS	Cover2	1.00	0.99	0.99	0.70	0.98	0.00	1.00	1.00	0.99	0.99	0.75	0.99	0.00	1.00
OS	Cover1	1.00	0.99	1.00	0.98	1.00	0.07	1.00	1.00	0.99	1.00	0.98	1.00	0.07	1.00
QALYs	Mean	117.08	108.61	125.26	171.36	105.01	54.77	117.62	120.72	114.12	138.06	179.36	111.24	55.45	121.74
QALYs	Bias	5.07	-3.40	13.24	59.35	-7.00	-57.24	5.60	8.71	2.11	26.05	67.35	-0.78	-56.56	9.73
QALYs	MC SE	0.36	0.48	0.90	1.87	0.47	0.13	0.39	0.61	0.59	1.09	1.50	0.62	0.14	0.46
QALYs	Rel Bias	0.05	-0.03	0.12	0.53	-0.06	-0.51	0.05	0.08	0.02	0.23	0.60	-0.01	-0.50	0.09
QALYs	empSE	11.05	14.73	27.63	57.61	14.47	4.08	11.97	18.77	18.16	33.53	46.12	19.20	4.29	14.21
QALYs	MC SE	0.25	0.34	0.63	1.32	0.33	0.09	0.27	0.43	0.42	0.77	1.06	0.44	0.10	0.33
QALYs	MSE	147.64	228.27	937.96	6837.54	258.30	3292.80	174.54	427.68	333.98	1801.81	6661.17	368.69	3217.83	296.33
QALYs	MC SE	5.41	7.29	57.95	258.26	9.04	15.07	7.21	100.01	71.45	139.48	223.30	72.30	15.63	18.42
QALYs	Rel P	0.00%	-43.71%	-84.01%	-96.32%	-41.72%	633.45%	-14.79%	0.00%	6.76%	-68.68%	-83.45%	-4.43%	1815.32%	74.47%
QALYs	ModelSE	25.76	32.88	52.21	65.13	35.47	5.07	26.30	25.76	32.88	52.21	65.13	35.47	5.07	26.30
QALYs	Cover2	1.00	0.99	0.99	0.70	0.98	0.00	1.00	1.00	1.00	1.00	0.76	0.99	0.00	1.00
QALYs	Cover1	1.00	0.99	1.00	0.98	1.00	0.07	1.00	1.00	0.99	1.00	0.98	1.00	0.07	1.00

Table 63 : Result for simulated scenario 41

	Scenario41	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	54.72	57.64	57.20	57.29	49.28	44.28	50.35	54.98	58.56	58.10	58.20	49.00	44.48	50.69
PFS	Bias	-1.24	1.67	1.24	1.33	-6.69	-11.68	-5.62	-0.99	2.59	2.14	2.24	-6.96	-11.49	-5.27
PFS	MC SE	0.16	0.22	0.22	0.22	0.16	0.11	0.16	0.16	0.23	0.23	0.23	0.17	0.11	0.17
PFS	Rel Bias	-2.2%	3.0%	2.2%	2.4%	-11.9%	-20.9%	-10.0%	-1.8%	4.6%	3.8%	4.0%	-12.4%	-20.5%	-9.4%
PFS	empSE	5.06	6.68	6.74	6.73	5.08	3.31	5.02	5.06	6.99	7.06	7.04	5.16	3.36	5.10
PFS	MC SE	0.12	0.15	0.15	0.15	0.12	0.08	0.12	0.12	0.16	0.16	0.16	0.12	0.08	0.12
PFS	MSE	27.16	47.37	46.92	46.98	70.50	147.39	56.72	26.53	55.59	54.35	54.53	75.01	143.22	53.83
PFS	MC SE	1.20	2.02	1.98	2.00	2.02	2.47	2.10	1.22	2.44	2.39	2.41	2.20	2.46	2.06
PFS	Rel P	0.00%	-42.53%	-43.55%	-43.34%	-0.64%	133.83%	1.69%	0.00%	-47.72%	-48.67%	-48.42%	-3.82%	126.30%	-1.85%
PFS	ModelSE	5.50	7.02	6.92	6.94	5.83	3.74	5.52	5.50	7.02	6.92	6.94	5.83	3.74	5.52
PFS	Cover2	0.92	0.94	0.94	0.94	0.61	0.17	0.75	0.92	0.95	0.94	0.94	0.60	0.18	0.77
PFS	Cover1	0.99	0.99	0.99	0.99	0.93	0.72	0.97	0.99	0.99	0.99	0.99	0.93	0.72	0.97
OS	Mean	97.30	94.40	104.37	175.41	95.56	62.24	99.94	101.43	98.46	114.91	188.05	100.39	62.60	103.78
OS	Bias	-0.20	-3.10	6.87	77.91	-1.95	-35.26	2.44	3.93	0.95	17.41	90.55	2.89	-34.90	6.27
OS	MC SE	0.39	0.36	0.45	1.88	0.41	0.13	0.38	0.85	0.58	0.84	1.71	0.56	0.13	0.44
OS	Rel Bias	- 0.00	- 0.03	0.07	0.80	- 0.02	- 0.36	0.03	0.04	0.01	0.18	0.93	0.03	- 0.36	0.06
OS	empSE	12.05	11.24	13.84	58.00	12.60	3.90	11.81	26.19	17.73	25.79	52.85	17.37	3.98	13.67
OS	MC SE	0.28	0.26	0.32	1.33	0.29	0.09	0.27	0.60	0.41	0.59	1.21	0.40	0.09	0.31
OS	MSE	145.17	135.73	238.63	9429.86	162.48	1258.27	145.38	700.84	314.96	967.61	10989.35	309.65	1233.92	226.01
OS	MC SE	0.28	0.26	0.32	1.33	0.29	0.09	0.27	242.60	77.91	88.56	371.16	29.91	9.01	17.47
OS	Rel P	0.00%	15.10%	-24.18%	-95.68%	-8.55%	857.00%	4.09%	0.00%	118.25%	3.16%	-75.43%	127.49%	4235.03%	267.22%
OS	ModelSE	32.54	25.16	45.48	78.85	26.73	4.67	24.04	32.54	25.16	45.48	78.85	26.73	4.67	24.04
OS	Cover2	1.00	0.96	0.99	0.77	0.95	0.00	1.00	1.00	0.97	0.99	0.81	0.96	0.00	1.00
OS	Cover1	1.00	1.00	1.00	0.97	1.00	0.07	1.00	1.00	1.00	1.00	0.97	1.00	0.07	1.00
QALYs	Mean	65.07	64.49	69.35	104.89	62.56	44.41	65.08	67.21	66.79	74.89	111.49	64.90	44.64	67.09
QALYs	Bias	-0.47	-1.05	3.81	39.35	-2.98	-21.13	-0.46	1.67	1.25	9.35	45.95	-0.64	-20.90	1.55
QALYs	MC SE	0.21	0.18	0.25	0.95	0.21	0.09	0.21	0.43	0.29	0.44	0.87	0.29	0.09	0.24
QALYs	Rel Bias	- 0.01	- 0.02	0.06	0.60	- 0.05	- 0.32	- 0.01	0.03	0.02	0.14	0.70	- 0.01	- 0.32	0.02
QALYs	empSE	6.42	5.66	7.67	29.42	6.46	2.80	6.48	13.36	8.89	13.47	26.86	8.79	2.85	7.28
QALYs	MC SE	0.15	0.13	0.18	0.68	0.15	0.06	0.15	0.31	0.20	0.31	0.62	0.20	0.07	0.17
QALYs	MSE	41.44	33.07	73.30	2413.19	50.54	454.43	42.23	181.12	80.45	268.70	2831.76	77.66	444.80	55.43
QALYs	MC SE	1.50	1.38	3.18	91.78	1.91	3.83	1.55	61.58	20.56	23.62	95.39	7.33	3.86	3.66
QALYs	Rel P	0.00%	28.93%	-29.90%	-95.23%	-1.07%	426.91%	-1.88%	0.00%	126.09%	-1.64%	-75.26%	130.86%	2093.29%	236.42%
QALYs	ModelSE	16.60	13.22	23.15	39.77	13.93	3.32	12.84	16.60	13.22	23.15	39.77	13.93	3.32	12.84
QALYs	Cover2	0.99	0.98	0.99	0.78	0.94	0.00	0.99	1.00	0.98	1.00	0.81	0.96	0.00	1.00
QALYs	Cover1	1.00	1.00	1.00	0.97	0.99	0.18	1.00	1.00	1.00	1.00	0.97	0.99	0.18	1.00

Table 64 : Result for simulated scenario 42

	Scenario42	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	52.77	56.99	56.96	56.94	48.45	45.37	49.21	52.83	57.86	57.82	57.81	48.30	45.54	49.65
PFS	Bias	-3.15	1.07	1.04	1.02	-7.47	-10.55	-6.72	-3.10	1.94	1.90	1.88	-7.62	-10.38	-6.27
PFS	MC SE	0.13	0.21	0.21	0.21	0.16	0.11	0.13	0.13	0.22	0.22	0.22	0.16	0.11	0.13
PFS	Rel Bias	-5.6%	1.9%	1.9%	1.8%	-13.4%	-18.9%	-12.0%	-5.5%	3.5%	3.4%	3.4%	-13.6%	-18.6%	-11.2%
PFS	empSE	3.99	6.45	6.50	6.57	4.84	3.41	4.07	4.01	6.79	6.84	6.91	4.82	3.46	4.13
PFS	MC SE	0.09	0.15	0.15	0.15	0.11	0.08	0.09	0.09	0.16	0.16	0.16	0.11	0.08	0.09
PFS	MSE	25.86	42.76	43.28	44.22	79.22	122.97	61.68	25.68	49.78	50.28	51.21	81.29	119.78	56.32
PFS	MC SE	1.02	1.65	1.75	1.78	2.07	2.29	1.86	0.98	2.00	2.11	2.14	2.10	2.29	1.80
PFS	Rel P	0.00%	-61.69%	-62.23%	-63.08%	-31.84%	37.19%	-3.82%	0.00%	-65.02%	-65.51%	-66.22%	-30.53%	34.54%	-5.29%
PFS	ModelSE	4.80	6.76	6.75	6.76	5.08	3.87	4.99	4.80	6.76	6.75	6.76	5.08	3.87	4.99
PFS	Cover2	0.88	0.93	0.93	0.93	0.52	0.26	0.70	0.88	0.93	0.94	0.93	0.51	0.27	0.74
PFS	Cover1	1.00	1.00	1.00	1.00	0.91	0.81	0.97	1.00	1.00	1.00	1.00	0.91	0.81	0.97
OS	Mean	70.92	69.04	75.46	112.62	64.62	53.74	71.78	71.72	71.39	78.66	120.35	66.43	53.94	73.09
OS	Bias	-2.07	-3.94	2.48	39.63	-8.36	-19.24	-1.21	-1.26	-1.60	5.67	47.37	-6.55	-19.04	0.11
OS	MC SE	0.28	0.18	0.26	0.98	0.18	0.12	0.28	0.29	0.52	0.50	0.91	0.35	0.12	0.30
OS	Rel Bias	- 0.03	- 0.05	0.03	0.54	- 0.11	- 0.26	- 0.02	- 0.02	- 0.02	0.08	0.65	- 0.09	- 0.26	0.00
OS	empSE	8.60	5.49	8.01	30.24	5.68	3.57	8.78	9.06	15.95	15.52	28.19	10.85	3.62	9.14
OS	MC SE	0.20	0.13	0.18	0.69	0.13	0.08	0.20	0.21	0.37	0.36	0.65	0.25	0.08	0.21
OS	MSE	78.22	45.64	70.18	2483.92	102.17	383.05	78.40	83.66	256.66	272.85	3037.35	160.48	375.62	83.49
OS	MC SE	0.20	0.13	0.18	0.69	0.13	0.08	0.20	2.42	99.55	86.49	104.42	52.97	4.40	3.49
OS	Rel P	0.00%	145.54%	15.44%	-91.90%	129.04%	481.49%	-3.90%	0.00%	-67.71%	-65.90%	-89.66%	-30.18%	526.11%	-1.70%
OS	ModelSE	9.59	21.23	21.66	44.93	14.27	4.05	10.09	9.59	21.23	21.66	44.93	14.27	4.05	10.09
OS	Cover2	0.92	0.89	0.98	0.89	0.72	0.00	0.94	0.93	0.91	0.99	0.90	0.75	0.01	0.97
OS	Cover1	1.00	0.99	1.00	0.97	0.95	0.43	1.00	1.00	0.99	1.00	0.97	0.95	0.43	1.00
QALYs	Mean	51.29	51.62	54.82	73.39	46.85	40.48	50.65	51.71	53.05	56.67	77.52	47.70	40.63	51.44
QALYs	Bias	-1.98	-1.65	1.55	20.12	-6.42	-12.79	-2.62	-1.56	-0.22	3.41	24.25	-5.56	-12.63	-1.83
QALYs	MC SE	0.16	0.14	0.18	0.51	0.12	0.09	0.16	0.17	0.28	0.28	0.48	0.19	0.09	0.16
QALYs	Rel Bias	- 0.04	- 0.03	0.03	0.38	- 0.12	- 0.24	- 0.05	- 0.03	- 0.00	0.06	0.46	- 0.10	- 0.24	- 0.03
QALYs	empSE	4.95	4.32	5.45	15.70	3.62	2.77	4.91	5.25	8.71	8.73	14.71	5.83	2.82	5.08
QALYs	MC SE	0.11	0.10	0.13	0.36	0.08	0.06	0.11	0.12	0.20	0.20	0.34	0.13	0.06	0.12
QALYs	MSE	28.37	21.40	32.04	651.02	54.34	171.21	30.91	29.98	75.84	87.69	804.19	64.86	167.56	29.14
QALYs	MC SE	0.80	0.73	1.22	27.78	1.47	2.26	0.97	0.87	25.60	21.94	27.58	13.00	2.27	0.97
QALYs	Rel P	0.00%	30.88%	-17.48%	-90.06%	86.64%	218.17%	1.68%	0.00%	-63.65%	-63.79%	-87.26%	-18.74%	247.72%	6.76%
QALYs	ModelSE	5.29	11.38	11.63	23.03	7.82	3.15	6.20	5.29	11.38	11.63	23.03	7.82	3.15	6.20
QALYs	Cover2	0.90	0.94	0.98	0.90	0.70	0.05	0.91	0.90	0.95	0.98	0.91	0.73	0.05	0.93
QALYs	Cover1	1.00	1.00	1.00	0.98	0.96	0.53	1.00	1.00	1.00	1.00	0.98	0.96	0.53	1.00

Table 65 : Result for simulated scenario 43

	Scenario43	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	61.80	62.17	62.03	62.05	58.07	49.31	58.72	62.23	62.81	62.67	62.69	56.86	49.45	59.12
PFS	Bias	-7.40	-7.03	-7.18	-7.15	-11.13	-19.89	-10.48	-6.97	-6.39	-6.54	-6.51	-12.34	-19.75	-10.08
PFS	MC SE	0.20	0.20	0.20	0.20	0.18	0.10	0.18	0.20	0.21	0.21	0.21	0.22	0.10	0.18
PFS	Rel Bias	-10.7%	-10.2%	-10.4%	-10.3%	-16.1%	-28.7%	-15.1%	-10.1%	-9.2%	-9.4%	-9.4%	-17.8%	-28.5%	-14.6%
PFS	empSE	6.03	6.28	6.28	6.23	5.66	3.14	5.51	6.10	6.50	6.51	6.45	6.83	3.17	5.67
PFS	MC SE	0.14	0.14	0.14	0.14	0.13	0.07	0.13	0.14	0.15	0.15	0.15	0.16	0.07	0.13
PFS	MSE	91.13	88.84	90.85	89.86	155.97	405.55	140.24	85.79	83.12	84.99	84.01	198.80	400.02	133.70
PFS	MC SE	2.90	2.91	2.97	2.90	3.53	3.99	3.47	2.84	2.92	2.98	2.91	5.78	3.99	3.43
PFS	Rel P	0.00%	-7.66%	-7.67%	-6.19%	13.62%	268.43%	19.70%	0.00%	-11.86%	-11.98%	-10.57%	-20.04%	270.14%	15.90%
PFS	ModelSE	5.63	5.89	5.86	5.86	8.70	3.16	5.41	5.63	5.89	5.86	5.86	8.70	3.16	5.41
PFS	Cover2	0.63	0.65	0.65	0.65	0.43	0.00	0.43	0.65	0.69	0.68	0.68	0.44	0.00	0.46
PFS	Cover1	0.93	0.93	0.93	0.93	0.79	0.12	0.89	0.93	0.93	0.93	0.93	0.79	0.12	0.89
OS	Mean	199.20	192.18	218.63	329.92	187.96	105.49	203.18	203.88	196.59	235.41	339.37	192.38	106.09	210.69
OS	Bias	5.62	-1.41	25.04	136.34	-5.63	-88.10	9.60	10.30	3.00	41.82	145.78	-1.20	-87.49	17.10
OS	MC SE	0.73	0.82	1.22	2.91	0.80	0.21	0.70	0.83	0.90	1.62	2.52	0.99	0.22	0.72
OS	Rel Bias	0.03	-0.01	0.13	0.70	-0.03	0.46	0.05	0.05	0.02	0.22	0.75	-0.01	0.45	0.09
OS	empSE	22.64	25.38	37.59	89.82	24.68	6.59	21.49	25.48	27.70	50.05	77.65	30.45	6.73	22.26
OS	MC SE	0.52	0.58	0.86	2.06	0.57	0.15	0.49	0.58	0.64	1.15	1.78	0.70	0.15	0.51
OS	MSE	543.82	645.64	2038.39	26648.16	640.26	7805.15	553.50	754.50	775.24	4251.08	27274.91	927.39	7700.42	787.32
OS	MC SE	0.52	0.58	0.86	2.06	0.57	0.15	0.49	46.80	32.48	360.69	761.90	38.47	38.05	38.23
OS	Rel P	0.00%	-20.42%	-63.70%	-93.64%	-15.83%	1080.85%	11.02%	0.00%	-15.37%	-74.08%	-89.23%	-29.96%	1332.23%	31.02%
OS	ModelSE	43.58	46.22	81.46	99.33	51.03	7.65	44.81	43.58	46.22	81.46	99.33	51.03	7.65	44.81
OS	Cover2	1.00	0.96	0.99	0.54	0.95	0.00	1.00	1.00	0.97	0.99	0.57	0.96	0.00	1.00
OS	Cover1	1.00	1.00	1.00	0.88	1.00	0.01	1.00	1.00	1.00	1.00	0.88	1.00	0.01	1.00
QALYs	Mean	118.14	114.74	127.92	183.58	111.40	67.54	119.21	120.61	117.13	136.50	188.49	113.25	67.88	123.08
QALYs	Bias	0.59	-2.81	10.37	66.02	-6.15	-50.02	1.65	3.06	-0.42	18.95	70.94	-4.30	-49.67	5.53
QALYs	MC SE	0.37	0.42	0.62	1.47	0.41	0.12	0.36	0.42	0.46	0.82	1.27	0.51	0.13	0.37
QALYs	Rel Bias	0.01	-0.02	0.09	0.56	-0.05	0.43	0.01	0.03	-0.00	0.16	0.60	-0.04	0.42	0.05
QALYs	empSE	11.51	12.93	19.04	45.23	12.54	3.85	10.96	12.96	14.12	25.35	39.19	15.87	3.92	11.39
QALYs	MC SE	0.26	0.30	0.44	1.04	0.29	0.09	0.25	0.30	0.32	0.58	0.90	0.36	0.09	0.26
QALYs	MSE	132.79	174.89	469.50	6403.12	194.99	2516.59	122.83	177.17	199.39	1001.12	6565.97	270.12	2482.60	160.09
QALYs	MC SE	4.04	5.59	25.07	199.43	6.67	12.46	4.15	10.95	8.14	88.00	187.51	11.71	12.61	8.01
QALYs	Rel P	0.00%	-20.68%	-63.41%	-93.52%	-15.71%	795.17%	10.28%	0.00%	-15.76%	-73.86%	-89.06%	-33.29%	990.80%	29.54%
QALYs	ModelSE	22.00	23.37	40.98	49.97	26.71	4.41	22.99	22.00	23.37	40.98	49.97	26.71	4.41	22.99
QALYs	Cover2	1.00	0.94	0.99	0.56	0.94	0.00	1.00	1.00	0.96	0.99	0.58	0.95	0.00	1.00
QALYs	Cover1	1.00	1.00	1.00	0.89	1.00	0.01	1.00	1.00	1.00	1.00	0.89	1.00	0.01	1.00

Table 66 : Result for simulated scenario 44

	Scenario44	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	61.21	62.71	62.48	62.42	57.43	49.98	57.57	61.58	63.37	63.14	63.09	56.49	50.12	57.93
PFS	Bias	-7.89	-6.39	-6.62	-6.68	-11.67	-19.12	-11.53	-7.52	-5.73	-5.96	-6.01	-12.61	-18.98	-11.17
PFS	MC SE	0.17	0.21	0.20	0.20	0.19	0.10	0.15	0.17	0.21	0.21	0.21	0.21	0.10	0.16
PFS	Rel Bias	-11.4%	-9.3%	-9.6%	-9.7%	-16.9%	-27.7%	-16.7%	-10.9%	-8.3%	-8.6%	-8.7%	-18.2%	-27.5%	-16.2%
PFS	empSE	5.33	6.42	6.31	6.27	5.72	3.12	4.74	5.33	6.62	6.51	6.47	6.57	3.15	4.80
PFS	MC SE	0.12	0.15	0.14	0.14	0.13	0.07	0.11	0.12	0.15	0.15	0.15	0.15	0.07	0.11
PFS	MSE	90.56	82.05	83.63	83.81	168.98	375.18	155.33	84.89	76.66	77.85	77.99	202.09	370.09	147.80
PFS	MC SE	2.77	2.81	2.82	2.82	3.78	3.81	3.48	2.67	2.76	2.76	2.76	5.15	3.82	3.42
PFS	Rel P	0.00%	-31.20%	-28.83%	-27.76%	-13.42%	191.52%	26.15%	0.00%	-35.19%	-32.95%	-32.06%	-34.04%	187.24%	23.33%
PFS	ModelSE	5.15	5.92	5.88	5.87	8.37	3.28	4.94	5.15	5.92	5.88	5.87	8.37	3.28	4.94
PFS	Cover2	0.60	0.69	0.68	0.68	0.40	0.00	0.34	0.62	0.72	0.71	0.71	0.40	0.00	0.36
PFS	Cover1	0.91	0.94	0.94	0.94	0.77	0.17	0.87	0.91	0.94	0.94	0.94	0.77	0.17	0.87
OS	Mean	126.48	122.91	134.18	216.81	122.01	80.63	128.34	127.81	125.67	140.17	222.95	124.34	80.89	130.58
OS	Bias	1.29	-2.28	8.99	91.62	-3.18	-44.56	3.15	2.62	0.48	14.98	97.76	-0.85	-44.30	5.39
OS	MC SE	0.51	0.43	0.52	1.74	0.42	0.13	0.50	0.53	0.52	0.89	1.57	0.56	0.13	0.51
OS	Rel Bias	0.01	-0.02	0.07	0.73	-0.03	0.36	0.03	0.02	0.00	0.12	0.78	-0.01	0.35	0.04
OS	empSE	15.71	13.21	16.03	53.66	12.86	3.88	15.50	16.29	16.03	27.34	48.48	17.16	3.93	15.69
OS	MC SE	0.36	0.30	0.37	1.23	0.30	0.09	0.36	0.37	0.37	0.63	1.11	0.39	0.09	0.36
OS	MSE	248.06	179.54	337.62	11271.40	175.21	2000.55	249.88	271.88	257.02	971.23	11904.81	294.87	1977.78	274.87
OS	MC SE	0.36	0.30	0.37	1.23	0.30	0.09	0.36	7.84	34.21	271.55	325.24	33.44	11.24	9.21
OS	Rel P	0.00%	41.34%	-4.02%	-91.43%	49.22%	1536.30%	2.68%	0.00%	3.20%	-64.52%	-88.71%	-9.91%	1617.22%	7.81%
OS	ModelSE	17.03	24.98	36.62	62.75	27.55	4.68	17.44	17.03	24.98	36.62	62.75	27.55	4.68	17.44
OS	Cover2	0.97	0.88	0.99	0.59	0.94	0.00	0.98	0.98	0.90	0.99	0.61	0.95	0.00	0.99
OS	Cover1	1.00	1.00	1.00	0.87	1.00	0.00	1.00	1.00	1.00	1.00	0.87	1.00	0.00	1.00
QALYs	Mean	81.60	80.26	85.84	127.13	78.23	55.31	81.44	82.38	81.85	89.03	130.40	79.12	55.48	82.67
QALYs	Bias	-1.72	-3.06	2.51	43.81	-5.09	-28.01	-1.88	-0.94	-1.48	5.70	47.08	-4.21	-27.84	-0.66
QALYs	MC SE	0.26	0.21	0.28	0.89	0.22	0.09	0.26	0.27	0.26	0.46	0.81	0.30	0.09	0.26
QALYs	Rel Bias	-0.02	-0.04	0.03	0.53	-0.06	0.34	-0.02	-0.01	-0.02	0.07	0.56	-0.05	0.33	-0.01
QALYs	empSE	8.06	6.57	8.69	27.37	6.71	2.70	7.97	8.39	8.06	14.16	24.87	9.21	2.73	8.07
QALYs	MC SE	0.18	0.15	0.20	0.63	0.15	0.06	0.18	0.19	0.19	0.32	0.57	0.21	0.06	0.19
QALYs	MSE	67.84	52.50	81.77	2667.52	70.82	792.07	67.03	71.25	67.15	232.78	2834.16	102.41	782.65	65.55
QALYs	MC SE	1.88	1.75	3.21	86.84	2.53	4.86	2.01	1.94	9.11	66.17	81.25	9.95	4.89	1.92
QALYs	Rel P	0.00%	50.37%	-14.03%	-91.33%	44.43%	794.03%	2.20%	0.00%	8.30%	-64.86%	-88.61%	-16.93%	847.63%	8.06%
QALYs	ModelSE	8.80	12.99	18.79	31.85	15.21	3.15	9.62	8.80	12.99	18.79	31.85	15.21	3.15	9.62
QALYs	Cover2	0.91	0.94	0.98	0.63	0.89	0.00	0.94	0.93	0.95	0.98	0.65	0.92	0.00	0.96
QALYs	Cover1	1.00	1.00	1.00	0.89	1.00	0.01	1.00	1.00	1.00	1.00	0.89	1.00	0.01	1.00

Table 67 : Result for simulated scenario 45

	Scenario45	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	61.16	63.02	63.16	62.97	56.43	50.48	57.34	61.54	63.72	63.85	63.66	56.35	50.59	57.57
PFS	Bias	-7.89	-6.03	-5.90	-6.08	-12.62	-18.57	-11.71	-7.51	-5.34	-5.20	-5.39	-12.70	-18.46	-11.48
PFS	MC SE	0.16	0.20	0.21	0.21	0.17	0.10	0.14	0.16	0.21	0.21	0.21	0.18	0.10	0.14
PFS	Rel Bias	-11.4%	-8.7%	-8.5%	-8.8%	-18.3%	-26.9%	-17.0%	-10.9%	-7.7%	-7.5%	-7.8%	-18.4%	-26.7%	-16.6%
PFS	empSE	4.86	6.22	6.36	6.37	5.38	3.06	4.31	4.90	6.41	6.55	6.55	5.64	3.08	4.34
PFS	MC SE	0.11	0.14	0.15	0.15	0.12	0.07	0.10	0.11	0.15	0.15	0.15	0.13	0.07	0.10
PFS	MSE	85.79	74.97	75.17	77.55	188.29	354.22	155.72	80.34	69.46	69.86	71.95	193.08	350.45	150.62
PFS	MC SE	2.61	2.67	2.56	2.68	3.97	3.66	3.36	2.52	2.56	2.47	2.58	4.34	3.67	3.32
PFS	Rel P	0.00%	-38.96%	-41.64%	-41.87%	-18.49%	151.94%	27.18%	0.00%	-41.51%	-44.01%	-44.10%	-24.50%	152.42%	27.68%
PFS	ModelSE	5.15	5.94	5.97	5.91	6.47	3.40	4.85	5.15	5.94	5.97	5.91	6.47	3.40	4.85
PFS	Cover2	0.60	0.72	0.71	0.71	0.32	0.00	0.34	0.62	0.75	0.75	0.74	0.32	0.00	0.35
PFS	Cover1	0.93	0.95	0.95	0.95	0.71	0.19	0.86	0.93	0.95	0.95	0.95	0.71	0.19	0.86
OS	Mean	92.58	83.15	92.29	141.86	80.55	64.90	94.36	93.44	84.57	94.16	145.94	82.11	65.05	95.33
OS	Bias	-0.96	-10.39	-1.25	48.32	-12.99	-28.64	0.82	-0.10	-8.97	0.61	52.40	-11.43	-28.49	1.79
OS	MC SE	0.33	0.17	0.31	0.92	0.17	0.10	0.33	0.34	0.37	0.37	0.88	0.41	0.11	0.34
OS	Rel Bias	-0.01	-0.11	0.01	0.52	-0.14	0.31	0.01	-0.00	0.10	0.01	0.56	-0.12	0.30	0.02
OS	empSE	10.20	5.28	9.45	28.31	5.26	3.23	10.19	10.52	11.28	11.47	27.16	12.53	3.26	10.36
OS	MC SE	0.23	0.12	0.22	0.65	0.12	0.07	0.23	0.24	0.26	0.26	0.62	0.29	0.07	0.24
OS	MSE	104.82	135.81	90.72	3134.90	196.38	830.57	104.33	110.66	207.54	131.86	3482.80	287.47	822.20	110.43
OS	MC SE	0.23	0.12	0.22	0.65	0.12	0.07	0.23	3.57	67.87	26.68	105.04	81.73	5.99	3.80
OS	Rel P	0.00%	272.63%	16.55%	-87.02%	276.62%	895.54%	0.23%	0.00%	-12.95%	-15.84%	-84.98%	-29.40%	941.28%	3.20%
OS	ModelSE	10.29	15.72	14.99	37.16	17.30	3.73	10.37	10.29	15.72	14.99	37.16	17.30	3.73	10.37
OS	Cover2	0.85	0.59	0.92	0.78	0.54	0.00	0.89	0.86	0.63	0.93	0.79	0.58	0.00	0.90
OS	Cover1	1.00	1.00	1.00	0.92	0.96	0.03	1.00	1.00	1.00	1.00	0.92	0.96	0.03	1.00
QALYs	Mean	64.64	60.48	65.09	89.82	57.20	47.60	64.38	65.18	61.40	66.23	92.07	57.96	47.70	64.94
QALYs	Bias	-2.85	-7.00	-2.40	22.33	-10.28	-19.89	-3.10	-2.30	-6.08	-1.25	24.58	-9.53	-19.78	-2.55
QALYs	MC SE	0.19	0.12	0.20	0.48	0.11	0.08	0.19	0.20	0.21	0.23	0.47	0.22	0.08	0.19
QALYs	Rel Bias	-0.04	-0.10	0.04	0.33	-0.15	-0.29	0.05	-0.03	-0.09	0.02	0.36	-0.14	-0.29	0.04
QALYs	empSE	5.89	3.82	6.12	14.90	3.35	2.48	5.72	6.12	6.38	7.02	14.48	6.72	2.50	5.81
QALYs	MC SE	0.14	0.09	0.14	0.34	0.08	0.06	0.13	0.14	0.15	0.16	0.33	0.15	0.06	0.13
QALYs	MSE	42.80	63.59	43.16	720.66	116.94	401.77	42.34	42.66	77.72	50.83	813.91	135.93	397.65	40.25
QALYs	MC SE	1.53	1.59	1.56	26.54	2.16	3.18	1.57	1.50	17.38	6.45	26.84	20.05	3.19	1.48
QALYs	Rel P	0.00%	138.53%	-7.26%	-84.36%	210.02%	463.92%	6.08%	0.00%	-8.21%	-24.17%	-82.17%	-17.28%	496.58%	10.67%
QALYs	ModelSE	5.49	8.67	8.48	19.28	9.64	2.84	6.34	5.49	8.67	8.48	19.28	9.64	2.84	6.34
QALYs	Cover2	0.79	0.65	0.87	0.83	0.41	0.00	0.84	0.80	0.70	0.89	0.83	0.46	0.00	0.86
QALYs	Cover1	0.95	0.99	1.00	0.96	0.93	0.06	0.99	0.95	0.99	1.00	0.96	0.93	0.06	0.99

Table 68 : Result for simulated scenario 46

	Scenario46	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	57.32	57.51	57.38	57.35	50.76	44.42	52.46	57.85	58.16	58.03	58.00	50.54	44.54	52.73
PFS	Bias	1.23	1.42	1.29	1.26	-5.33	-11.67	-3.63	1.76	2.07	1.94	1.91	-5.56	-11.55	-3.36
PFS	MC SE	0.18	0.19	0.19	0.18	0.15	0.09	0.14	0.19	0.19	0.19	0.19	0.15	0.09	0.15
PFS	Rel Bias	2.2%	2.5%	2.3%	2.3%	-9.5%	-20.8%	-6.5%	3.1%	3.7%	3.5%	3.4%	-9.9%	-20.6%	-6.0%
PFS	empSE	5.68	5.76	5.71	5.68	4.47	2.83	4.46	5.78	5.93	5.89	5.86	4.48	2.85	4.53
PFS	MC SE	0.13	0.13	0.13	0.13	0.10	0.06	0.10	0.13	0.14	0.14	0.13	0.10	0.07	0.10
PFS	MSE	33.71	35.10	34.28	33.87	48.37	144.20	33.10	36.49	39.40	38.38	37.92	50.92	141.60	31.78
PFS	MC SE	1.54	1.62	1.59	1.57	1.49	2.12	1.25	1.69	1.84	1.82	1.79	1.56	2.11	1.22
PFS	Rel P	0.00%	-2.69%	-1.26%	-0.23%	61.30%	302.84%	61.93%	0.00%	-4.91%	-3.55%	-2.57%	66.70%	310.57%	62.98%
PFS	ModelSE	5.26	5.40	5.38	5.37	4.76	2.84	4.53	5.26	5.40	5.38	5.37	4.76	2.84	4.53
PFS	Cover2	0.91	0.92	0.92	0.92	0.63	0.05	0.79	0.91	0.91	0.92	0.92	0.63	0.06	0.79
PFS	Cover1	1.00	1.00	1.00	1.00	0.93	0.47	0.98	1.00	1.00	1.00	1.00	0.93	0.47	0.98
OS	Mean	203.66	190.49	225.62	307.54	187.20	100.80	201.64	207.33	196.21	242.58	323.33	195.94	101.75	206.95
OS	Bias	7.19	-5.98	29.15	111.07	-9.27	-95.66	5.18	10.86	-0.26	46.11	126.86	-0.53	-94.72	10.48
OS	MC SE	0.68	0.85	1.44	2.99	0.81	0.24	0.69	0.90	0.92	1.73	2.58	0.95	0.25	0.78
OS	Rel Bias	0.04	-0.03	0.15	0.57	-0.05	0.49	0.03	0.06	-0.00	0.23	0.65	-0.00	0.48	0.05
OS	empSE	20.82	26.20	44.36	92.22	25.09	7.43	21.34	27.88	28.51	53.38	79.67	29.34	7.68	23.93
OS	MC SE	0.48	0.60	1.02	2.12	0.58	0.17	0.49	0.64	0.65	1.23	1.83	0.67	0.18	0.55
OS	MSE	484.89	721.38	2815.30	20833.27	714.94	9206.95	481.68	894.42	811.82	4972.81	22432.62	860.24	9031.54	681.86
OS	MC SE	0.48	0.60	1.02	2.12	0.58	0.17	0.49	247.98	46.47	373.84	785.44	43.70	46.98	47.93
OS	Rel P	0.00%	-36.81%	-77.96%	-94.90%	-31.14%	684.75%	-4.77%	0.00%	-4.34%	-72.72%	-87.75%	-9.70%	1218.45%	35.75%
OS	ModelSE	39.09	50.02	84.16	106.71	53.24	9.12	38.76	39.09	50.02	84.16	106.71	53.24	9.12	38.76
OS	Cover2	1.00	0.98	0.98	0.62	0.99	0.00	1.00	1.00	0.99	0.98	0.64	0.99	0.00	1.00
OS	Cover1	1.00	1.00	1.00	0.94	1.00	0.03	1.00	1.00	1.00	1.00	0.94	1.00	0.03	1.00
QALYs	Mean	119.03	112.50	130.03	170.98	108.83	63.73	116.56	121.02	115.55	138.70	179.06	113.13	64.23	119.29
QALYs	Bias	3.96	-2.57	14.96	55.91	-6.23	-51.33	1.50	5.96	0.49	23.63	64.00	-1.93	-50.83	4.23
QALYs	MC SE	0.34	0.43	0.72	1.50	0.41	0.13	0.35	0.46	0.47	0.87	1.30	0.48	0.14	0.39
QALYs	Rel Bias	0.03	-0.02	0.13	0.49	-0.05	0.45	0.01	0.05	0.00	0.21	0.56	-0.02	0.44	0.04
QALYs	empSE	10.60	13.24	22.31	46.18	12.69	4.13	10.87	14.08	14.39	26.78	39.92	14.81	4.26	12.11
QALYs	MC SE	0.24	0.30	0.51	1.06	0.29	0.09	0.25	0.32	0.33	0.61	0.92	0.34	0.10	0.28
QALYs	MSE	127.97	181.80	720.99	5257.03	199.76	2652.21	120.28	233.43	207.13	1274.98	5687.99	222.72	2601.56	164.35
QALYs	MC SE	4.88	5.66	39.76	184.76	6.52	13.73	4.31	62.94	11.66	94.02	197.81	10.53	14.00	11.38
QALYs	Rel P	0.00%	-35.93%	-77.42%	-94.73%	-30.24%	557.81%	-4.89%	0.00%	-4.31%	-72.36%	-87.56%	-9.60%	993.60%	35.19%
QALYs	ModelSE	19.68	25.15	42.23	53.49	26.86	4.99	19.81	19.68	25.15	42.23	53.49	26.86	4.99	19.81
QALYs	Cover2	1.00	0.98	0.98	0.62	0.98	0.00	1.00	1.00	0.99	0.98	0.65	0.99	0.00	1.00
QALYs	Cover1	1.00	1.00	1.00	0.94	1.00	0.03	1.00	1.00	1.00	1.00	0.94	1.00	0.03	1.00

Table 69 : Result for simulated scenario 47

	Scenario47	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	55.50	57.72	57.56	57.57	50.49	45.17	50.92	55.76	58.38	58.22	58.22	50.36	45.28	51.15
PFS	Bias	-0.48	1.74	1.58	1.59	-5.49	-10.81	-5.06	-0.23	2.39	2.24	2.24	-5.62	-10.70	-4.84
PFS	MC SE	0.15	0.19	0.19	0.19	0.15	0.09	0.13	0.15	0.20	0.20	0.20	0.15	0.09	0.13
PFS	Rel Bias	-0.9%	3.1%	2.8%	2.8%	-9.8%	-19.3%	-9.0%	-0.4%	4.3%	4.0%	4.0%	-10.0%	-19.1%	-8.6%
PFS	empSE	4.64	5.94	5.91	5.89	4.55	2.84	4.01	4.64	6.13	6.10	6.08	4.59	2.87	4.07
PFS	MC SE	0.11	0.14	0.14	0.14	0.10	0.07	0.09	0.11	0.14	0.14	0.14	0.11	0.07	0.09
PFS	MSE	21.70	38.29	37.42	37.19	50.81	124.94	41.67	21.58	43.27	42.17	41.93	52.69	122.68	39.93
PFS	MC SE	0.89	1.78	1.75	1.75	1.57	1.97	1.40	0.89	2.03	1.99	1.99	1.62	1.97	1.37
PFS	Rel P	0.00%	-39.14%	-38.50%	-38.08%	3.82%	166.71%	33.63%	0.00%	-42.66%	-42.09%	-41.68%	2.13%	161.49%	30.08%
PFS	ModelSE	4.45	5.45	5.42	5.43	4.52	2.96	4.22	4.45	5.45	5.42	5.43	4.52	2.96	4.22
PFS	Cover2	0.91	0.92	0.92	0.92	0.61	0.09	0.70	0.91	0.91	0.91	0.91	0.60	0.10	0.72
PFS	Cover1	1.00	1.00	1.00	1.00	0.92	0.57	0.97	1.00	1.00	1.00	1.00	0.92	0.57	0.97
OS	Mean	107.03	106.91	117.99	182.28	108.92	72.43	107.69	108.84	109.91	124.61	191.49	112.91	72.70	110.02
OS	Bias	-0.54	-0.67	10.42	74.71	1.35	-35.15	0.12	1.27	2.34	17.03	83.91	5.33	-34.87	2.44
OS	MC SE	0.43	0.39	0.46	1.58	0.41	0.12	0.41	0.55	0.55	0.69	1.50	0.50	0.13	0.45
OS	Rel Bias	- 0.01	- 0.01	0.10	0.69	0.01	- 0.33	0.00	0.01	0.02	0.16	0.78	0.05	- 0.32	0.02
OS	empSE	13.41	11.91	14.23	48.64	12.77	3.83	12.74	16.97	16.85	21.32	46.10	15.38	3.88	13.79
OS	MC SE	0.31	0.27	0.33	1.12	0.29	0.09	0.29	0.39	0.39	0.49	1.06	0.35	0.09	0.32
OS	MSE	179.80	142.19	310.85	7944.05	164.67	1249.95	162.16	289.24	289.01	743.96	9164.20	264.78	1231.02	195.90
OS	MC SE	0.31	0.27	0.33	1.12	0.29	0.09	0.29	68.75	69.85	77.65	294.45	13.45	8.77	10.43
OS	Rel P	0.00%	26.64%	-11.29%	-92.40%	10.23%	1122.84%	10.71%	0.00%	1.45%	-36.63%	-86.45%	21.71%	1812.44%	51.45%
OS	ModelSE	20.29	23.06	34.84	62.08	22.16	4.51	17.68	20.29	23.06	34.84	62.08	22.16	4.51	17.68
OS	Cover2	0.98	0.94	1.00	0.64	0.97	0.00	0.99	0.98	0.96	1.00	0.65	0.97	0.00	0.99
OS	Cover1	1.00	1.00	1.00	0.89	1.00	0.03	1.00	1.00	1.00	1.00	0.89	1.00	0.03	1.00
QALYs	Mean	70.17	70.77	76.26	108.41	69.61	49.76	69.12	71.15	72.47	79.77	113.21	71.56	49.94	70.35
QALYs	Bias	-0.41	0.19	5.68	37.83	-0.97	-20.82	-1.46	0.57	1.89	9.19	42.63	0.98	-20.64	-0.23
QALYs	MC SE	0.23	0.19	0.25	0.80	0.21	0.08	0.22	0.28	0.27	0.36	0.76	0.25	0.08	0.23
QALYs	Rel Bias	- 0.01	0.00	0.08	0.54	- 0.01	- 0.29	- 0.02	0.01	0.03	0.13	0.60	0.01	- 0.29	- 0.00
QALYs	empSE	6.94	5.80	7.71	24.76	6.51	2.58	6.66	8.69	8.44	11.15	23.51	7.72	2.61	7.16
QALYs	MC SE	0.16	0.13	0.18	0.57	0.15	0.06	0.15	0.20	0.19	0.26	0.54	0.18	0.06	0.16
QALYs	MSE	48.34	33.63	91.73	2043.30	43.23	439.97	46.41	75.74	74.73	208.66	2369.38	60.51	433.00	51.27
QALYs	MC SE	1.68	1.38	3.77	69.32	1.55	3.48	1.50	16.90	19.29	19.41	75.61	2.91	3.49	2.31
QALYs	Rel P	0.00%	43.37%	-18.96%	-92.13%	13.92%	626.19%	8.76%	0.00%	5.97%	-39.31%	-86.34%	26.64%	1010.19%	47.25%
QALYs	ModelSE	10.42	11.93	17.75	31.33	11.48	2.96	9.51	10.42	11.93	17.75	31.33	11.48	2.96	9.51
QALYs	Cover2	0.97	0.99	1.00	0.64	0.96	0.00	0.97	0.97	1.00	1.00	0.66	0.97	0.00	0.98
QALYs	Cover1	1.00	1.00	1.00	0.89	1.00	0.06	1.00	1.00	1.00	1.00	0.89	1.00	0.06	1.00

Table 70 : Result for simulated scenario 48

	Scenario48	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	53.53	57.45	57.58	57.45	50.28	45.91	49.86	53.67	58.08	58.22	58.09	50.27	45.99	50.08
PFS	Bias	-2.40	1.52	1.65	1.52	-5.66	-10.02	-6.07	-2.26	2.15	2.29	2.16	-5.66	-9.94	-5.85
PFS	MC SE	0.12	0.19	0.19	0.20	0.15	0.09	0.11	0.13	0.20	0.20	0.20	0.15	0.09	0.11
PFS	Rel Bias	-4.3%	2.7%	3.0%	2.7%	-10.1%	-17.9%	-10.9%	-4.0%	3.8%	4.1%	3.9%	-10.1%	-17.8%	-10.5%
PFS	empSE	3.84	5.93	5.97	6.05	4.52	2.77	3.41	3.89	6.13	6.17	6.25	4.60	2.79	3.45
PFS	MC SE	0.09	0.14	0.14	0.14	0.10	0.06	0.08	0.09	0.14	0.14	0.14	0.11	0.06	0.08
PFS	MSE	20.48	37.44	38.31	38.93	52.42	108.15	48.46	20.21	42.12	43.21	43.72	53.20	106.64	46.11
PFS	MC SE	0.79	1.66	1.75	1.84	1.58	1.79	1.39	0.78	1.88	1.98	2.08	1.62	1.79	1.36
PFS	Rel P	0.00%	-58.15%	-58.66%	-59.83%	-28.01%	91.98%	26.58%	0.00%	-59.70%	-60.21%	-61.33%	-28.44%	93.68%	26.91%
PFS	ModelSE	3.91	5.33	5.36	5.34	4.34	3.03	3.88	3.91	5.33	5.36	5.34	4.34	3.03	3.88
PFS	Cover2	0.84	0.89	0.89	0.88	0.59	0.12	0.62	0.84	0.88	0.88	0.87	0.59	0.13	0.64
PFS	Cover1	0.99	1.00	1.00	1.00	0.91	0.64	0.95	0.99	1.00	1.00	1.00	0.91	0.64	0.95
OS	Mean	76.68	75.18	83.49	119.44	72.44	59.65	77.15	77.23	76.20	85.43	122.61	73.77	59.78	77.90
OS	Bias	-2.26	-3.76	4.55	40.50	-6.50	-19.29	-1.79	-1.71	-2.74	6.49	43.67	-5.17	-19.16	-1.04
OS	MC SE	0.31	0.17	0.26	0.80	0.17	0.10	0.31	0.32	0.24	0.36	0.75	0.21	0.10	0.32
OS	Rel Bias	- 0.03	- 0.05	0.06	0.51	- 0.08	- 0.24	- 0.02	- 0.02	- 0.03	0.08	0.55	- 0.07	- 0.24	- 0.01
OS	empSE	9.50	5.33	8.14	24.53	5.18	3.03	9.67	9.85	7.47	11.07	23.09	6.47	3.05	9.89
OS	MC SE	0.22	0.12	0.19	0.56	0.12	0.07	0.22	0.23	0.17	0.25	0.53	0.15	0.07	0.23
OS	MSE	95.35	42.57	86.81	2241.68	68.99	381.24	96.58	99.76	63.23	164.63	2439.68	68.46	376.34	98.78
OS	MC SE	0.22	0.12	0.19	0.56	0.12	0.07	0.22	2.66	23.14	29.19	74.46	2.27	3.78	2.85
OS	Rel P	0.00%	217.58%	36.47%	-84.99%	237.10%	885.48%	-3.37%	0.00%	73.78%	-20.93%	-81.81%	131.92%	939.42%	-0.89%
OS	ModelSE	7.90	10.58	15.45	30.89	8.88	3.45	7.82	7.90	10.58	15.45	30.89	8.88	3.45	7.82
OS	Cover2	0.80	0.85	0.98	0.76	0.73	0.00	0.83	0.81	0.87	0.99	0.77	0.74	0.00	0.85
OS	Cover1	1.00	1.00	1.00	0.92	0.99	0.16	1.00	1.00	1.00	1.00	0.92	0.99	0.16	1.00
QALYs	Mean	54.40	54.82	59.02	76.96	51.30	43.60	53.53	54.72	55.53	60.18	78.73	51.97	43.69	53.98
QALYs	Bias	-1.85	-1.43	2.77	20.71	-4.95	-12.65	-2.72	-1.53	-0.72	3.93	22.48	-4.28	-12.56	-2.27
QALYs	MC SE	0.17	0.13	0.17	0.42	0.10	0.07	0.17	0.18	0.15	0.21	0.40	0.11	0.07	0.17
QALYs	Rel Bias	- 0.03	- 0.03	0.05	0.37	- 0.09	- 0.22	- 0.05	- 0.03	- 0.01	0.07	0.40	- 0.08	- 0.22	- 0.04
QALYs	empSE	5.31	3.88	5.28	12.86	3.04	2.28	5.24	5.54	4.70	6.58	12.23	3.53	2.30	5.37
QALYs	MC SE	0.12	0.09	0.12	0.30	0.07	0.05	0.12	0.13	0.11	0.15	0.28	0.08	0.05	0.12
QALYs	MSE	31.56	17.10	35.52	593.96	33.67	165.27	34.83	32.98	22.57	58.68	654.89	30.81	163.09	33.98
QALYs	MC SE	0.85	0.63	1.39	20.45	0.91	1.87	0.99	0.91	5.90	7.79	20.32	0.90	1.87	0.95
QALYs	Rel P	0.00%	86.76%	1.04%	-82.96%	205.57%	440.97%	2.49%	0.00%	38.91%	-29.11%	-79.50%	145.67%	478.44%	6.30%
QALYs	ModelSE	4.31	6.19	8.43	15.97	5.22	2.58	4.83	4.31	6.19	8.43	15.97	5.22	2.58	4.83
QALYs	Cover2	0.77	0.93	0.97	0.77	0.75	0.00	0.79	0.78	0.95	0.97	0.78	0.77	0.00	0.80
QALYs	Cover1	0.97	1.00	1.00	0.93	0.99	0.29	1.00	0.97	1.00	1.00	0.93	0.99	0.29	1.00

Table 71 : Result for simulated scenario 49

	Scenario49	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	66.34	66.38	66.28	66.42	63.93	57.71	63.61	66.95	67.02	66.92	67.06	63.19	57.84	64.69
PFS	Bias	-2.96	-2.92	-3.02	-2.87	-5.37	-11.58	-5.68	-2.35	-2.28	-2.38	-2.23	-6.11	-11.46	-4.60
PFS	MC SE	0.21	0.21	0.21	0.21	0.20	0.12	0.20	0.22	0.22	0.22	0.22	0.23	0.12	0.21
PFS	Rel Bias	-4.3%	-4.2%	-4.4%	-4.1%	-7.8%	-16.7%	-8.2%	-3.4%	-3.3%	-3.4%	-3.2%	-8.8%	-16.5%	-6.6%
PFS	empSE	6.44	6.45	6.46	6.44	6.28	3.79	6.07	6.67	6.68	6.71	6.68	7.12	3.81	6.32
PFS	MC SE	0.15	0.15	0.15	0.15	0.14	0.09	0.14	0.15	0.15	0.15	0.15	0.16	0.09	0.15
PFS	MSE	50.21	50.00	50.83	49.64	68.24	148.54	69.07	49.93	49.83	50.57	49.56	87.98	145.77	61.08
PFS	MC SE	2.06	2.04	2.07	2.04	2.37	2.87	2.34	2.15	2.14	2.16	2.14	3.54	2.86	2.19
PFS	Rel P	0.00%	-0.14%	-0.70%	0.16%	5.20%	188.69%	12.70%	0.00%	-0.49%	-1.11%	-0.35%	-12.37%	206.27%	11.31%
PFS	ModelSE	5.98	6.01	6.01	6.03	8.42	3.79	5.61	5.98	6.01	6.01	6.03	8.42	3.79	5.61
PFS	Cover2	0.84	0.84	0.83	0.84	0.72	0.18	0.71	0.85	0.85	0.85	0.85	0.72	0.19	0.76
PFS	Cover1	0.97	0.97	0.97	0.97	0.94	0.68	0.97	0.97	0.97	0.97	0.97	0.94	0.68	0.97
OS	Mean	365.92	336.06	399.32	454.53	330.31	246.07	356.87	367.68	335.06	403.95	460.25	331.25	247.61	364.72
OS	Bias	1.41	-28.44	34.81	90.02	-34.20	-118.44	-7.64	3.17	-29.45	39.44	95.74	-33.26	-116.90	0.21
OS	MC SE	1.01	0.90	1.74	2.23	0.88	0.64	0.96	1.00	0.92	1.64	2.30	1.06	0.65	0.96
OS	Rel Bias	0.00	-0.08	0.10	0.25	-0.09	-0.32	-0.02	0.01	-0.08	0.11	0.26	-0.09	0.32	0.00
OS	empSE	31.05	27.85	53.49	68.74	27.02	19.76	29.58	30.85	28.43	50.62	70.91	32.81	20.14	29.58
OS	MC SE	0.71	0.64	1.23	1.58	0.62	0.45	0.68	0.71	0.65	1.16	1.63	0.75	0.46	0.68
OS	MSE	964.86	1584.13	4069.89	12824.63	1898.72	14417.77	932.69	960.80	1674.29	4115.59	14188.90	2181.81	14071.16	873.84
OS	MC SE	0.71	0.64	1.23	1.58	0.62	0.45	0.68	32.02	59.90	233.90	601.81	103.06	150.99	33.65
OS	Rel P	0.00%	24.24%	-66.31%	-79.60%	32.02%	146.88%	10.13%	0.00%	17.78%	-62.86%	-81.07%	-11.61%	134.60%	8.80%
OS	ModelSE	48.88	45.61	79.93	83.58	54.24	22.71	47.93	48.88	45.61	79.93	83.58	54.24	22.71	47.93
OS	Cover2	1.00	0.91	0.90	0.61	0.90	0.00	0.99	1.00	0.91	0.91	0.61	0.92	0.00	1.00
OS	Cover1	1.00	1.00	0.99	0.95	0.99	0.31	1.00	1.00	1.00	0.99	0.95	0.99	0.31	1.00
QALYs	Mean	202.86	187.95	219.54	247.19	184.33	140.35	197.52	203.93	187.64	222.05	250.24	184.58	141.16	201.77
QALYs	Bias	-0.18	-15.10	16.50	44.15	-18.71	-62.69	-5.53	0.88	-15.41	19.01	47.20	-18.46	-61.89	-1.28
QALYs	MC SE	0.52	0.46	0.87	1.13	0.45	0.34	0.49	0.51	0.47	0.83	1.16	0.56	0.34	0.50
QALYs	Rel Bias	0.00	-0.07	0.08	0.22	-0.09	-0.31	-0.03	0.00	-0.08	0.09	0.23	-0.09	0.30	0.01
QALYs	empSE	15.92	14.12	26.96	34.81	13.81	10.34	15.24	15.82	14.39	25.60	35.88	17.14	10.54	15.27
QALYs	MC SE	0.37	0.32	0.62	0.80	0.32	0.24	0.35	0.36	0.33	0.59	0.82	0.39	0.24	0.35
QALYs	MSE	253.25	426.95	998.11	3159.77	540.55	4037.42	262.50	250.71	444.10	1016.01	3513.71	634.31	3940.97	234.59
QALYs	MC SE	8.60	14.38	43.81	100.20	17.18	41.65	10.11	8.53	15.79	58.28	150.82	30.38	41.88	8.90
QALYs	Rel P	0.00%	27.22%	-65.12%	-79.08%	32.91%	137.05%	9.16%	0.00%	20.90%	-61.83%	-80.56%	-14.84%	125.43%	7.28%
QALYs	ModelSE	24.54	23.03	40.23	42.09	28.43	11.84	24.48	24.54	23.03	40.23	42.09	28.43	11.84	24.48
QALYs	Cover2	1.00	0.91	0.90	0.62	0.89	0.00	0.99	1.00	0.91	0.92	0.63	0.91	0.00	1.00
QALYs	Cover1	1.00	0.99	0.99	0.96	0.99	0.29	1.00	1.00	0.99	0.99	0.96	0.99	0.29	1.00

Table 72 : Result for simulated scenario 50

	Scenario50	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	65.85	66.26	66.11	66.22	63.52	57.29	62.37	66.30	66.89	66.74	66.85	62.84	57.41	62.84
PFS	Bias	-3.36	-2.96	-3.10	-2.99	-5.70	-11.92	-6.85	-2.91	-2.33	-2.48	-2.37	-6.38	-11.80	-6.38
PFS	MC SE	0.19	0.20	0.20	0.20	0.20	0.12	0.17	0.19	0.21	0.21	0.21	0.24	0.12	0.17
PFS	Rel Bias	-4.9%	-4.3%	-4.5%	-4.3%	-8.2%	-17.2%	-9.9%	-4.2%	-3.4%	-3.6%	-3.4%	-9.2%	-17.1%	-9.2%
PFS	empSE	5.90	6.25	6.22	6.24	6.23	3.66	5.30	5.96	6.45	6.42	6.44	7.32	3.68	5.35
PFS	MC SE	0.14	0.14	0.14	0.14	0.14	0.08	0.12	0.14	0.15	0.15	0.15	0.17	0.08	0.12
PFS	MSE	46.06	47.72	48.23	47.85	71.23	155.48	74.95	44.01	47.01	47.27	47.08	94.18	152.87	69.21
PFS	MC SE	1.88	1.98	1.99	1.97	2.43	2.83	2.45	1.83	2.01	2.02	2.01	3.93	2.82	2.33
PFS	Rel P	0.00%	-10.88%	-10.02%	-10.65%	-10.42%	159.47%	23.81%	0.00%	-14.55%	-13.57%	-14.31%	-33.55%	162.96%	24.53%
PFS	ModelSE	5.53	5.92	5.87	5.91	8.59	3.72	5.10	5.53	5.92	5.87	5.91	8.59	3.72	5.10
PFS	Cover2	0.82	0.84	0.84	0.84	0.72	0.14	0.65	0.84	0.86	0.85	0.85	0.72	0.15	0.67
PFS	Cover1	0.97	0.98	0.98	0.98	0.93	0.64	0.95	0.97	0.98	0.98	0.98	0.93	0.64	0.95
OS	Mean	202.47	203.43	228.56	296.13	201.64	151.01	200.83	203.49	203.61	233.32	297.84	201.73	151.46	203.60
OS	Bias	-0.21	0.76	25.88	93.46	-1.04	-51.67	-1.84	0.81	0.93	30.64	95.16	-0.95	-51.22	0.92
OS	MC SE	0.74	0.76	0.80	1.54	0.69	0.27	0.69	0.75	0.77	0.85	1.45	0.84	0.27	0.69
OS	Rel Bias	- 0.00	0.00	0.13	0.46	- 0.01	- 0.25	- 0.01	0.00	0.00	0.15	0.47	- 0.00	- 0.25	0.00
OS	empSE	22.78	23.41	24.72	47.38	21.37	8.26	21.18	23.15	23.82	26.13	44.79	25.95	8.35	21.19
OS	MC SE	0.52	0.54	0.57	1.09	0.49	0.19	0.49	0.53	0.55	0.60	1.03	0.60	0.19	0.49
OS	MSE	518.49	548.05	1280.08	10976.31	457.48	2738.12	451.66	536.03	567.61	1620.75	11059.97	673.83	2693.45	449.30
OS	MC SE	0.52	0.54	0.57	1.09	0.49	0.19	0.49	15.86	17.80	70.23	313.60	27.72	27.65	13.93
OS	Rel P	0.00%	-5.30%	-15.05%	-76.88%	13.59%	660.37%	15.66%	0.00%	-5.54%	-21.51%	-73.29%	-20.44%	667.79%	19.38%
OS	ModelSE	20.86	22.27	41.83	50.47	29.98	9.73	21.18	20.86	22.27	41.83	50.47	29.98	9.73	21.18
OS	Cover2	0.96	0.87	0.95	0.53	0.90	0.00	0.95	0.96	0.88	0.94	0.53	0.91	0.00	0.97
OS	Cover1	1.00	1.00	0.99	0.79	1.00	0.23	1.00	1.00	1.00	0.99	0.79	1.00	0.23	1.00
QALYs	Mean	120.99	121.59	134.11	167.93	119.88	92.69	119.13	121.64	121.87	136.68	168.97	119.71	92.95	120.65
QALYs	Bias	-1.11	-0.51	12.01	45.83	-2.23	-29.41	-2.98	-0.47	-0.23	14.57	46.87	-2.39	-29.15	-1.45
QALYs	MC SE	0.37	0.37	0.41	0.79	0.34	0.15	0.35	0.38	0.38	0.44	0.75	0.43	0.16	0.35
QALYs	Rel Bias	- 0.01	- 0.00	0.10	0.38	- 0.02	- 0.24	- 0.02	- 0.00	- 0.00	0.12	0.38	- 0.02	- 0.24	- 0.01
QALYs	empSE	11.44	11.45	12.77	24.30	10.46	4.76	10.79	11.65	11.65	13.60	23.06	13.22	4.82	10.80
QALYs	MC SE	0.26	0.26	0.29	0.56	0.24	0.11	0.25	0.27	0.27	0.31	0.53	0.30	0.11	0.25
QALYs	MSE	131.96	131.29	307.07	2690.37	114.34	887.72	125.08	135.86	135.62	397.07	2728.07	180.21	873.04	118.55
QALYs	MC SE	3.65	4.05	13.48	76.35	3.58	9.06	3.69	3.91	4.44	17.98	79.71	9.14	9.08	3.54
QALYs	Rel P	0.00%	-0.24%	-19.73%	-77.84%	19.50%	476.69%	12.46%	0.00%	0.05%	-26.54%	-74.46%	-22.27%	485.51%	16.49%
QALYs	ModelSE	10.66	11.66	21.39	25.78	16.72	5.54	11.40	10.66	11.66	21.39	25.78	16.72	5.54	11.40
QALYs	Cover2	0.93	0.91	0.95	0.54	0.90	0.00	0.94	0.94	0.91	0.95	0.55	0.92	0.00	0.96
QALYs	Cover1	1.00	1.00	0.99	0.81	1.00	0.22	1.00	1.00	1.00	0.99	0.81	1.00	0.22	1.00

Table 73 : Result for simulated scenario 51

	Scenario51	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	65.22	66.57	66.75	66.48	63.69	58.03	61.26	65.63	67.19	67.38	67.10	62.99	58.14	61.67
PFS	Bias	-3.90	-2.56	-2.38	-2.65	-5.43	-11.09	-7.87	-3.50	-1.93	-1.75	-2.02	-6.13	-10.98	-7.45
PFS	MC SE	0.18	0.20	0.20	0.20	0.21	0.12	0.16	0.18	0.21	0.21	0.20	0.25	0.12	0.16
PFS	Rel Bias	-5.6%	-3.7%	-3.4%	-3.8%	-7.9%	-16.0%	-11.4%	-5.1%	-2.8%	-2.5%	-2.9%	-8.9%	-15.9%	-10.8%
PFS	empSE	5.61	6.29	6.15	6.13	6.44	3.62	4.93	5.68	6.47	6.33	6.31	7.59	3.65	5.01
PFS	MC SE	0.13	0.14	0.14	0.14	0.15	0.08	0.11	0.13	0.15	0.15	0.14	0.17	0.08	0.12
PFS	MSE	46.66	46.05	43.47	44.54	70.87	136.11	86.17	44.39	45.59	43.04	43.83	95.23	133.89	80.64
PFS	MC SE	1.86	1.95	1.86	1.88	2.45	2.67	2.58	1.80	2.00	1.91	1.91	4.30	2.67	2.50
PFS	Rel P	0.00%	-20.42%	-16.86%	-16.23%	-24.00%	139.93%	29.52%	0.00%	-23.15%	-19.53%	-19.05%	-44.14%	141.50%	28.17%
PFS	ModelSE	5.23	5.88	5.88	5.84	8.81	3.79	4.80	5.23	5.88	5.88	5.84	8.81	3.79	4.80
PFS	Cover2	0.80	0.85	0.86	0.85	0.73	0.20	0.56	0.81	0.87	0.88	0.87	0.73	0.20	0.58
PFS	Cover1	0.97	0.98	0.98	0.98	0.94	0.72	0.93	0.97	0.98	0.98	0.98	0.94	0.72	0.93
OS	Mean	145.63	142.49	156.35	203.40	140.47	114.42	144.79	146.50	142.81	157.55	203.83	139.90	114.65	146.23
OS	Bias	2.63	-0.51	13.34	60.39	-2.54	-28.58	1.78	3.50	-0.20	14.54	60.83	-3.10	-28.35	3.23
OS	MC SE	0.58	0.50	0.45	0.85	0.47	0.18	0.56	0.60	0.52	0.47	0.82	0.58	0.18	0.58
OS	Rel Bias	0.02	-0.00	0.09	0.42	-0.02	0.20	0.01	0.02	-0.00	0.10	0.43	-0.02	0.20	0.02
OS	empSE	17.89	15.52	13.96	26.26	14.58	5.44	17.32	18.36	15.96	14.54	25.24	17.77	5.48	17.73
OS	MC SE	0.41	0.36	0.32	0.60	0.33	0.12	0.40	0.42	0.37	0.33	0.58	0.41	0.13	0.41
OS	MSE	326.64	240.81	372.56	4336.24	218.86	846.60	302.87	348.85	254.38	422.65	4335.86	324.91	833.73	324.41
OS	MC SE	0.41	0.36	0.32	0.60	0.33	0.12	0.40	10.95	12.98	15.07	106.41	16.08	10.01	10.53
OS	Rel P	0.00%	32.92%	64.34%	-53.58%	50.50%	983.16%	6.69%	0.00%	32.35%	59.42%	-47.08%	6.77%	1024.04%	7.21%
OS	ModelSE	13.00	14.23	18.08	29.88	20.10	6.24	13.05	13.00	14.23	18.08	29.88	20.10	6.24	13.05
OS	Cover2	0.77	0.79	0.98	0.50	0.81	0.00	0.82	0.76	0.80	0.97	0.49	0.81	0.00	0.82
OS	Cover1	0.99	0.99	0.99	0.80	0.99	0.27	1.00	0.99	0.99	0.99	0.80	0.99	0.27	1.00
QALYs	Mean	92.38	91.22	98.20	121.64	89.34	74.62	90.77	92.94	91.56	98.99	122.05	88.85	74.77	91.62
QALYs	Bias	0.14	-1.02	5.96	29.40	-2.90	-17.62	-1.47	0.70	-0.68	6.75	29.81	-3.39	-17.47	-0.62
QALYs	MC SE	0.30	0.25	0.25	0.45	0.24	0.11	0.30	0.31	0.26	0.27	0.44	0.31	0.12	0.30
QALYs	Rel Bias	0.00	-0.01	0.06	0.32	-0.03	0.19	-0.02	0.01	-0.01	0.07	0.32	-0.04	0.19	-0.01
QALYs	empSE	9.37	7.81	7.84	13.96	7.43	3.54	9.15	9.65	8.09	8.18	13.54	9.53	3.56	9.37
QALYs	MC SE	0.22	0.18	0.18	0.32	0.17	0.08	0.21	0.22	0.19	0.19	0.31	0.22	0.08	0.22
QALYs	MSE	87.77	61.93	96.94	1059.27	63.54	322.95	85.71	93.49	65.88	112.43	1071.56	102.21	317.88	88.07
QALYs	MC SE	2.76	2.43	3.55	28.92	2.15	4.04	2.68	2.97	3.75	4.15	28.13	5.85	4.03	2.75
QALYs	Rel P	0.00%	44.13%	42.83%	-54.93%	59.12%	602.02%	5.02%	0.00%	42.15%	39.01%	-49.21%	2.53%	632.68%	6.06%
QALYs	ModelSE	6.77	7.95	9.97	15.78	12.13	4.01	7.50	6.77	7.95	9.97	15.78	12.13	4.01	7.50
QALYs	Cover2	0.76	0.87	0.99	0.59	0.86	0.00	0.81	0.75	0.87	0.99	0.56	0.87	0.00	0.83
QALYs	Cover1	0.95	0.99	0.99	0.85	0.99	0.32	0.99	0.95	0.99	0.99	0.85	0.99	0.32	0.99

Table 74 : Result for simulated scenario 52

	Scenario52	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	59.12	59.16	59.07	59.17	54.15	50.19	55.83	59.62	59.69	59.60	59.70	54.09	50.26	56.08
PFS	Bias	3.02	3.06	2.97	3.07	-1.94	-5.91	-0.27	3.52	3.59	3.50	3.60	-2.01	-5.84	-0.02
PFS	MC SE	0.18	0.18	0.18	0.18	0.16	0.10	0.15	0.18	0.18	0.18	0.18	0.16	0.10	0.15
PFS	Rel Bias	5.4%	5.5%	5.3%	5.5%	-3.5%	-10.5%	-0.5%	6.3%	6.4%	6.2%	6.4%	-3.6%	-10.4%	0.0%
PFS	empSE	5.44	5.47	5.44	5.42	4.78	3.13	4.65	5.56	5.59	5.57	5.55	4.89	3.15	4.66
PFS	MC SE	0.12	0.13	0.12	0.12	0.11	0.07	0.11	0.13	0.13	0.13	0.13	0.11	0.07	0.11
PFS	MSE	38.63	39.22	38.44	38.81	26.62	44.75	21.68	43.32	44.14	43.27	43.75	27.95	43.96	21.68
PFS	MC SE	1.80	1.83	1.81	1.84	1.13	1.25	0.98	2.01	2.03	2.02	2.05	1.20	1.24	1.02
PFS	Rel P	0.00%	-1.08%	-0.31%	0.48%	29.23%	201.28%	36.59%	0.00%	-0.99%	-0.34%	0.51%	29.33%	212.00%	42.59%
PFS	ModelSE	4.84	4.85	4.84	4.86	4.28	3.10	4.31	4.84	4.85	4.84	4.86	4.28	3.10	4.31
PFS	Cover2	0.91	0.91	0.91	0.91	0.82	0.49	0.91	0.88	0.88	0.88	0.88	0.82	0.51	0.92
PFS	Cover1	1.00	1.00	1.00	1.00	0.98	0.90	1.00	1.00	1.00	1.00	1.00	0.98	0.90	1.00
OS	Mean	337.53	329.71	398.29	427.32	327.13	229.10	327.63	339.93	330.76	403.07	439.25	332.24	231.58	331.13
OS	Bias	-34.30	-42.13	26.45	55.48	-44.70	-142.74	-44.20	-31.90	-41.07	31.23	67.42	-39.60	-140.26	-40.70
OS	MC SE	1.02	0.92	1.71	2.20	0.89	0.72	1.04	1.01	0.89	1.84	2.35	0.92	0.74	1.01
OS	Rel Bias	- 0.09	- 0.11	0.07	0.15	- 0.12	- 0.38	- 0.12	- 0.09	- 0.11	0.08	0.18	- 0.11	- 0.38	- 0.11
OS	empSE	31.32	28.28	52.70	67.94	27.56	22.11	32.14	31.06	27.52	56.60	72.53	28.25	22.79	31.16
OS	MC SE	0.72	0.65	1.21	1.56	0.63	0.51	0.74	0.71	0.63	1.30	1.66	0.65	0.52	0.72
OS	MSE	2156.34	2573.85	3474.60	7689.56	2757.45	20863.62	2985.73	1981.46	2443.55	4175.62	9800.04	2365.66	20192.36	2626.79
OS	MC SE	0.72	0.65	1.21	1.56	0.63	0.51	0.74	63.35	73.35	326.55	546.11	73.61	203.85	77.71
OS	Rel P	0.00%	22.62%	-64.70%	-78.76%	29.06%	100.57%	-5.06%	0.00%	27.37%	-69.89%	-81.66%	20.84%	85.65%	-0.68%
OS	ModelSE	52.95	52.76	83.50	91.01	54.02	26.17	49.97	52.95	52.76	83.50	91.01	54.02	26.17	49.97
OS	Cover2	0.90	0.94	0.92	0.74	0.93	0.00	0.79	0.91	0.95	0.93	0.75	0.95	0.00	0.81
OS	Cover1	1.00	1.00	1.00	0.97	1.00	0.35	0.98	1.00	1.00	1.00	0.97	1.00	0.35	0.98
QALYs	Mean	186.50	182.60	216.87	231.41	179.81	129.60	180.57	187.85	183.29	219.42	237.54	182.34	130.87	182.39
QALYs	Bias	-16.25	-20.15	14.12	28.66	-22.94	-73.14	-22.18	-14.90	-19.46	16.67	34.79	-20.40	-71.88	-20.36
QALYs	MC SE	0.51	0.47	0.86	1.11	0.45	0.37	0.53	0.51	0.45	0.92	1.18	0.47	0.38	0.51
QALYs	Rel Bias	- 0.08	- 0.10	0.07	0.14	- 0.11	- 0.36	- 0.11	- 0.07	- 0.10	0.08	0.17	- 0.10	- 0.35	- 0.10
QALYs	empSE	15.79	14.35	26.56	34.18	13.99	11.35	16.27	15.66	14.00	28.47	36.43	14.34	11.70	15.77
QALYs	MC SE	0.36	0.33	0.61	0.78	0.32	0.26	0.37	0.36	0.32	0.65	0.84	0.33	0.27	0.36
QALYs	MSE	513.03	611.53	904.23	1988.46	721.40	5478.88	756.46	466.76	574.36	1087.76	2536.18	621.87	5303.61	662.77
QALYs	MC SE	16.11	18.85	43.73	72.58	20.94	53.15	21.93	15.36	17.79	83.23	138.14	19.24	53.75	19.88
QALYs	Rel P	0.00%	21.12%	-64.66%	-78.65%	27.50%	93.49%	-5.79%	0.00%	25.12%	-69.76%	-81.53%	19.16%	79.18%	-1.41%
QALYs	ModelSE	26.55	26.42	41.85	45.61	27.17	13.41	25.37	26.55	26.42	41.85	45.61	27.17	13.41	25.37
QALYs	Cover2	0.91	0.95	0.92	0.73	0.93	0.00	0.80	0.92	0.95	0.93	0.74	0.95	0.00	0.83
QALYs	Cover1	1.00	1.00	1.00	0.97	1.00	0.34	0.99	1.00	1.00	1.00	0.97	1.00	0.34	0.99

Table 75 : Result for simulated scenario 53

	Scenario53	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	58.95	59.32	59.23	59.27	54.08	49.83	54.43	59.35	59.86	59.77	59.81	54.06	49.90	54.67
PFS	Bias	2.87	3.24	3.15	3.19	-2.00	-6.25	-1.65	3.27	3.78	3.69	3.73	-2.02	-6.18	-1.41
PFS	MC SE	0.17	0.18	0.18	0.18	0.16	0.10	0.14	0.17	0.18	0.18	0.18	0.16	0.10	0.14
PFS	Rel Bias	5.1%	5.8%	5.6%	5.7%	-3.6%	-11.2%	-2.9%	5.8%	6.7%	6.6%	6.7%	-3.6%	-11.0%	-2.5%
PFS	empSE	5.20	5.43	5.43	5.41	4.83	3.04	4.32	5.26	5.56	5.56	5.53	5.00	3.05	4.33
PFS	MC SE	0.12	0.12	0.12	0.12	0.11	0.07	0.10	0.12	0.13	0.13	0.13	0.11	0.07	0.10
PFS	MSE	35.20	40.00	39.42	39.38	27.33	48.31	21.35	38.34	45.15	44.46	44.44	29.04	47.46	20.68
PFS	MC SE	1.54	1.82	1.81	1.80	1.13	1.29	0.87	1.66	2.02	2.02	2.00	1.27	1.28	0.85
PFS	Rel P	0.00%	-8.51%	-8.51%	-7.65%	15.58%	192.72%	44.72%	0.00%	-10.36%	-10.33%	-9.42%	10.79%	196.77%	47.86%
PFS	ModelSE	4.61	4.87	4.87	4.87	4.32	3.07	4.08	4.61	4.87	4.87	4.87	4.32	3.07	4.08
PFS	Cover2	0.91	0.91	0.91	0.92	0.81	0.47	0.87	0.89	0.88	0.88	0.89	0.80	0.48	0.87
PFS	Cover1	1.00	0.99	0.99	0.99	0.97	0.87	0.99	1.00	0.99	0.99	0.99	0.97	0.87	0.99
OS	Mean	172.20	169.60	197.64	255.58	170.80	127.55	168.19	173.15	170.95	204.01	260.79	173.38	128.05	170.36
OS	Bias	0.04	-2.56	25.48	83.42	-1.36	-44.62	-3.98	0.99	-1.21	31.85	88.63	1.22	-44.11	-1.80
OS	MC SE	0.59	0.62	0.73	1.47	0.60	0.25	0.54	0.60	0.63	0.80	1.45	0.69	0.25	0.54
OS	Rel Bias	0.00	-0.01	0.15	0.48	-0.01	-0.26	-0.02	0.01	-0.01	0.19	0.51	0.01	-0.26	-0.01
OS	empSE	18.08	19.10	22.61	45.41	18.61	7.65	16.60	18.58	19.41	24.61	44.66	21.13	7.73	16.76
OS	MC SE	0.42	0.44	0.52	1.04	0.43	0.18	0.38	0.43	0.45	0.56	1.03	0.48	0.18	0.38
OS	MSE	326.63	371.14	1160.07	9018.16	347.74	2049.03	290.94	346.00	377.95	1619.28	9847.76	447.41	2005.62	283.82
OS	MC SE	0.42	0.44	0.52	1.04	0.43	0.18	0.38	13.87	12.77	84.63	328.14	14.52	22.08	10.13
OS	Rel P	0.00%	-10.41%	-36.05%	-84.14%	-5.57%	458.52%	18.72%	0.00%	-8.36%	-42.95%	-82.68%	-22.63%	477.68%	22.97%
OS	ModelSE	18.63	18.98	40.81	52.94	21.42	8.95	18.33	18.63	18.98	40.81	52.94	21.42	8.95	18.33
OS	Cover2	0.96	0.89	0.95	0.51	0.94	0.00	0.94	0.97	0.92	0.95	0.52	0.93	0.00	0.97
OS	Cover1	1.00	1.00	0.99	0.80	1.00	0.30	1.00	1.00	1.00	0.99	0.80	1.00	0.30	1.00
QALYs	Mean	103.78	102.60	116.59	145.57	101.63	78.72	100.42	104.38	103.43	119.94	148.34	102.91	79.00	101.58
QALYs	Bias	0.88	-0.31	13.69	42.67	-1.28	-24.18	-2.48	1.48	0.53	17.03	45.43	0.00	-23.91	-1.32
QALYs	MC SE	0.31	0.31	0.37	0.75	0.30	0.14	0.28	0.32	0.32	0.41	0.74	0.34	0.14	0.29
QALYs	Rel Bias	0.01	0.00	0.13	0.41	-0.01	-0.24	-0.02	0.01	0.01	0.17	0.44	0.00	-0.23	-0.01
QALYs	empSE	9.47	9.57	11.54	23.06	9.36	4.31	8.72	9.74	9.74	12.62	22.72	10.56	4.36	8.81
QALYs	MC SE	0.22	0.22	0.26	0.53	0.21	0.10	0.20	0.22	0.22	0.29	0.52	0.24	0.10	0.20
QALYs	MSE	90.44	91.50	320.45	2351.38	89.15	603.43	82.05	96.92	94.99	449.28	2580.09	111.44	590.60	79.28
QALYs	MC SE	3.18	3.18	12.54	68.19	2.66	6.78	2.68	3.89	3.61	22.14	84.71	3.50	6.77	2.75
QALYs	Rel P	0.00%	-1.90%	-32.64%	-83.11%	2.47%	382.79%	18.16%	0.00%	0.03%	-40.48%	-81.63%	-14.99%	399.52%	22.19%
QALYs	ModelSE	9.51	9.81	20.67	26.76	11.07	4.99	9.78	9.51	9.81	20.67	26.76	11.07	4.99	9.78
QALYs	Cover2	0.97	0.95	0.95	0.51	0.94	0.00	0.94	0.97	0.96	0.94	0.52	0.93	0.00	0.96
QALYs	Cover1	1.00	1.00	0.98	0.81	1.00	0.31	1.00	1.00	1.00	0.98	0.81	1.00	0.31	1.00

Table 76 : Result for simulated scenario 54

	Scenario54	**Deterministic**							**Probabilistic**						
		PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model	PSM	STM	STM - Log	STM - Time	MSM	Li's model	Fu's model
PFS	Mean	57.28	59.48	59.63	59.43	54.21	50.37	53.02	57.53	60.02	60.17	59.97	54.24	50.46	53.15
PFS	Bias	1.29	3.49	3.63	3.44	-1.78	-5.62	-2.97	1.54	4.03	4.17	3.98	-1.75	-5.54	-2.84
PFS	MC SE	0.15	0.18	0.18	0.18	0.16	0.10	0.13	0.15	0.19	0.18	0.18	0.17	0.10	0.13
PFS	Rel Bias	2.3%	6.2%	6.5%	6.1%	-3.2%	-10.0%	-5.3%	2.8%	7.2%	7.5%	7.1%	-3.1%	-9.9%	-5.1%
PFS	empSE	4.52	5.63	5.51	5.52	5.00	3.02	3.90	4.53	5.75	5.64	5.65	5.19	3.04	3.91
PFS	MC SE	0.10	0.13	0.13	0.13	0.11	0.07	0.09	0.10	0.13	0.13	0.13	0.12	0.07	0.09
PFS	MSE	22.06	43.81	43.56	42.31	28.11	40.66	24.00	22.85	49.25	49.15	47.67	29.91	39.85	23.37
PFS	MC SE	1.01	2.02	1.98	1.95	1.22	1.16	0.94	1.04	2.24	2.19	2.16	1.36	1.15	0.92
PFS	Rel P	0.00%	-35.53%	-32.82%	-33.12%	-18.23%	124.15%	34.48%	0.00%	-37.99%	-35.45%	-35.72%	-23.75%	122.54%	34.02%
PFS	ModelSE	4.00	4.88	4.89	4.87	4.46	3.13	3.74	4.00	4.88	4.89	4.87	4.46	3.13	3.74
PFS	Cover2	0.91	0.89	0.89	0.90	0.83	0.54	0.81	0.91	0.86	0.86	0.87	0.83	0.55	0.81
PFS	Cover1	1.00	0.99	0.99	0.99	0.97	0.90	0.98	1.00	0.99	0.99	0.99	0.97	0.90	0.98
OS	Mean	112.17	113.11	126.14	163.78	109.92	93.67	110.82	112.75	114.23	127.99	165.32	111.09	93.85	111.64
OS	Bias	-1.30	-0.36	12.68	50.31	-3.54	-19.80	-2.64	-0.72	0.77	14.52	51.85	-2.38	-19.61	-1.82
OS	MC SE	0.41	0.40	0.38	0.73	0.31	0.15	0.38	0.43	0.44	0.40	0.71	0.39	0.15	0.39
OS	Rel Bias	-0.01	0.00	0.11	0.44	-0.03	-0.17	0.02	-0.01	0.01	0.13	0.46	-0.02	-0.17	0.02
OS	empSE	12.71	12.33	11.72	22.44	9.61	4.61	11.76	13.15	13.66	12.38	21.85	12.09	4.63	12.08
OS	MC SE	0.29	0.28	0.27	0.52	0.22	0.11	0.27	0.30	0.31	0.28	0.50	0.28	0.11	0.28
OS	MSE	163.08	152.06	297.90	3034.21	104.72	413.22	145.13	173.34	187.08	364.08	3165.74	151.61	406.08	149.20
OS	MC SE	0.29	0.28	0.27	0.52	0.22	0.11	0.27	7.74	29.40	13.19	79.00	27.95	5.92	5.86
OS	Rel P	0.00%	6.23%	17.69%	-67.91%	75.11%	660.28%	16.82%	0.00%	-7.33%	12.88%	-63.77%	18.41%	706.32%	18.48%
OS	ModelSE	9.76	12.56	14.49	25.81	13.40	5.24	9.61	9.76	12.56	14.49	25.81	13.40	5.24	9.61
OS	Cover2	0.80	0.79	0.95	0.51	0.83	0.03	0.83	0.80	0.80	0.92	0.50	0.82	0.05	0.84
OS	Cover1	0.99	0.98	0.99	0.83	1.00	0.48	1.00	0.99	0.98	0.99	0.83	1.00	0.48	1.00
QALYs	Mean	73.27	74.40	80.96	99.72	71.22	61.95	71.32	73.63	75.12	82.04	100.65	71.82	62.06	71.76
QALYs	Bias	-0.26	0.87	7.43	26.19	-2.31	-11.58	-2.21	0.10	1.59	8.51	27.12	-1.71	-11.47	-1.77
QALYs	MC SE	0.22	0.21	0.21	0.38	0.16	0.10	0.21	0.23	0.23	0.23	0.38	0.20	0.10	0.21
QALYs	Rel Bias	0.00	0.01	0.10	0.36	-0.03	-0.16	0.03	0.00	0.02	0.12	0.37	-0.02	-0.16	0.02
QALYs	empSE	6.77	6.56	6.59	11.82	4.97	2.98	6.32	7.01	7.23	6.97	11.60	6.17	2.99	6.49
QALYs	MC SE	0.16	0.15	0.15	0.27	0.11	0.07	0.15	0.16	0.17	0.16	0.27	0.14	0.07	0.15
QALYs	MSE	45.91	43.69	98.56	825.33	29.96	143.07	44.81	49.06	54.77	120.97	870.00	41.01	140.45	45.24
QALYs	MC SE	1.92	2.46	3.40	21.73	1.03	2.27	1.52	2.17	8.05	4.28	21.95	7.43	2.25	1.60
QALYs	Rel P	0.00%	6.77%	5.73%	-67.15%	86.06%	417.06%	14.83%	0.00%	-6.08%	1.21%	-63.54%	28.85%	447.72%	16.44%
QALYs	ModelSE	5.15	6.86	7.90	13.45	7.28	3.34	5.59	5.15	6.86	7.90	13.45	7.28	3.34	5.59
QALYs	Cover2	0.81	0.89	0.92	0.52	0.87	0.08	0.84	0.81	0.89	0.89	0.50	0.87	0.08	0.85
QALYs	Cover1	0.97	0.98	0.99	0.82	0.99	0.57	1.00	0.97	0.98	0.99	0.82	0.99	0.57	1.00

Appendix 14 : Predicted mean time in health state using different assumption about treatment effect

Table 77 : Predicted health state sojourn time and QALYs - Dataset 3 [base-case highlighted in Yellow and Greeen]

	Control				Intervention				Incremental			
	PFS	PD	LY	QALY	PFS	PD	LY	QALY	PFS	PD	LY	QALY
TRUTH	**150.93**	**106.26**	**257.18**	**173.87**	**190.75**	**106.74**	**297.50**	**205.98**	**39.83**	**0.49**	**40.32**	**32.11**
STRINGENT CURVE SELECTION CRITERIA												
PSM BC	140.28	94.30	234.57	159.37	171.43	114.62	286.05	194.45	31.15	20.33	51.48	35.08
PSM SC1	140.22	93.14	233.36	158.74	171.53	117.01	288.54	195.73	31.31	23.87	55.18	36.98
PSM SC2	140.28	94.30	234.57	159.37	170.02	93.76	263.78	182.89	29.74	-0.54	29.20	23.52
STM BC	140.91	119.16	260.07	172.31	172.21	125.38	297.60	200.46	31.30	6.22	37.52	28.15
STM SC1	140.91	99.70	240.62	162.58	172.21	140.80	313.01	208.17	31.30	41.09	72.39	45.59
STM SC1	140.91	95.93	236.84	160.70	172.21	156.50	328.72	216.02	31.30	60.57	91.87	55.33
STMlog B	140.91	120.30	261.22	172.88	172.21	126.99	299.21	201.27	31.30	6.69	37.99	28.38
STMlog S	140.91	96.41	237.33	160.94	172.21	135.51	307.73	205.53	31.30	39.10	70.40	44.59
STMlog S	140.91	94.91	235.83	160.19	172.21	137.62	309.83	206.58	31.30	42.71	74.00	46.39
LESS STRINGENT CURVE SELECTION CRITERIA												
PSM BC	140.91	148.31	289.22	186.89	172.21	183.06	355.27	229.30	31.30	34.75	66.05	42.41
PSM SC1	140.91	139.14	280.06	182.30	172.21	200.81	373.03	238.18	31.30	61.67	92.97	55.87
PSM SC2	140.91	148.31	289.22	186.89	172.21	155.01	327.23	215.28	31.30	6.70	38.00	28.39
STM BC	140.91	119.16	260.07	172.31	172.21	125.38	297.60	200.46	31.30	6.22	37.52	28.15
STM SC1	140.91	99.70	240.62	162.58	172.21	140.80	313.01	208.17	31.30	41.09	72.39	45.59
STM SC1	140.91	101.57	242.48	163.52	172.21	159.92	332.14	217.73	31.30	58.35	89.65	54.22
STMlog B	140.91	120.30	261.22	172.88	172.21	126.99	299.21	201.27	31.30	6.69	37.99	28.38
STMlog S	140.91	104.05	244.96	164.76	172.21	145.87	318.08	210.70	31.30	41.82	73.12	45.95
STMlog S	140.91	103.81	244.73	164.64	172.21	143.93	316.14	209.74	31.30	40.11	71.41	45.10

Abbreviations : BC : base-case ; LY: life years; PD: progressive disease; PFS: progression-free survival; PSM : partitioned survival model; QALYs: quality adjusted life years; SC: scenario; STM: state-transition model; STMlog: state-transition using log of time

Table 78 : Predicted health state sojourn time and QALYs - Dataset 4 [base-case highlighted in Yellow and Green]

	Control				Intervention				Incremental			
	PFS	PD	LY	QALY	PFS	PD	LY	QALY	PFS	PD	LY	QALY
TRUTH	**153.57**	**137.76**	**291.33**	**191.74**	**203.04**	**95.76**	**298.80**	**210.31**	**49.47**	**-42.01**	**7.47**	**18.58**
STRINGENT CURVE SELECTION CRITERIA												
PSM BC	143.94	157.72	301.67	194.02	191.06	145.13	336.19	225.41	47.11	-12.59	34.52	31.39
PSM SC1	143.94	157.44	301.38	193.87	191.06	145.35	336.41	225.52	47.11	-12.08	35.03	31.65
PSM SC2	143.94	157.72	301.67	194.02	191.06	135.38	326.44	220.54	47.11	-22.34	24.77	26.52
STM BC	143.94	113.95	257.89	172.13	191.06	106.07	297.13	205.88	47.11	-7.88	39.23	33.75
STM SC1	143.94	118.16	262.10	174.23	191.06	104.19	295.24	204.94	47.11	-13.97	33.14	30.71
STM SC1	143.94	133.60	277.55	181.96	191.06	98.44	289.50	202.07	47.11	-35.16	11.95	20.11
STMlog B	143.94	148.87	292.81	189.59	191.06	164.30	355.36	235.00	47.11	15.43	62.54	45.41
STMlog S	143.94	162.06	306.01	196.19	191.06	159.59	350.65	232.64	47.11	-2.47	44.65	36.46
STMlog S	143.94	173.91	317.85	202.11	191.06	83.68	274.74	194.69	47.11	-90.23	-43.11	-7.42
LESS STRINGENT CURVE SELECTION CRITERIA												
PSM BC	143.94	157.72	301.67	194.02	191.06	145.13	336.19	225.41	47.11	-12.59	34.52	31.39
PSM SC1	143.94	157.44	301.38	193.87	191.06	145.35	336.41	225.52	47.11	-12.08	35.03	31.65
PSM SC2	143.94	157.72	301.67	194.02	191.06	135.38	326.44	220.54	47.11	-22.34	24.77	26.52
STM BC	143.94	113.95	257.89	172.13	191.06	106.07	297.13	205.88	47.11	-7.88	39.23	33.75
STM SC1	143.94	118.16	262.10	174.23	191.06	104.19	295.24	204.94	47.11	-13.97	33.14	30.71
STM SC1	143.94	133.60	277.55	181.96	191.06	98.44	289.50	202.07	47.11	-35.16	11.95	20.11
STMlog B	143.94	148.87	292.81	189.59	191.06	164.30	355.36	235.00	47.11	15.43	62.54	45.41
STMlog S	143.94	162.06	306.01	196.19	191.06	159.59	350.65	232.64	47.11	-2.47	44.65	36.46
STMlog S	143.94	193.16	337.10	211.73	191.06	92.44	283.50	199.07	47.11	-100.72	-53.60	-12.67

Abbreviations : BC : base-case ; LY: life years; PD: progressive disease; PFS: progression-free survival; PSM : partitioned survival model; QALYs: quality adjusted life years; SC: scenario; STM: state-transition model; STMlog: state-transition using log of time

Table 79 : Predicted health state sojourn time and QALYs - Dataset 7 [base-case highlighted in Yellow and Green]

	Control				Intervention				Incremental			
	PFS	PD	LY	QALY	PFS	PD	LY	QALY	PFS	PD	LY	QALY
TRUTH	**219.51**	**112.24**	**331.75**	**231.73**	**205.78**	**66.15**	**271.93**	**197.70**	**-13.73**	**-46.09**	**-59.82**	**-34.03**
STRINGENT CURVE SELECTION CRITERIA												
PSM BC	176.98	89.78	266.77	186.48	173.59	60.58	234.18	169.17	-3.39	-29.20	-32.59	-17.31
PSM SC1	176.98	106.24	283.22	194.70	173.60	72.50	246.11	175.13	-3.38	-33.73	-37.11	-19.57
PSM SC2	176.98	89.78	266.77	186.48	173.60	71.52	245.12	174.64	-3.38	-18.27	-21.65	-11.84
STM BC	176.98	82.39	259.37	182.78	173.60	87.23	260.83	182.50	-3.38	4.84	1.46	-0.28
STM SC1	176.98	95.20	272.18	189.19	173.60	81.66	255.26	179.71	-3.38	-13.54	-16.92	-9.47
STM SC1	176.98	148.67	325.66	215.92	173.60	57.52	231.12	167.64	-3.38	-91.15	-94.53	-48.28
STMlog B	176.98	100.02	277.01	191.60	173.60	104.77	278.37	191.27	-3.38	4.74	1.36	-0.33
STMlog S	176.98	90.30	267.28	186.74	173.60	70.05	243.66	173.91	-3.38	-20.25	-23.63	-12.83
STMlog S	176.98	105.28	282.26	194.23	173.60	96.72	270.32	187.24	-3.38	-8.56	-11.94	-6.99
LESS STRINGENT CURVE SELECTION CRITERIA												
PSM BC	176.98	89.78	266.77	186.48	173.59	60.58	234.18	169.17	-3.39	-29.20	-32.59	-17.31
PSM SC1	176.98	106.24	283.22	194.70	173.60	72.50	246.11	175.13	-3.38	-33.73	-37.11	-19.57
PSM SC2	176.98	89.78	266.77	186.48	173.60	71.52	245.12	174.64	-3.38	-18.27	-21.65	-11.84
STM BC	176.98	82.39	259.37	182.78	173.60	87.23	260.83	182.50	-3.38	4.84	1.46	-0.28
STM SC1	176.98	95.20	272.18	189.19	173.60	81.66	255.26	179.71	-3.38	-13.54	-16.92	-9.47
STM SC1	176.98	148.67	325.66	215.92	173.60	57.52	231.12	167.64	-3.38	-91.15	-94.53	-48.28
STMlog B	176.98	100.02	277.01	191.60	173.60	104.77	278.37	191.27	-3.38	4.74	1.36	-0.33
STMlog S	176.98	90.30	267.28	186.74	173.60	70.05	243.66	173.91	-3.38	-20.25	-23.63	-12.83
STMlog S	176.98	105.28	282.26	194.23	173.60	96.72	270.32	187.24	-3.38	-8.56	-11.94	-6.99

Abbreviations : BC : base-case ; LY: life years; PD: progressive disease; PFS: progression-free survival; PSM : partitioned survival model; QALYs: quality adjusted life years; SC: scenario; STM: state-transition model; STMlog: state-transition using log of time

Table 80 : Predicted health state sojourn time and QALYs – Dataset 12 [base-case highlighted in Yellow and Green]

	Control				Intervention				Incremental			
	PFS	PD	LY	QALY	PFS	PD	LY	QALY	PFS	PD	LY	QALY
TRUTH	**94.74**	**147.56**	**242.30**	**149.57**	**147.27**	**111.16**	**258.43**	**173.39**	**52.53**	**-36.40**	**16.13**	**23.82**
STRINGENT CURVE SELECTION CRITERIA												
PSM BC	104.44	118.86	223.31	142.99	151.09	85.52	236.61	163.63	46.65	-33.34	13.30	20.64
PSM SC1	104.88	123.94	228.82	145.87	150.11	82.85	232.96	161.51	45.23	-41.09	4.14	15.64
PSM SC2	104.44	118.86	223.31	142.99	NA	NA	NA	NA	NA	NA	NA	NA
STM BC	106.68	114.83	221.51	142.76	160.63	107.53	268.16	182.27	53.95	-7.30	46.65	39.51
STM SC1	106.68	127.89	234.57	149.29	160.63	97.57	258.19	177.29	53.95	-30.32	23.63	28.00
STM SC1	106.68	116.66	223.33	143.67	160.63	99.40	260.03	178.20	53.95	-17.26	36.69	34.53
STMlog B	106.68	137.89	244.56	154.28	160.63	144.42	305.05	200.71	53.95	6.53	60.49	46.43
STMlog S	106.68	146.26	252.94	158.47	160.63	122.91	283.54	189.96	53.95	-23.35	30.61	31.49
STMlog S	106.68	130.21	236.89	150.45	160.63	170.24	330.87	213.62	53.95	40.03	93.98	63.18
LESS STRINGENT CURVE SELECTION CRITERIA												
PSM BC	105.92	135.55	241.47	152.51	156.19	100.63	256.82	175.27	50.27	-34.92	15.36	22.76
PSM SC1	106.08	140.98	247.06	155.35	155.65	97.38	253.04	173.21	49.58	-43.60	5.98	17.86
PSM SC2	105.92	135.55	241.47	152.51	155.24	94.00	249.24	171.19	49.32	-41.55	7.77	18.68
STM BC	106.68	114.83	221.51	142.76	160.63	107.53	268.16	182.27	53.95	-7.30	46.65	39.51
STM SC1	106.68	127.89	234.57	149.29	160.63	97.57	258.19	177.29	53.95	-30.32	23.63	28.00
STM SC1	106.68	116.66	223.33	143.67	160.63	99.40	260.03	178.20	53.95	-17.26	36.69	34.53
STMlog B	106.68	137.89	244.56	154.28	160.63	144.42	305.05	200.71	53.95	6.53	60.49	46.43
STMlog S	106.68	162.07	268.74	166.37	160.63	130.88	291.51	193.94	53.95	-31.19	22.76	27.57
STMlog S	106.68	157.79	264.47	164.24	160.63	130.58	291.21	193.79	53.95	-27.22	26.74	29.55

Abbreviations : BC : base-case ; LY: life years; PD: progressive disease; PFS: progression-free survival; PSM : partitioned survival model; QALYs: quality adjusted life years; SC: scenario; STM: state-transition model; STMlog: state-transition using log of time

Table 81 : Predicted health state sojourn time and QALYs – Dataset 18 [base-case highlighted in Yellow and Green]

	Control				Intervention				Incremental			
	PFS	PD	LY	QALY	PFS	PD	LY	QALY	PFS	PD	LY	QALY
TRUTH	**129.23**	**241.13**	**370.36**	**223.95**	**167.02**	**243.26**	**410.27**	**255.24**	**37.79**	**2.12**	**39.91**	**31.29**
STRINGENT CURVE SELECTION CRITERIA												
PSM BC	120.67	210.55	331.22	201.81	176.86	241.79	418.64	262.38	56.19	31.23	87.43	60.57
PSM SC1	120.64	201.25	321.90	197.14	176.90	252.12	429.02	267.58	56.26	50.86	107.12	70.44
PSM SC2	120.67	210.55	331.22	201.81	176.40	185.83	362.23	234.03	55.73	-24.72	31.01	32.22
STM BC	120.70	232.29	353.00	212.71	177.00	231.99	408.98	257.59	56.29	-0.30	55.99	44.88
STM SC1	120.70	225.72	346.42	209.42	177.00	236.89	413.89	260.04	56.29	11.17	67.46	50.62
STM SC1	120.70	225.72	346.42	209.42	177.00	236.89	413.89	260.04	56.29	11.17	67.46	50.62
STMlog B	120.70	253.21	373.92	223.17	177.00	295.12	472.11	289.16	56.29	41.91	98.20	65.99
STMlog S	120.70	269.79	390.49	231.46	177.00	289.51	466.51	286.35	56.29	19.72	76.01	54.89
STMlog S	120.70	262.39	383.10	227.76	177.00	394.51	571.51	338.85	56.29	132.12	188.41	111.09
LESS STRINGENT CURVE SELECTION CRITERIA												
PSM BC	120.67	210.55	331.22	201.81	176.86	241.79	418.64	262.38	56.19	31.23	87.43	60.57
PSM SC1	120.70	295.43	416.13	244.28	177.00	378.81	555.81	331.00	56.29	83.39	139.68	86.73
PSM SC2	120.67	210.55	331.22	201.81	176.40	185.83	362.23	234.03	55.73	-24.72	31.01	32.22
STM BC	120.70	207.44	328.14	200.28	177.00	207.17	384.16	245.18	56.29	-0.27	56.02	44.90
STM SC1	120.70	204.05	324.76	198.59	177.00	209.92	386.91	246.56	56.29	5.87	62.16	47.97
STM SC1	120.70	334.55	455.26	263.84	177.00	411.58	588.58	347.39	56.29	77.03	133.32	83.55
STMlog B	120.70	285.22	405.92	239.17	177.00	341.08	518.07	312.13	56.29	55.86	112.15	72.96
STMlog S	120.70	269.79	390.49	231.46	177.00	289.51	466.51	286.35	56.29	19.72	76.01	54.89
STMlog S	120.70	262.39	383.10	227.76	177.00	394.51	571.51	338.85	56.29	132.12	188.41	111.09

Abbreviations : BC : base-case ; LY: life years; PD: progressive disease; PFS: progression-free survival; PSM : partitioned survival model; QALYs: quality adjusted life years; SC: scenario; STM: state-transition model; STMlog: state-transition using log of time

Table 82 : Predicted health state sojourn time and QALYs – Dataset 20 [base-case highlighted in Yellow and Green]

	Control				Intervention				Incremental			
	PFS	PD	LY	QALY	PFS	PD	LY	QALY	PFS	PD	LY	QALY
TRUTH	**101.31**	**52.85**	**154.16**	**107.47**	**152.52**	**61.92**	**214.44**	**152.98**	**51.21**	**9.08**	**60.28**	**45.50**
STRINGENT CURVE SELECTION CRITERIA												
PSM BC	115.31	72.61	187.92	128.55	178.49	158.24	336.73	221.91	63.18	85.64	148.82	93.36
PSM SC1	115.31	72.00	187.31	128.25	178.49	159.19	337.68	222.39	63.18	87.19	150.37	94.14
PSM SC2	115.31	72.61	187.92	128.55	178.49	114.83	293.32	200.21	63.18	42.22	105.40	71.66
STM SC1	115.31	54.47	169.79	119.49	178.49	50.84	229.33	168.21	63.18	-3.63	59.55	48.73
STM BC	115.31	39.25	154.57	111.88	178.49	81.47	259.97	183.53	63.18	42.22	105.40	71.65
STM SC1	115.31	37.41	152.73	110.96	178.49	86.15	264.64	185.87	63.18	48.74	111.92	74.91
STMlog S	115.31	96.24	211.56	140.37	178.49	121.04	299.54	203.32	63.18	24.80	87.98	62.94
STMlog B	115.31	58.29	173.60	121.40	178.49	134.31	312.80	209.95	63.18	76.02	139.20	88.55
STMlog S	115.31	57.10	172.41	120.80	178.49	90.52	269.01	188.05	63.18	33.42	96.60	67.25
LESS STRINGENT CURVE SELECTION CRITERIA												
PSM BC	115.31	72.61	187.92	128.55	178.49	158.24	336.73	221.91	63.18	85.64	148.82	93.36
PSM SC1	115.31	72.00	187.31	128.25	178.49	159.19	337.68	222.39	63.18	87.19	150.37	94.14
PSM SC2	115.31	72.61	187.92	128.55	178.49	114.83	293.32	200.21	63.18	42.22	105.40	71.66
STM SC1	115.31	54.47	169.79	119.49	178.49	50.84	229.33	168.21	63.18	-3.63	59.55	48.73
STM BC	115.31	39.25	154.57	111.88	178.49	81.47	259.97	183.53	63.18	42.22	105.40	71.65
STM SC1	115.31	37.41	152.73	110.96	178.49	86.15	264.64	185.87	63.18	48.74	111.92	74.91
STMlog S	115.31	96.24	211.56	140.37	178.49	121.04	299.54	203.32	63.18	24.80	87.98	62.94
STMlog B	115.31	58.29	173.60	121.40	178.49	134.31	312.80	209.95	63.18	76.02	139.20	88.55
STMlog S	115.31	57.10	172.41	120.80	178.49	90.52	269.01	188.05	63.18	33.42	96.60	67.25

Abbreviations : BC : base-case ; LY: life years; PD: progressive disease; PFS: progression-free survival; PSM : partitioned survival model; QALYs: quality adjusted life years; SC: scenario; STM: state-transition model; STMlog: state-transition using log of time

Appendix 15 : Parametric distribution selected for the stringent and less stringent analysis

Table 83 : Summary of distributions selected – Stringent definition

	Dataset 3	Dataset 4	Dataset 7	Dataset 12	Dataset 18	Dataset 20
Time point used for PFS selection criteria	413.55	448	449.6	290.1	377	305.25
Time point used for OS selection criteria	582	821.3	750.4	516.1	929.95	441.25
Difference in PPS (p-value)	0.328	0.640	0.272	0.314	0.824	0.006
Selected OS distribution	Gamma	Log-logistic	Weibull	Weibull	Gamma	Generalised Gamma
Selected PFS distribution	Log-normal	Log-logistic	Gamma	Log-normal	Log-normal	Log-normal
Selected prePS distribution	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
Selected PPS distribution for the unadjusted STM	Gamma	Log-logistic	Log-logistic	Gamma	Exponential	Log-normal
Selected PPS distribution for the adjusted STM	Gamma	Log-logistic	Log-logistic	Gamma	Gompertz	Gamma

Table 84 : Summary of distributions selected – Less stringent definition

	Dataset 3	Dataset 4	Dataset 7	Dataset 12	Dataset 18	Dataset 20
Time point used for PFS selection criteria	298.7	338.4	378.8	245.9	272.1	235
Time point used for OS selection criteria	512.5	663.6	629.4	473.9	733.5	386.5
Difference in PPS (p-value)	0.328	0.640	0.272	0.314	0.824	0.006
Selected OS distribution	Log-normal	Log-logistic	Weibull	Gamma	Gamma	Log-logistic
Selected PFS distribution	Log-normal	Log-logistic	Gamma	Log-normal	Log-normal	Log-normal
Selected prePS distribution	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
Selected PPS distribution for the unadjusted STM	Gamma	Log-logistic	Log-logistic	Gamma	Gamma	Log-normal
Selected PPS distribution for the adjusted STM	Gamma	Log-logistic	Log-logistic	Gamma	Gamma	Log-normal