The analysis of overall survival in a changing clinical world: the methodological implications of relapsing cancers in randomised controlled trials

Miss Kara-Louise Royle

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of Leeds

School of Medicine

April 2026

Intellectual Property and Publication Statements

I, Kara-Louise Royle, confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contribution along with the contribution of the other authors to this work has been explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

Study One publication (Chapter 2)

Royle KL, Meads D, Visser-Rogers JK, White IR, Cairns DA. How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review. Trials. 2023 Nov 6;24(1):708
https://doi.org/10.1186/s13063-023-07730-1

Under supervision from DM, JKV-R, IRW and DAC, K-LR conceived the idea for the review, planned the project, performed the screening, extracted the data, and drafted the manuscript. All authors read and approved the final manuscript.

Study Two publication (Chapter 3)

Royle, KL., Wheatstone, P., Meads, D. et al. How should overall survival be analysed in randomised clinical trials in cancer if participants receive subsequent treatment lines? A stakeholder consultation. Trials ***26****, 434 (2025).*
https://doi.org/10.1186/s13063-025-09148-3

Under supervision from DM, JKV-R, IRW and DAC, K-LR devised and conducted the qualitative analysis, and drafted the manuscript. PW helped to recruit to and was an active member of both the co-design and stakeholder group. He also facilitated the focus group. All authors read and approved the final manuscript.

Study Three publication (Chapter 4)

Royle KL, Meads D, Visser-Rogers JK, Cairns DA, White IR. What Would Happen if Everyone Received the Same Subsequent Therapy? A Hypothetical Estimand for Overall Survival in the Presence of Treatment Discontinuation

Submitted to Statistics in Medicine 29th January 2026.
Under supervision from DM, JKV-R, DAC and IRW, K-LR devised and conducted the quantitative analysis, and drafted the manuscript. All authors read and approved the final manuscript.

Conference Abstracts and Other Dissemination

Study One and Study Two were accepted as poster presentations at the 2022 NIHR academy members conference and the Society of Clinical Trials conference 2023 respectively.

Study Three was accepted as an oral presentation at the Royal Statistical Society 2024 conference and poster presentations at the International Clinical Trials Methodology conference 2024, the Society of Clinical Trials conference 2025, the PSI conference 2025 and the International Society for Clinical Biostatistics conference 2025.

The patient and public involvement of this work was presented by K-LR as a case study during the “PPIE for statistical methodology research workshop” on the 26th February 2024, hosted by the Improving Statistical Literacy Working Group of the NIHR Statistics Group, NIHR Leicester BRC and NIHR ARC East Midlands. A summary of the workshop was published in Statistics in Medicine “*Incorporation of Patient and Public Involvement in Statistical Methodology Research: Summary of Workshop Proceedings*” https://doi.org/10.1002/sim.70159.The PPI work was also included during an invited session at the Royal Statistical Society 2024 conference entitled “*Patient and public involvement (PPI) for medical statistics methodology research: why bother?*” and an invited Cancer Research UK webinar entitled “*Maximising the impact of data webinar series: whose data is it anyway?”*.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. All publications resulting from this thesis have been published open access, and with a Creative Commons Attribution (CC BY) license.

The right of Kara-Louise Royle to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

Thesis Structure

I (K-LR) chose to use an alternative style of thesis, including published material, to maximise the outputs from this PhD. There are three studies, which include published material. An introduction (Chapter 1) precedes the three studies. Studies One and Two (Chapter 2 and Chapter 3) have been published in peer-reviewed journals. Study Three (Chapter 4) has been submitted to a peer-reviewed journal and is under review. A discussion (Chapter 5) follows the studies to bind the thesis into one piece of work. The three studies have been presented exactly as published or submitted, other than minor formatting and typographical changes to ensure consistency of presentation. Each chapter includes its own list of references and the supplementary material covering methods, tables and figures is included in the appendices as per University guidance. Note that for Study Two hyperlinks are provided to some supplementary information for full transparency, e.g. the full questionnaire, which are not necessary for examination but were published.

Acknowledgements

Firstly I would like to thank my supervisors; David Cairns, Ian White, David Meads and Jennifer Visser-Rogers. Each of you brought something different to the last five years and I couldn't have done it without you. David (C), your confidence in me is something I will always appreciate, and thank you for letting me be me and bringing me out of the many spirals over the years. I probably don't say that enough. Thank you Ian for answering a random email from someone you didn't know six years ago, giving me your time, support and learning to translate what was in my head into actual English at times. David M, thank you for bringing the HE perspective and not minding too much when I refer to you as Meads in meetings to differentiate between Davids. Thank you Jen for being a role model in terms of public engagement and inspiring the next generation of researchers through your work with schools.

Thank you to also the NIHR who funded this research, the members of my stakeholder advisory group, co-design group and participants in my stakeholder consultation and those of the studies' of data I used. Without all of you none of this would have been possible.

This PhD has given me many opportunities but I am most grateful for a friendship I have made through it. Erin I'm so glad you sat next to the strange girl sewing at the first APTs. I'm looking forward to all our future type 2 fun adventures. Anna, you have been with me since the start, picking me up when I am down and reminding me who the F I am. Maths Matt, babes, I couldn't have got through my first degree without you and I'm glad stabbing you with a pen didn't stop our friendship continuing all these years. The three of you will always be my chosen siblings.

To my actual sibling, Sam. Thank you for being a third parent so I could have a childhood. For being amazing at everything to encourage friendly sibling rivalry and keeping me modest.

Dad. I am who I am because of you. Sweary, sarcastic and very independent. You taught me that sometimes shit happens and you have to get on with things. I hope your sacrifices over the years are worth it.

Special thanks go to my unhelpful ginger ghost writer (cat) Ralphie. Who despite his best efforts (and multiple attempts) did not delete all my work when it was lunch or snack or general attention seeking time.

Finally, thank you to the A-Level tutor and college friends who persuaded me to do a maths degree in the first place, and set me on this path.

Abstract

Overall survival (OS) is defined as the time from randomisation to death from any cause. It is a definitive endpoint in phase III clinical trials to determine whether an experimental intervention improves participant outcomes compared to a control intervention. However, during clinical trial follow-up, participants, in any arm, may discontinue their trial intervention and begin subsequent anti-cancer therapy, which may improve outcomes.

This PhD thesis aimed to understand whether researchers were considering subsequent anti-cancer therapy in their analysis of OS, what stakeholders (statisticians and other data analysts, payers and industry partners, healthcare professionals, patient, carers and the public) thought about subsequent anti-cancer therapy and OS, and to develop new, or extend existing, methodology which assessed OS whilst accounting for subsequent anti-cancer therapy in randomised clinical trials.

Study One was a systematic literature review investigating how OS is reported. The review found that subsequent anti-cancer therapy was generally acknowledged as a potential limitation, but not addressed through sensitivity analysis. **Study Two** was a stakeholder consultation incorporating an online questionnaire and focus group. The questionnaire had 103 respondents, and seven participants attended the focus group. From the stakeholder consultation it was concluded that stakeholders were interested in the problem of subsequent anti-cancer therapy. This translated into an alternative hypothetical estimand which was investigated in **Study Three**. This was a simulation study assessing whether existing statistical analysis methods could be used to estimate the estimand: The treatment effect if everyone who did discontinue before death, went on to receive the same subsequent anti-cancer therapy. The existing methods were also applied to a real data case study. From the simulation and case study, the existing two-stage estimation method was concluded to be the most appropriate. These findings will now be disseminated to the consenting participants from study two and further research will be agreed collaboratively.

Contents

Intellectual Property and Publication Statements I
Conference Abstracts and Other Dissemination II
Thesis Structure III
Acknowledgements IV
Abstract V
Contents VI
Table of Figures XI
Table of Tables XII
Table of Appendix Figures XIII
Table of Appendix Tables XV
Acronyms XVII
Chapter 1 Introduction 1
1.1 Chapter Overview 1
1.2 Overall Survival 1
1.2.1 Implementation and Impact 1
1.2.2 The Issue of Subsequent Anti-Cancer Therapy 2
1.2.3 Current Approaches to Subsequent Anti-Cancer Therapy 2
1.2.4 Recent Examples of Subsequent Anti-Cancer Therapy and Clinical Trials 3
1.2.5 PhD Rationale 5
1.3 Aims and Objectives 6
1.3.1 Aim 6
1.3.2 Objectives 6
1.4 Overview of Relevant Methodology 7
1.4.1 Consensus Methods 7
1.4.2 Estimands 7
1.4.3 Time-to-Event Analysis 9
1.4.4 Causal Inference 13
1.4.5 Simulation Studies 15
1.4.6 Patient and Public Involvement 18
1.5 Thesis Summary 19
1.6 Chapter 1 References 19

Chapter 2 How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review.....22
2.1 Abstract.....22
2.1.1 Background.....22
2.1.2 Methods.....22
2.1.3 Results.....23
2.1.4 Discussion.....23
2.1.5 Funding.....23
2.2 Key Words.....23
2.3 Introduction.....23
2.4 Methods.....25
2.4.1 Search strategy.....25
2.4.2 Initial screening process.....28
2.4.3 Reading process.....28
2.5 Results.....30
2.5.1 Overview and sample size.....30
2.5.2 Clinical effectiveness papers.....30
2.5.3 Cost-effectiveness papers.....36
2.5.4 Methodology papers.....36
2.6 Discussion.....36
2.7 Conclusion.....40
2.8 Declarations.....41
2.8.1 Ethics approval and consent to participate.....41
2.8.2 Consent for publication.....41
2.8.3 Availability of data and material.....41
2.8.4 Competing Interests.....41
2.8.5 Funding.....41
2.8.6 Author contributions.....41
2.8.7 Acknowledgements.....41
2.9 Chapter 2 References.....41
Chapter 3 How should overall survival be analysed in randomised clinical trials in cancer if participants receive subsequent treatment lines? A stakeholder consultation.....46
3.1 Abstract.....46
3.1.1 Background.....46
3.1.2 Methods.....46
3.1.3 Results.....47

3.1.4 Discussion....47
3.1.5 Funding....47
3.2 Key Words....47
3.3 Introduction....48
3.4 Methods....50
3.4.1 Online Questionnaire....51
3.4.1.1 Development and Testing....51
3.4.1.2 Recruitment....56
3.4.1.3 Analysis....56
3.4.2 Focus Group....57
3.4.2.1 Design....57
3.4.2.2 Format....58
3.4.2.3 Analysis....58
3.5 Results....58
3.5.1 Online Questionnaire....58
3.5.1.1 Demographics....59
3.5.1.2 Additional Questions....60
3.5.1.3 Question of Interest....62
3.5.1.4 Information Required....63
3.5.1.5 Assumptions and Answer....67
3.5.1.6 Summary....67
3.5.2 Focus Group....67
3.5.2.1 Discussion 1: The question of interest....68
3.5.2.2 Discussion 2: Data Items....68
3.5.2.3 Discussion 3: Data Collection and Recording....69
3.6 Discussion....70
3.7 Conclusion....74
3.8 Declarations....74
3.8.1 Ethics approval and consent to participate....74
3.8.2 Consent for publication....74
3.8.3 Availability of data and material....74
3.8.4 Competing Interests....75
3.8.5 Funding....75
3.8.6 Author contributions....75
3.8.7 Acknowledgements....75
3.9 Chapter 3 References....75

Chapter 4 What Would Happen if Everyone Received the Same Subsequent Therapy? A Hypothetical Estimand for Overall Survival in the Presence of Treatment Discontinuation 77
4.1 Abstract 77
4.2 Key Words 78
4.3 Introduction 78
4.4 Overview of Simulation Study 79
4.4.1 Aims 79
4.4.2 Data-Generating Mechanism 80
4.4.2.1 Overview 80
4.4.2.2 Data Generating Process 80
4.4.2.3 Parameter Specification 81
4.4.3 Estimand 85
4.4.4 Methods 85
4.4.4.1 Assumptions 89
4.4.5 Performance Measures 89
4.4.6 Implementation of ADEMP 90
4.5 Simulation Study Results 91
4.5.1 Individual Method Implementation and Fit 91
4.5.2 Descriptive Cross-Method Performance 91
4.5.3 Overall Cross-Method Performance 94
4.5.4 Sensitivity Analysis 96
4.5.5 Summary 96
4.6 Case Study 97
4.6.1 Aim 97
4.6.2 Trial Overview and Estimand 97
4.6.3 Methods 98
4.6.4 Results 100
4.7 Discussion 103
4.8 Declarations 106
4.8.1 Ethics approval and consent to participate 106
4.8.2 Consent for publication 106
4.8.3 Availability of data and material 106
4.8.4 Competing Interests 107
4.8.5 Funding 107
4.8.6 Author contributions 107

4.8.7 Acknowledgements 107
4.9 Chapter 4 References 107
Chapter 5 Discussion 110
5.1 Thesis Overview 110
5.2 Developments in and Contributions to the Existing Literature and Infrastructure 111
5.2.1 Treatment Switching and Uptake of New Statistical Methodology 111
5.2.2 Healthcare Systems Data Access and Usage 119
5.2.3 Patient and Public Involvement in Statistical Methodology Research 120
5.2.4 Research Participation in Statistical Methodology Research 122
5.3 Recommendations for Cancer Clinical Trials 124
5.4 Extensions to Other Areas 125
5.5 Limitations 126
5.6 Future work 127
5.7 Conclusion 128
5.8 Chapter 5 References 128
Appendix A Ethical Approval 133
Appendix B Chapter 2 Supplementary Material 140
B.1 Appendix B References 188
Appendix C Chapter 3 Supplementary Material 200
Appendix D Chapter 4 Supplementary Material 221
D.1 Additional Methods Information 221
D.2 Preliminary Investigations 229
D.3 Additional Simulation Results – Under the Constant Treatment Effect Assumption 245
D.4 Simulation Results – Relaxing the Constant Treatment Effect Assumption 274
D.5 Additional Case Study Results 303
Appendix E Chapter 5 Appendix 304
E.1 Appendix E References 310

Table of Figures

Figure 1.1: Overall Survival....1
Figure 1.2: An Example of Overall Survival Showing Subsequent Therapy....2
Figure 1.3: Example of a Kaplan-Meier Curve....11
Figure 1.4: Example of a Directed Acyclic Graph (DAG)....14
Figure 2.1: Screening Process for abstracts....27
Figure 2.2: Was a Logrank Test Presented in the Main Text?....31
Figure 2.3: Was a Cox-PH Model Presented in the Main Text?....32
Figure 3.1: Questionnaire Structure....53
Figure 3.2: Bar Chart Showing the Number of Statisticians or Other Data Analysts who Collect Information During Trial Follow-up....61
Figure 3.3: Bar Chart Showing Statisticians and Other Data Analysts' Awareness of Methodology....62
Figure 3.4: Summary of Question of Interest Results....63
Figure 3.5: Summary of Data Collection Responses....64
Figure 3.6: Summary of Collecting Data and Recording Information Results....66
Figure 4.1: Visual Representation of the Simulation Study....82
Figure 4.2: Swarm Plot of Estimates across the Scenarios Generated Under the Constant Treatment Effect Assumption....94
Figure 4.3: Lollipop Plots of Performance Measures from Simulations Generated under the Constant Treatment Effect Assumption....95
Figure 4.4: Kaplan-Meier Survival Curves for the Simple Methods....101
Figure 4.5: Kaplan-Meier Survival Curves for the Advanced Methods....101
Figure 4.6: STAR trial: Summary of Model Results....102
Figure 5.1: Example of Reporting Subsequent Treatment Lines from IKEMA....117

Table of Tables

Table 1.1: The Impact of Different Intercurrent Event Approaches on the Research Question 9

Table 1.2: Different Performance Measures in a Simulation Study 16

Table 2.1: Literature Search Terms 26

Table 2.2: Information Extracted from Each Article Type 29

Table 2.3: Summary of the Clinical Effectiveness papers by endpoint type.....33

Table 3.1:Methods of Dealing with an Intercurrent Event 50

Table 3.2:The Definitions of the Stakeholder Groups.....51

Table 3.3: Summary of the Online Questionnaire.....54

Table 3.5: Questionnaire Decision Guidelines 57

Table 3.6: Summary of Questionnaire Respondents.....59

Table 4.1: Simulation study scenario descriptions, rationale, and parameter changes from the base scenario.....83

Table 4.2: Estimand of Interest for the Simulation Study.....85

Table 4.3: Mean (SD) of the number of events used in the Cox Model averaged across the repetitions. 93

Table 4.4: Estimand for STAR Real Data Case Study 98

Table 4.5: STAR trial: Second Line Treatment Type 100

Table 5.1: Screening Results from Literature Search Update 115

Table 5.2: Template Estimand Framework for Hypothetical Estimation.....124

Table of Appendix Figures

Appendix Figure C.1: Flow Chart of Implemented Methods 211

Appendix Figure C.2: Clinical Trials Tweetorial 212

Appendix Figure C.3: Statistics Tweetorial 213

Appendix Figure C.4: Survival Analysis Tweetorial 214

Appendix Figure C.5: Summary of Statistical Assumptions Results 215

Appendix Figure C.6: Summary of Presentation of Results 217

Appendix Figure C.7: Post-Focus Group Questionnaire Results – Data Items 218

Appendix Figure C.8: Post-Focus Group Questionnaire Results – Collecting Information 219

Appendix Figure C.9: Post-Focus Group Questionnaire Results – Recording Information 220

Appendix Figure D.1: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - BaseScenario with and without the constant treatment effect assumption 230

Appendix Figure D.2: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - HarmTrialTrt with and without the constant treatment effect assumption 232

Appendix Figure D.3:Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - NoEffectTrialTrt with and without the constant treatment effect assumption 232

Appendix Figure D.4: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - VeryBenTrialTrt with and without the constant treatment effect assumption 233

Appendix Figure D.5: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment -ThreeMonthlyFup with and without the constant treatment effect assumption 233

Appendix Figure D.6: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - MajorityExpTrtB with and without the constant treatment effect assumption 234

Appendix Figure D.7: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - MajorityDiscontExp with and without the constant treatment effect assumption 234

Appendix Figure D.8: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - EarlyDiscontinuation with and without the constant treatment effect assumption 235

Appendix Figure D.9: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - LateDiscontinuation with and without the constant treatment effect assumption 235

Appendix Figure D.10: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - MajorityDiscontA with and without the constant treatment effect assumption 236

Appendix Figure D.11:Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - VeryBenSubTrt with and without the constant treatment effect assumption.....236

Appendix Figure D.12: Asymptotic Performance of the ITT Method – With and without the constant treatment effect assumption.....238

Appendix Figure D.13: Asymptotic Performance of the IPCW Method – With and without the constant treatment effect assumption.....239

Appendix Figure D.14: Asymptotic Performance of the Two Stage Method using a Weibull AFT Model – With and without the constant treatment effect assumption 240

Appendix Figure D.15: Asymptotic Performance of the TSE Method using a Generalised Gamma AFT Model – With and without the constant treatment effect assumption.....241

Appendix Figure D.16: Plots Investigating the Simulation Study Implementation - Base Scenario 242

Appendix Figure D.17: Range Plot of the Unstabilised and Stabilised Weights from the IPCW Methods Across the Simulations and Scenarios - With the constant treatment effect assumption.....245

Appendix Figure D.18: Scatter Plots of the Estimates and their Standard Errors Across All Scenarios - With the constant treatment effect assumption 247

Appendix Figure D.19: Bland-Altman Plots of Estimates Across the Simulations - With the constant treatment effect assumption.....256

Appendix Figure D.20: Lollipop Plots of Performance Measures from Simulations - With the constant treatment effect assumption.....262

Appendix Figure D.21: Zip Plots Estimates Across the Simulations - With the constant treatment effect assumption.....268

Appendix Figure D.22: Swarm Plot of Estimates Across the Scenarios - Without the constant treatment effect assumption.....276

Appendix Figure D.23: Scatter Plots of the Estimates and their Standard Errors Across All Scenarios - Without the constant treatment effect assumption ...277

Appendix Figure D.24: Range Plot of the Unstabilised and Stabilised weights from the IPCW Methods Across the Simulations and Scenarios - Without the constant treatment effect assumption.....282

Appendix Figure D.25: Bland-Altman Plots of Estimates Across the Simulations - Without the constant treatment effect assumption.....287

Appendix Figure D.26: Lollipop Plots of Performance Measures from Simulations - Without the constant treatment effect assumption.....293

Appendix Figure D.27: Zip Plots Estimates Across the Simulations - Without the constant treatment effect assumption.....298

Table of Appendix Tables

Appendix Table B.1: PRISMA Abstract Checklist....140
Appendix Table B.2: PRISMA Reporting Checklist 141
Appendix Table B.3: Summarised Clinical Effectiveness Papers 144
Appendix Table B.4: Summary of the Clinical Effectiveness Papers by Year of Publication 175
Appendix Table B.5: Summary of the Effectiveness Papers by Cancer 180
Appendix Table B.6: Summarised Cost-effectiveness Papers....184
Appendix Table B.7: Summarised Methodology / Review papers 185
Appendix Table C.1: Checklist for Reporting Results of Internet E-Surveys (CHERRIES)....200
Appendix Table C.2: COREQ (Consolidated criteria for REporting Qualitative research) Checklist 203
Appendix Table C.3: Recipients of Targeted Emails 207
Appendix Table C.4: Summary of Invited Respondents 210
Appendix Table D.1: Simulation Study Parameter and Variable Descriptions for the Base Scenario 221
Appendix Table D.2: Cox Model for OS from Time 0 to the First Follow-up Timepoint – One Million Participants Base Scenario with the constant treatment effect assumption. 230
Appendix Table D.3: True Hazard Ratio for Each Scenario - With and without the constant treatment effect assumption....237
Appendix Table D.4: Parameter Estimates from Repetition 3/10 Base Scenario DGM 243
Appendix Table D.5: Parameter Estimates from Base Scenario Replication Check 244
Appendix Table D.6: AIC Goodness of Fit Comparisons for TSE AFT Models....253
Appendix Table D.7: Mean (SD) of Theta from the TSE Methods Calculated Across the Simulations 254
Appendix Table D.8: Percentage of Simulations Randomisation Allocation Violates Proportional Hazards Within Each Implemented Method 255
Appendix Table D.9: Mean (SD) of the Number of Events used in the Cox Model Averaged Across the Repetitions - Without the constant treatment effect assumption 274
Appendix Table D.10: AIC Goodness of Fit Comparisons for TSE AFT Models – Without the constant treatment effect assumption....283
Appendix Table D.11: Mean (SD) of Theta from the TSE Methods Calculated Across the Simulations 284
Appendix Table D.12: Percentage of Simulations Randomisation Allocation Violates Proportional Hazards Within Each Implemented Method 285

Appendix Table D.13: STAR trial: Follow-up Treatment Overview 303
Appendix Table E.1: Summarised Papers from Updated Literature Review.....304

Acronyms

ADEMP	Aim, Data generation, Estimand, Methods, Performance measures
AFT	Accelerated failure time
AIC	Akaike information criterion
ARTs	Androgen receptor signalling–targeted agents
ASCT	Autologous stem cell transplant
CCS	Conventional continuation strategy
CHERRIES	Checklist for Reporting Results of Internet E-Surveys
CI	Confidence interval
CoP	Community of practice
COREQ	Consolidated criteria for reporting qualitative research
Cox-PH	Cox proportional hazards model
CTU	Clinical trials unit
DAGs	Directed acyclic graphs
DARS	Data access request service
DFIS	Drug-free interval strategy
DGM	Data-generating mechanism
DUCkS	Data utility comparison studies
EORTC	European Organisation for Research and Treatment of Cancer
HR	Hazard ratio
HTA	Health technology assessment
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IPTW	Inverse proportional treatment weighting
IP	Internet protocol
IPCW	Inverse probability of censoring weights
IPCWN	Inverse probability of censoring weights with unstabilised weights
IPCWNNUC	Inverse probability of censoring weights with unstabilised weights and no unmeasured confounding
IPCWS	Inverse probability of censoring weights with stabilised weights.
IPCWSNUC	Inverse probability of censoring weights with stabilised weights and no unmeasured confounding
IQR	Inter-quartile range
IRB	Independent research board
ISRCTN	International standard randomised controlled trial number
ITT	Intention-to-treat
LICTR	Leeds Institute of Clinical Trials Research
MC	Monte Carlo
MHRA	Medicines and Healthcare products Regulatory Agency
NICE	National Institute for Health and Care Excellence
NDRS	National Disease Registration Service
NHS	National Health Service
NIHR	National Institute for Health and Care Research

NUC	No unmeasured confounding
OS	Overall survival
PPI-SMART	Patient and Public Involvement for Statistical Methods Research Techniques
PFS	Progression-free survival
PHE	Public Health England
PIS	Patient information sheet
PP	Per-protocol
PPCens	Per-protocol censor
PPEx	Per-protocol exclude
PPI	Patient and public involvement
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RAM	RAND/UCLA appropriateness method
RCC	Renal cell carcimona
RCT	Randomised controlled trial
RFS	Relapse-free survival
RNG	Random number generator
RPSFTM	Rank preserving structural failure time model
SAG	Stakeholder advisory group
SD	Standard deviation
SE	Standard error
SAIL	Secure anonymised information linkage
SMART	Sequential, multiple assignment, randomized trial
TKI	Tyrosine kinase inhibitor
TSD	Technical support documentation
TSE	Two-stage estimation
TSEGNoR	Two-stage estimation generalised gamma without recensoring
TSEGNUCNoR	Two-stage estimation generalised gamma without recensoring and no unmeasured confounding
TSEGR	Two-stage estimation generalised gamma with recensoring
TSENUCGR	Two-stage estimation generalised gamma with recensoring and no unmeasured confounding
TSEWNoR	Two-stage estimation Weibull without recensoring
TSEWNUCNoR	Two-stage estimation Weibull without recensoring and no unmeasured confounding
TSEWNUCR	Two-stage estimation Weibull with recensoring
TSEWR	Two-stage estimation Weibull with recensoring and no unmeasured confounding
UKCRC	United Kingdom Clinical Research Collaboration

Chapter 1 Introduction

1.1 Chapter Overview

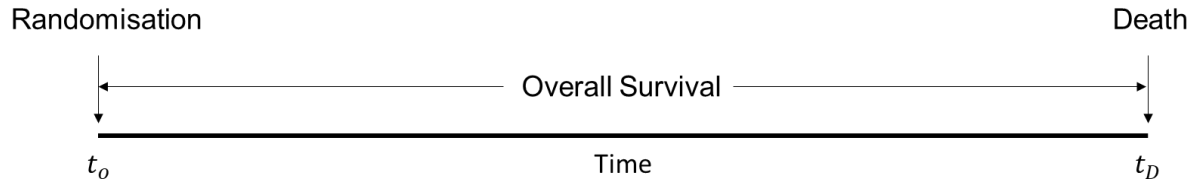
This chapter introduces the topic of the thesis: overall survival in cancer clinical trials when participants stop their clinical trial treatment prior to death and begin subsequent anti-cancer therapy. It also presents the aims and objectives of the thesis. Finally, it provides an overview of both the relevant methodology, required to meet the objectives, and the remaining chapters of the thesis.

1.2 Overall Survival

1.2.1 Implementation and Impact

Overall survival (OS) is a definitive endpoint for randomised controlled clinical trials in oncology. It is defined as the time from randomisation (t_0) to death (t_D) from any cause (**Figure 1.1**).

Figure 1.1: Overall Survival



The analysis of OS is implemented by statisticians and other data analysts e.g. health economists. The results are then used:

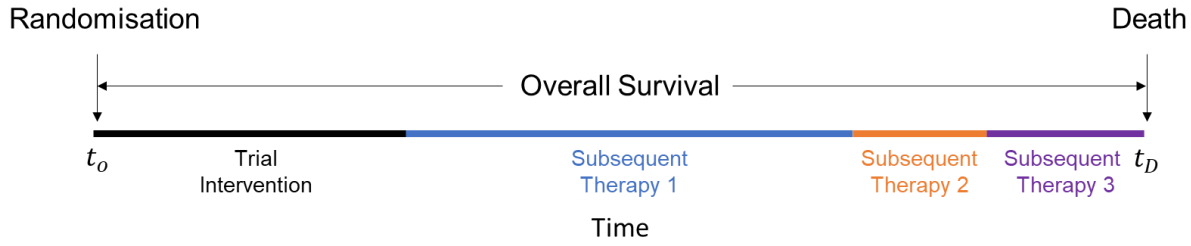
1. To assess clinical effectiveness of experimental interventions.
2. Alongside safety information by pharmaceutical companies to apply for product licences, also referred to as marketing authorisations, from regulatory bodies such as the MHRA (Medicines and Healthcare products Regulatory Agency).
3. In conjunction with other evidence in appraisals by organisations, such as NICE (National Institute for Health and Care Excellence), to determine whether interventions should be used in the National Health Service (NHS).

Therefore, statisticians and other data analysts, healthcare professionals, payers and industry partners, and patients and the public all have an interest or concern with OS i.e. are stakeholders.

1.2.2 The Issue of Subsequent Anti-Cancer Therapy

The analysis of OS requires clinical trial participants to be followed up for what can be a considerable length of time. This poses a methodological problem, particularly in relapsing cancers, as participants may discontinue their clinical trial intervention prior to death. This could be at relapse or progression, or earlier, due to withdrawal as a result of toxicity, patient choice, or a treatment stopping rule. Clinical trial participants may then go on to receive potentially multiple and different subsequent lines of anti-cancer therapy following completion of their clinical trial intervention (**Figure 1.2**).

Figure 1.2: An Example of Overall Survival Showing Subsequent Therapy



These subsequent therapies will likely also aim to improve OS. Therefore, the uptake of subsequent therapy may impact the OS assessment of the experimental clinical trial intervention by reducing the number of OS events and in turn reduce the power available to detect, if present, a statistically significant difference in OS between the clinical trial experimental and control interventions. This could lead to a beneficial intervention not being recommended for use in the NHS, if the control arm participants have been rescued by their subsequent therapy, resulting in patients potentially missing out on additional life-years in society, at work and with their families. The converse could also be true i.e. a non-beneficial experimental intervention being accepted, if the experimental arm have been rescued by their subsequent therapy. This issue with interpretation makes the ITT analysis of OS a less than satisfactory approach when participants go on to subsequent therapies.

1.2.3 Current Approaches to Subsequent Anti-Cancer Therapy

One approach to mitigate the potential impact and incorrect conclusions being made as a result of subsequent anti-cancer therapy, is to use surrogate endpoints. ICH E9 defines a surrogate endpoint as "*A variable that provides an indirect measurement of effect in situations where direct measurement of clinical effect is not feasible or practical.*"[1] From a practical perspective, surrogate endpoints measured earlier in the treatment pathway, such as progression-free survival (PFS), should not be influenced by confounding introduced by subsequent anti-cancer therapy. Therefore, it is common for a clinical trial to have PFS as the primary endpoint and OS as a secondary

endpoint. However, OS is still of interest, and preferred for the appraisal of treatments[2]. Indeed if OS is not available, analysts will use PFS to estimate OS in appraisals, which increases the uncertainty in any conclusions made. Therefore, methodological solutions which analyse OS and account for subsequent anti-cancer therapy have been proposed.

At PhD outset the methodological solutions recommended focused on the specific scenario of treatment switching, defined as participants in the control arm receiving the experimental intervention as a subsequent anti-cancer therapy, for which there are a number of approaches researchers can implement. Two approaches are to either censor participants at the point of switch or simply exclude those who switch. However, these approaches are acknowledged as simplistic and do not utilise all the information available. Another approach is the inverse probability of censoring weights (IPCW) method, which aims to create a scenario where no switching occurs. The method works by censoring switchers at the point of switch and assigning weights to the remaining individuals. The weights are based on finding characteristics which predict whether a switch will occur. An alternative approach is the two-stage estimation (TSE) method which incorporates an accelerated failure time model to derive counterfactual survival times as if no one switched (see Sections 1.4.3 and 1.4.4). However, both of these methods (IPCW and TSE) require information at the point of switch, which is not always available to researchers. A recent review of these methods found that they are only applicable in certain situations and require a number of assumptions.[3] It concluded that no one approach could be recommended and that a decision on a case-by-case basis was needed but that, where treatment switching occurred, adjustment methods should be considered.

1.2.4 Recent Examples of Subsequent Anti-Cancer Therapy and Clinical Trials

At clinical trial design it may be possible to estimate when a participant may start a subsequent anti-cancer therapy, but it may not be possible to predict the therapy they receive. Consider a clinical trial investigating experimental intervention “E” for newly diagnosed patients for which there are limited effective subsequent treatment options available on relapse. Then the following three scenarios represent plausible changes to a participant’s treatment pathway during a clinical trial.

1. The introduction of a new and improved intervention at a later line of therapy.
2. The introduction or availability of the same intervention “E” at a later line of therapy.

3. A patient's entire treatment pathway consisting of multiple sequential participation in clinical trials.

At PhD outset each of these three scenarios had been observed recently at the Leeds Institute of Clinical Trials Research (LICTR).

Scenario one was observed in an RCT in renal cell carcinoma (RCC). At trial outset there were limited second-line therapies available for patients. However five new therapies were approved by NICE in five years[4], where the new therapies would be available to participants in ongoing clinical trials such as STAR (ISRCTN06473203). This demonstrates how emerging evidence and changing practice impacts ongoing clinical trials. For example, beyond LICTR trials, the approval of sunitinib, sorafenib, and everolimus was reasoned to have impacted the overall survival (OS) results of AVERON[5] and prompted a post-hoc retrospective analysis considering subsequent lines of therapy.[6]

Scenario two has been observed in Myeloma. Myeloma X (ISRCTN60123120) considered the efficacy of a second autologous stem cell transplant (ASCT). However, a number of participants in the control arm received an ASCT as part of their subsequent anti-cancer therapy[7], which is unsurprising given the guidance changed during the course of follow-up recommending a second ASCT post-relapse under certain scenarios.[8] Similarly, in Myeloma XI/XI+ (ISRCTN49407852) the efficacy of continuous lenalidomide therapy was compared with observation. However, continuous lenalidomide is recommended by NICE after two prior lines therapies.[9] Approximately 40% of participants in the control arm went on to receive lenalidomide after relapse, which on publication, was suggested to have impacted on the results of OS.[10]

Scenario three has been observed in myeloma clinical trials conducted at LICTR. Participants in Myeloma XI/XI+ went on to Myeloma XII (ISRCTN10038996) or MUK5 (ISRCTN17354232) for their second-line therapy and some then participated in clinical trials for subsequent lines such as MUK7 (NCT02406222) or MUK4 (NCT01720875). It is thought that ignoring a participant's involvement in a later clinical trial in the analysis of an earlier clinical trial, could confound the intervention effect due to them receiving above NHS standard of care in a later line of therapy.

To investigate the experience of other researchers, at PhD outset three separate scoping reviews were conducted to investigate whether treatment switching approaches were implemented in practice. The scoping reviews were conducted on the 12th May 2020 using Pubmed. Each of the reviews included the filters:

- The title contains overall survival
- Free full text

- Clinical trial
- Randomised controlled trial
- Last ten years

in addition to one of the following relapsing cancers

1. Myeloma
2. Renal cell carcinoma
3. Breast cancer

The cancers for the scoping review were chosen based on prior experience and awareness of existing clinical trial research. This restriction was removed for the systematic literature presented in **Chapter 2**.

Of the six papers identified for myeloma, one was excluded as it referred to primary lung cancer. Four of the five remaining papers acknowledged subsequent anti-cancer therapy as a limitation in the discussion but only one considered OS in the context of subsequent therapies.[11] Twelve papers were identified for renal cancer, five papers were excluded due to not presenting the OS results of a randomised clinical trial. All of the remaining seven papers either mentioned subsequent therapies in the discussion or mentioned subsequent therapies in the discussion and presented the lines of therapy participants received in the results. However, one paper was specifically focused on treatment switching, this paper censored participants at the point of the switch.[6] Finally, twenty-five papers were identified for breast cancer. Of these, seven papers were excluded due to them not presenting the OS results of a randomised clinical trial in breast cancer and an additional seven did not clearly mention subsequent lines in the manuscript. Of the remaining eleven papers, the majority stated subsequent anti-cancer therapy as a reason for a non-significant result. Two papers implemented treatment switching techniques: one paper censored participants at the time of their switch[12] and another applied IPCW.[13]

Therefore, the scoping review concluded that treatment switching methods are not widely used, that subsequent anti-cancer therapy is generally stated as a potential limitation to be considered in interpretation of results and the effect of subsequent anti-cancer therapy is generally not quantified through primary or sensitivity analysis.

1.2.5 PhD Rationale

The first-hand experience of researchers within LICTR, and the scoping review, highlight a potential gap in the literature. The definition of treatment switching at PhD outset introduced in Section 1.2.3 only considers scenario two described in Section 1.2.4, whereas the other two scenarios have been shown to be equally plausible. In addition, the scoping review indicated that researchers are potentially interested in

considering the uptake of subsequent anti-cancer therapy as a whole, rather than the current simple definition of treatment switching.

Methodological research is therefore needed to support the rapidly evolving clinical landscape. At PhD outset there were 250 clinical trials in cancer listed as “ongoing” on the ISCRTN database.[14] In addition, in the year before this PhD grant application was submitted (January 2019 and January 2020), 20 technology appraisal guidelines were updated by NICE to advise on the treatment of cancer.[15] Furthermore, there are calls for this guidance to be published quickly so that patients can get access to new interventions faster as part of the NHS, where, once approved, interventions should be available to patients within three months.[16] Therefore, the variety of clinical trials and availability of approved interventions will only serve to increase the incidence of scenarios one and three (Section 1.2.4). Thus, without methodological research there is a risk that incorrect conclusions on the impact that an experimental intervention has on participant outcomes will be published, and the time and resources it takes for a phase III confirmatory clinical trial to report OS results wasted.

1.3 Aims and Objectives

1.3.1 Aim

The overarching aim of this thesis is to develop new, or extend existing, methodology which assesses OS whilst accounting for subsequent lines of anti-cancer therapy in randomised clinical trials.

1.3.2 Objectives

1. To determine how OS is currently being analysed in the presence of subsequent anti-cancer therapy, identify the key issues with the current approaches and what the main priorities are according to all stakeholders.
2. To build on existing methods, or develop new methodology, to address the main priorities as identified by stakeholders.
3. To assess the methods identified in objective 2 in a simulation study and recent clinical trial data.
4. To ensure that the developed method is available and accessible to all stakeholders.

The following section introduces relevant methodology to meet these objectives which is considered assumed knowledge in Chapters 2,3 and 4.

1.4 Overview of Relevant Methodology

1.4.1 Consensus Methods

Objective 1 is addressed in part by a stakeholder consultation. The stakeholder consultation utilises consensus methodology. Consensus methodology is a type of qualitative research used to find formal agreement between different stakeholder groups. Different types of consensus methods have been implemented within the literature to gain opinions on both clinical and methodological projects.

The three common consensus methods are: Delphi[17], nominal group[18] and RAND/UCLA appropriateness.[19] The Delphi method is an iterative process where participants score different questions until consensus is reached. There is no interaction between participants but the scores of each question can be shared between rounds. One example of how the Delphi method has been used is to determine how patients and the public should be included in the development of statistical analysis plans.[20] The nominal group method is a structured meeting where all participants provide their own thoughts on a topic, any required clarification is obtained, and then all suggestions are ranked or rated as required. The nominal group method has been implemented in the past to obtain stakeholder perspectives on how to improve efficiency in the care of cancer patients.[21] The RAND/UCLA appropriateness method can be seen as a combination of the Delphi and nominal group methods. Participants first complete a questionnaire and then meet to discuss any aspects which did not meet consensus, where rounds can be repeated as required. It has been used previously to determine quality measures in the care of older adults.[22]

The stakeholder consultation implemented a modified RAND/UCLA appropriateness method, which incorporated both an online questionnaire and focus group to obtain consensus on the project direction and associated topics with the different stakeholder groups. The stakeholder consultation is described in Chapter 3.

1.4.2 Estimands

The findings of the literature review and stakeholder consultation were combined to understand current practices and opinions on the uptake of subsequent anti-cancer therapy as well as derive the estimand of interest.

The estimand framework is a systematic approach for researchers to clearly document and explain what they are investigating in a clinical trial, and has been incorporated into regulatory guidance.[23, 24] The framework includes five components:

1. Treatment Conditions: The treatment strategies / interventions being compared.

2. Population: The population of patients targeted by the question.
3. Endpoint: The outcome measure for each participant in the clinical trial.
4. Population Level Summary: The summary measure used to compare the outcomes for each participant.
5. Intercurrent Events: Post-randomisation events which affect the interpretation or occurrence of the outcome.

There are five different methods to address intercurrent events:

1. Treatment Policy: The intercurrent event is considered part of the treatment strategy.
2. Hypothetical: The treatment effect for the hypothetical scenario where the intercurrent event is assumed to not have occurred is estimated.
3. Principal Stratum: The treatment effect in which the intercurrent event would (or would not) occur is estimated.
4. Composite: The intercurrent event is incorporated into the endpoint definition.
5. While on treatment / alive: The endpoint prior to treatment discontinuation is of interest.

The choice of a method to address an intercurrent event impacts the research question being investigated and therefore the treatment effect being estimated. For example, considering the topic of this thesis, trial treatment discontinuation and the uptake of subsequent anti-cancer therapy is an intercurrent event and the research question being answered changes depending on the approach taken (Table 1.1).

Table 1.1: The Impact of Different Intercurrent Event Approaches on the Research Question

Approach	Research Question¥
Treatment Policy	How does the experimental treatment extend overall survival compared to the control treatment; regardless of whether patients receive subsequent anti-cancer therapy?
Hypothetical	How would the experimental treatment extend overall survival compared to the control treatment, if no one stopped their trial treatment prior to death?
Principal Stratum	How does the experimental treatment extend overall survival compared to the control treatment, in patients who do not require subsequent anti-cancer therapy?
Composite	How does the experimental treatment extend the time to the requirement of subsequent anti-cancer therapy or death from any cause compared to the control treatment?
While on treatment / alive	How does the experimental treatment extend the time to treatment discontinuation or death from any cause compared to the control treatment?

¥Note that the question specified focuses on the impact of the intercurrent event. A complete research question should also include the population of interest, clear details of the experimental and control treatments as well as the population level summary measure.

Following the literature review and stakeholder consultation an alternative hypothetical estimand was identified. This is detailed in Chapter 4.

1.4.3 Time-to-Event Analysis

As previously introduced, the topic under consideration in this thesis is the analysis of OS. To analyse OS, statisticians and other data analysts use time-to-event analysis methods.[25]

Time-to-event analysis, also referred to as survival analysis, describes statistical methods which consider not only whether an individual has experienced the event of interest e.g. death, but also when it occurred in the context of clinical trial follow-up e.g. time from randomisation. Individuals for whom it is unknown when the event occurred are “censored” at the time they were last known to be event-free. For example, an individual withdrawing from further data collection, i.e. they are “lost-to-follow-up”, is an example of right censoring which is the type of censoring considered in this thesis.

In general, time-to-event analysis methods consider the survivor function (the probability an individual is event-free past a certain time) and the hazard function (the probability an individual experiences the event of interest at a particular time, provided the individual has reached that time, expressed as a rate). In order to compare two groups of individuals e.g. clinical trial arms, using either the survivor or hazard function, one group is assigned as the reference group to which a group of interest is being compared to e.g. the experimental intervention vs the control intervention in a clinical trial. A time ratio can be used to compare two survivor functions. A time ratio greater than one indicates that it takes longer for the group of interest to experience the event of interest compared to the reference group. A hazard ratio is used to compare two hazard functions. A hazard ratio greater than one indicates the group of interest has an increased hazard of the event of interest compared to the reference group.

One common method used to estimate and visualise the survivor function is the Kaplan-Meier survivor function. This method places no distributional assumptions on the survivor function. Instead it considers the number of individuals who could have an event at a particular time and compares that with the number of individuals who do have an event at that time to obtain the proportion of individuals event-free at each time (Equation 1.1). In randomised clinical trials, the Kaplan-Meier survival function is often estimated separately for each clinical trial arm.

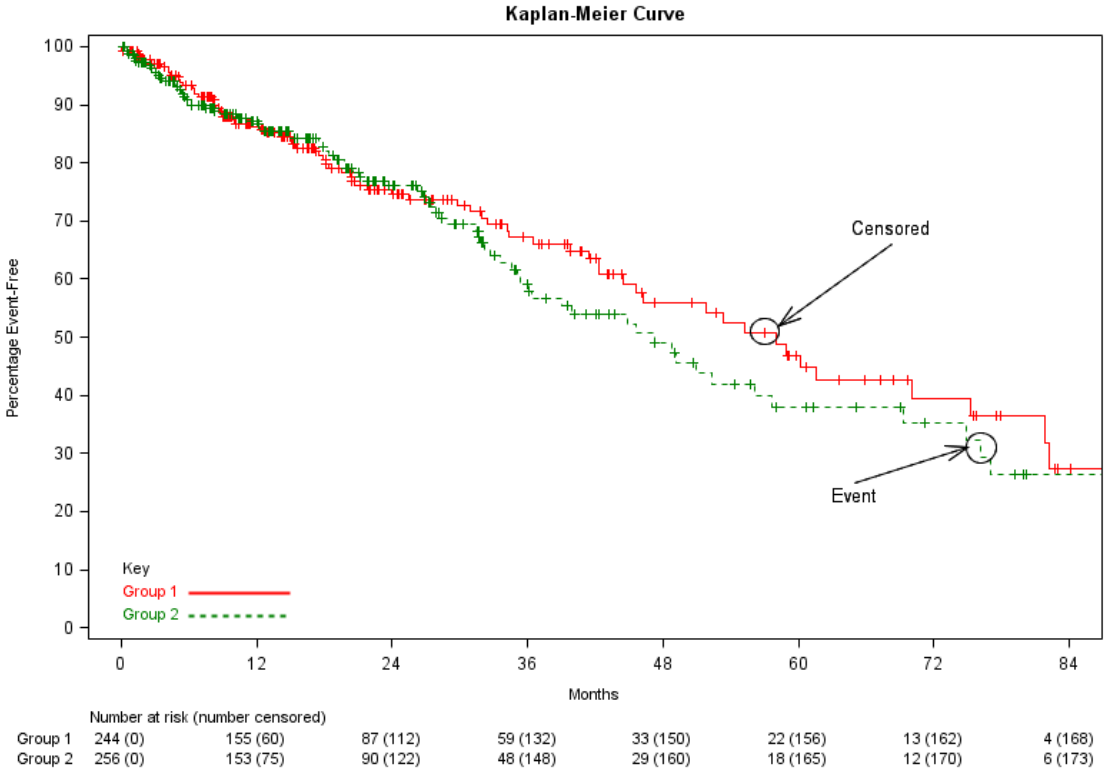
Equation 1.1: Kaplan-Meier Estimate of the Survivor Function

Suppose N individuals in a clinical trial are followed-up from time t_o to time T and D of the N experience the event of interest at $t_1, t_2, \ldots t_D$, where the number of events at t_i is denoted by d_i and the number of individuals event-free and at risk just before t_i is denoted by n_i. Then, assuming all events are independent, the Kaplan-Meier survival function is given by:

$$\hat{S}(t) = \prod_{i=1}^{D} \frac{(n_i - d_i)}{n_i}$$

The Kaplan-Meier estimate is a step function which can be used to describe a time-to-event outcome. A commonly used descriptive measure is the median time i.e. the time at which 50% of participants are event-free. The step-function can also be plotted to visually display the estimated survivor function over time. This is more commonly referred to as a Kaplan-Meier curve (Figure 1.3). Kaplan-Meier curves may also include: a risk table which details, the number at risk (n_i) at each time (t_i); annotations of when individuals are censored and confidence intervals. However, a recent survey found that opinions on incorporating these extensions vary.[26]

Figure 1.3: Example of a Kaplan-Meier Curve



Whilst the Kaplan-Meier curve can be used to visualise differences between different groups of individuals. Formal hypothesis testing requires alternative methods. One such test is the logrank test. The null hypothesis is that there is no difference in the hazard of an event between two groups. The alternative hypothesis is that there is a difference in the hazard of an event between the two groups. The test statistic is calculated (Equation 1.2) and compared to an appropriate critical value.[25] As a result of the skewed nature of the chi-squared distribution to implement a two-sided test at the α% significance level, the test statistic must be above the $1-\alpha$ upper-tail critical value for a conclusion of statistical significance. For example a two-sided test at a 5% level of significance comparing two randomisation arms would require a test statistic above 3.84 to be concluded as statistically significant.

Equation 1.2: Logrank Test Statistic

Suppose N individuals in a clinical trial are followed-up from time t_o to time T, and D of the N experience the event of interest at $t_1, t_2, \ldots t_D$. Let the number of events at t_i be denoted by d_i, and the number of individuals event-free and at risk just before t_i be denoted by n_i. Then, suppose the individuals are split into two groups with $n_{1i}\, n_{2i}, d_{1i}, d_{2i}$. Then assuming all events are independent, the logrank test statistic is given by:

$$\frac{\left\{\sum_i \left(d_{1i} - \frac{n_{1i}d_i}{n_i}\right)\right\}^2}{\sum_i \frac{n_{1i}n_{2i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}} \sim \chi_1^2$$

where χ_1^2 denotes the chi-squared distribution with one degree of freedom.

The logrank test is a type of univariable analysis. There may be other explanatory variables which are thought to affect the outcome of interest as well as potentially the explanatory variable of interest itself e.g. stratification factors in a randomised controlled clinical trial. It is recommended in clinical trial guidelines that a clinical trial which uses stratification factors should account for these additional explanatory variables, sometimes also referred to as confounders, in their analysis.[1] In this context, the inclusion of additional explanatory variables aims to increase precision and power to ensure that any significant results are due to the explanatory variable of interest, and not because another variable correlated with the outcome of interest has not been considered. Any analysis which includes other explanatory variables (covariates, confounders, predictors, independent variables) is called adjusted or multivariable. Note that these additional variables may vary over time. In this case they are referred to as time dependent variables or time dependent covariates. For example, in cancer, serial tumour measurements, may be considered time dependent variables.

In time-to-event analysis, most courses introduce the Cox proportional hazards model as a method to conduct adjusted analysis. The Cox proportional hazards model assumes that the hazard function for an individual can be written as a multiplicative product of a function of the explanatory variables and a baseline hazard function. The baseline hazard function is the hazard when all explanatory variables are zero. Hazard ratios are estimated from the model and used to compare between groups. An alternative method for conducting adjusted analysis is the accelerated failure time (AFT) model. The AFT model does not assume proportional hazards. However, it does assume that explanatory variables have a multiplicative effect on an individual's survival time. It also requires distributional assumptions to be placed on the observed survival times. In the AFT model, time ratios are estimated from the model and used to

compare between groups. The methodology behind both of these models is detailed in Chapter 4.

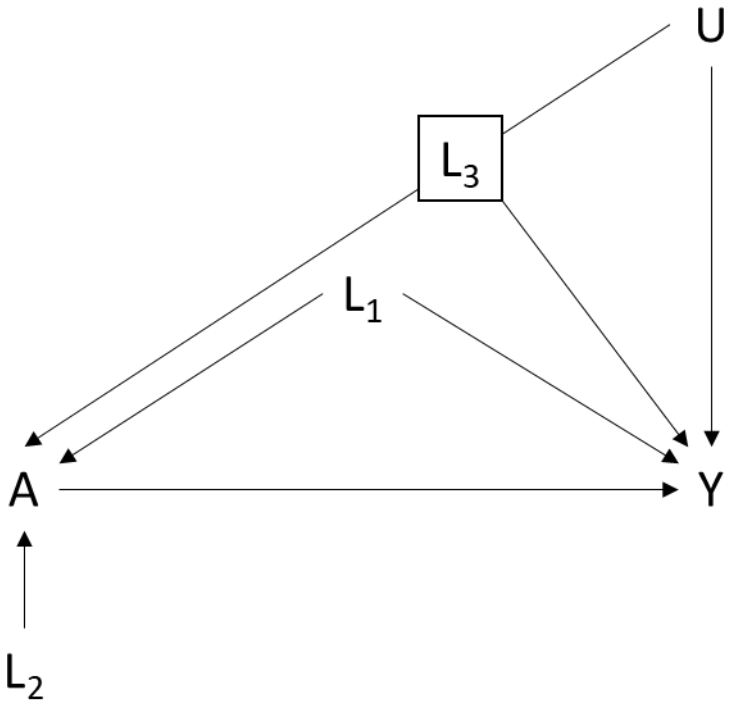
The commonality of the different time-to-event analysis methods being implemented for OS within the literature is described in Chapter 2. As part of the stakeholder consultation, stakeholders were also asked their experience and opinion of various methods and associated assumptions. This is described in Chapter 3. Chapter 4 implements the methods.

1.4.4 Causal Inference

As stated earlier, the setting of this thesis is randomised cancer clinical trials. In randomised clinical trials we are investigating whether an experimental intervention has an effect on participant outcomes compared to a control intervention. This investigation can be seen as the second “rung” of the causal ladder which was introduced by Pearl[27] as a way to explain causal inference. It is one “rung” above simply observing correlations, researchers are intervening to identify whether doing something different changes an outcome i.e. is there a causal relationship.

Pearl, along with colleagues Greenland and Robins, also introduced how to visualise causal relationships using directed acyclic graphs (DAGs) (Figure 1.4).[28] Directional arrows are drawn or omitted to show how one variable is assumed to influence or not influence another. It may be that the relationship between one action (A) and outcome (Y) are of interest, but other variables (L) or unknowns (U) are known to cause the action and outcome of interest. DAGs are used to determine what statistical methods are most appropriate. For example, in Figure 1.4, L_1 is an example of a confounder, as it directionally affects both the action (A) and outcome (Y), which should be controlled for in any analysis e.g. by stratification in randomised trials and/or multivariable modelling so that the true causal relationship of A on Y can be estimated. However, the DAG also contains the unknown U which also should be controlled for. Unfortunately, this is not measured. This scenario is known as unmeasured confounding. In this scenario researchers aim to identify alternative confounders (L_3) which can be measured to represent the unmeasured confounder. For example, it may be that the presence of a genetic mutation influences an individual’s response to treatment and their chance of survival, but genome sequencing is not available. Instead, family history of a condition related to the mutation could be collected and incorporated into the analysis. In Figure 1.4, L_2 is an example of a variable assumed to influence A but not Y and as such does not need to be considered in the estimation of the causal relationship between A and Y. DAGs can also be used in other contexts to show how data is generated.

Figure 1.4: Example of a Directed Acyclic Graph (DAG)



Randomised clinical trials are the gold standard for determining whether causal relationships exist. However, as indicated above, intercurrent events such as treatment discontinuation and the uptake of subsequent anti-cancer therapy can affect the interpretation of the results or occurrence of the outcome of interest. The third and final "rung" of the causal ladder, utilised when considering the hypothetical strategy of dealing with intercurrent events, is counterfactuals, i.e. what would have happened to a participant in a different scenario than to that which truly happened. A causal inference approach considering counterfactuals is adopted throughout this thesis.

1.4.5 Simulation Studies

The estimand identified from the literature review and stakeholder consultation was used to meet Objectives 2 and 3 through a simulation study.

A simulation study is a structured approach to determine how well a statistical method performs in different scenarios. They have been used in past methodological research to consider; approaches to handling missing data[29], sample size estimation[30] and the development of methods to consider treatment switching.[31-33] A simulation study generally implemented by:

1. Creating data based on a set of assumptions and a known parameter of interest (the true estimand). This can be visualised using a DAG. To generate the data, simulation studies rely on the computer generating a sequence of pseudo-random numbers. The starting value of this sequence can be fixed by setting a random number seed. The value of the sequence after a certain number of iterations can be obtained using a second sequence called the random number generator (RNG) state.
2. Applying different methods to the data created in step one and using them to estimate the parameter of interest.
3. Repeating steps one and two for a pre-specified number of replications (also referred to as repetitions).
4. Assessing the performance of the methods considering all iterations (Table 1.2). The choice of performance measures depends on the aim and objective of the simulation study. For example, in a randomised clinical trial; power, coverage and bias are important. Due to the reliance on random number generation uncertainty in the performance measures is quantified using the Monte-Carlo standard error.

This implementation can be structured using the ADEMP framework.[34] The ADEMP framework considers the **A**im of the simulation study, how the **D**ata will be generated, the **E**stimand of interest, the **M**ethods to be applied in the simulation study and the **P**erformance measures which will be used to compare the methods.

Table 1.2: Different Performance Measures in a Simulation Study

Performance Measure	Definition	Derivation	Monte-Carlo Standard Error¥
Bias	The difference between the average estimate over the simulations and the true estimand. A value of zero indicates a non-biased estimate.	$\widehat{Bias} = \frac{1}{N}\sum_i (\hat{\theta}_i - \theta)$	$\sqrt{\frac{1}{N(N-1)}\sum_i (\hat{\theta}_i - \bar{\theta})^2}$
Coverage	The percentage of simulations that the true estimand lies within the upper and lower bound of the confidence interval of the estimate. A higher percentage indicates good coverage.	$\widehat{Cover} = \frac{1}{N}\sum_i I(\hat{\theta}_{i.low} \leq \theta \leq \hat{\theta}_{i.high})$	$\sqrt{\frac{\widehat{Cover} \times (1 - \widehat{Cover})}{N}}$
Power (Rejection Percent)	The percentage of simulations that the p-value for the estimate is concluded to be statistically significant. A higher percentage indicates a higher power.	$\widehat{Power} = \frac{1}{N}\sum_i I(p_i \leq \alpha)$	$\sqrt{\frac{\widehat{Power} \times (1 - \widehat{Power})}{N}}$
Empirical Standard Error	A measure of how much each estimate from each repetition differs from the average of all the estimates across the repetitions. Lower values indicate that on average the estimated values are closer to the average estimate i.e. reduced spread across repetitions.	$\widehat{EmpSE} = \sqrt{\frac{1}{N-1}\sum_i (\hat{\theta}_i - \bar{\theta})^2}$	$\frac{\widehat{EmpSE}}{\sqrt{2(N-1)}}$
Mean Square Error	The average of the difference between the true estimand and the estimate from each repetition squared. Lower values indicate that the methods are good at estimating the true estimand	$\widehat{MSE} = \frac{1}{N}\sum_i (\hat{\theta}_i - \theta)^2$	$\sqrt{\frac{\sum_i \left[\left(\hat{\theta}_i - \theta\right)^2 - \widehat{MSE}\right]^2}{N(N-1)}}$

Performance Measure	Definition	Derivation	Monte-Carlo Standard Error¥
Average Model Standard Error	The square root of the sum of the estimated variance averaged across all repetitions Lower values indicate reduced variation within the models.	$\widehat{ModSE} = \sqrt{\frac{1}{N}\sum_i \widehat{Var}(\hat{\theta}_i)}$	$\sqrt{\frac{\widehat{Var}\left(\widehat{Var}(\hat{\theta})\right)}{4N \times \widehat{ModSE}^2}}$
Relative Error	The average model standard error compared to the average empirical standard error on the percent scale. Negative values indicate a higher empirical standard error.	$\widehat{RelError} = 100\left(\frac{\widehat{ModSE}}{\widehat{EmpSE}} - 1\right)$	$100\left(\frac{\widehat{ModSE}}{\widehat{EmpSE}}\right)\sqrt{\frac{\widehat{Var}\left(\widehat{Var}(\hat{\theta})\right)}{4N \times \widehat{ModSE}^4} + \frac{1}{2(N-1)}}$
Relative Precision	The average empirical variance of a reference model e.g. ITT, compared to the average empirical variance of the alternative method in question on the percent scale. A higher value indicates increased precision (smaller empirical variance) compared to the reference model.	$\widehat{RelPrec} = 100\left(\left(\frac{\widehat{EmpSE}_{ref}}{\widehat{EmpSE}_{model}}\right)^2 - 1\right)$	$200\left(\frac{\widehat{EmpSE}_{ref}}{\widehat{EmpSE}_{model}}\right)^2\sqrt{\frac{1 - Corr\left(\hat{\theta}_{ref}, \hat{\theta}_{model}\right)^2}{N-1}}$
Bias-Eliminated Coverage	The percentage of simulations that the average of the estimates across the repetitions lies within the upper and lower bound of the confidence interval of the estimate. A higher percentage indicates good coverage.	$\widehat{Cover.B} = \frac{1}{N}\sum_i I(\hat{\theta}_{i.low} \leq \bar{\theta} \leq \hat{\theta}_{i.high})$	$\sqrt{\frac{\widehat{Cover.B} \times (1 - \widehat{Cover.B})}{N}}$

θ=True value of the estimand; $\hat{\theta}_i$=Value of the estimate from the i^{th} repetition; N = Number of repetitions; α = Nominal significance level; p_i=the p-value of the significance test for $\hat{\theta}_i$. $\widehat{Var}(\hat{\theta}_i)$= Variance of $\hat{\theta}_i$, $\bar{\theta} = \frac{1}{N}\sum\hat{\theta}_i$, ¥ = Monte-Carlo Standard Errors are approximate for relative precision, average model standard error and relative error; $\widehat{Var}\left(\widehat{Var}(\hat{\theta})\right) = \frac{1}{N-1}\sum_i\{\widehat{Var}(\hat{\theta}_i) - \frac{1}{N}\sum_j \widehat{Var}(\hat{\theta}_j)\}^2$.

Recent guidance has been published on how to use simulation studies to evaluate statistical methods[34] and also check a simulation study to ensure it has been implemented as expected.[35] Some of the key recommendations are:

1. To pre-specify and save the random number sequences and associated RNG stream state so that the results are reproducible.
2. To produce descriptive statistics to ensure the data is being generated as expected.
3. To conduct the analysis on a small number of repetitions first to check for any implementation errors, specifically to make sure different results are being generated each time.
4. To use graphical methods to determine any outliers which require further investigation.

These recommendations were used to design the simulation study implemented in this thesis. In the simulation study the performance of existing time-to-event analysis methods, used for treatment switching, to estimate the estimand of interest were assessed based on bias, coverage and power. This is described in Chapter 4.

1.4.6 Patient and Public Involvement

In order to ensure all information is disseminated to, interpretable and of interest to all stakeholder groups, patient and public involvement (PPI) was embedded throughout the PhD.

The National Institute for Health and Care Research (NIHR) define public involvement as "*research being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them*." [36] i.e. PPI describes the inclusion of members of the public with or without lived experience of the research topic in the running of the research. It is thought to make the research more accessible to the public audience and ensure that the patient voice is considered throughout the project. PPI members may: design and input into research proposals considering whether the research questions are of interest to patients; review documentation intended for a patient audience to assess whether the language is appropriate and attend meetings to discuss decisions which affect patients or the project outcomes. The practice of PPI is endorsed by funders and researchers in all areas of health and social care research.[37]

Therefore, in addition to the usual supervisory requirements for a PhD this work was also supported by a stakeholder advisory group (SAG) which included representatives from; healthcare, PPI, National Institute for Heath and Care Excellence (NICE) and health economics. The SAG primarily supported the development of the online questionnaire and focus group materials. Further support was obtained from a co-

design group of five PPI representatives who inputted into the design, format and contents of the online questionnaire. Throughout the research, newsletters were sent to the SAG, co-design group and consenting questionnaire participants to provide updates and allow for comments.

1.5 Thesis Summary

The remaining chapters of this thesis detail the systematic literature review (Chapter 2), stakeholder consultation (Chapter 3), simulation study and real data example (Chapter 4). These three chapters are then brought together and considered in the context of the current research landscape (Chapter 5). The strengths and limitations of the work overall are also discussed; including the availability and accessibility of the method and data required (Objective 4) along with areas for future research.

1.6 Chapter 1 References

1. EMA. ICH Topic E9 Statistical Principles for Clinical Trials, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf (1998, accessed 6th September 2022).
2. Davis S, Tappenden P and Cantrell A. NICE Decision Support Unit Methods Development. *A Review of Studies Examining the Relationship between Progression-Free Survival and Overall Survival in Advanced or Metastatic Cancer*. London: National Institute for Health and Care Excellence (NICE) Copyright © 2012 National Institute for Health and Clinical Excellence, unless otherwise stated. All rights reserved., 2012.
3. Latimer NR and Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
4. Tannir NM, Pal SK and Atkins MB. Second-Line Treatment Landscape for Renal Cell Carcinoma: A Comprehensive Review. *The Oncologist* 2018; 23: 540-555. DOI: 10.1634/theoncologist.2017-0534.
5. Escudier B, Bellmunt J, Négrier S, et al. Phase III Trial of Bevacizumab Plus Interferon Alfa-2a in Patients With Metastatic Renal Cell Carcinoma (AVOREN): Final Analysis of Overall Survival. 2010; 28: 2144-2150. DOI: 10.1200/jco.2009.26.7849.
6. Bracarda S, Bellmunt J, Melichar B, et al. Overall survival in patients with metastatic renal cell carcinoma initially treated with bevacizumab plus interferon-α2a and subsequent therapy with tyrosine kinase inhibitors: a retrospective analysis of the phase III AVOREN trial. 2011; 107: 214-219. DOI: 10.1111/j.1464-410X.2010.09707.x.
7. Cook G, Royle K-L, O'Connor S, et al. The impact of cytogenetics on duration of response and overall survival in patients with relapsed multiple myeloma (long-term follow-up results from BSBMT/UKMF Myeloma X Relapse [Intensive]): a randomised, open-label, phase 3 trial. *British journal of haematology* 2019; 185: 450-467.
8. NICE. Myeloma: diagnosis and management https://www.nice.org.uk/guidance/ng35/resources/myeloma-diagnosis-and-management-pdf-1837394042821 (2016, accessed 9th April 2026).
9. NICE. Lenalidomide for the treatment of multiple myeloma in people who have received at least 2 prior therapies Technology appraisal guidance [TA171], https://www.nice.org.uk/guidance/ta171 (2009, accessed 29th July 2020).
10. Jackson GH, Davies FE, Pawlyn C, et al. Lenalidomide maintenance versus observation for patients with newly diagnosed multiple myeloma (Myeloma XI): a multicentre, open-label, randomised, phase 3 trial. *The Lancet Oncology* 2019; 20: 57-73. 2018/12/14. DOI: 10.1016/S1470-2045(18)30687-9.

11. Miguel JFS, Schlag R, Khuageva NK, et al. Persistent Overall Survival Benefit and No Increased Risk of Second Malignancies With Bortezomib-Melphalan-Prednisone Versus Melphalan-Prednisone in Patients With Previously Untreated Multiple Myeloma. *JCO* 2013; 31: 448-455. DOI: 10.1200/jco.2012.41.6180.
12. Diéras V, Miles D, Verma S, et al. Trastuzumab emtansine versus capecitabine plus lapatinib in patients with previously treated HER2-positive advanced breast cancer (EMILIA): a descriptive analysis of final overall survival results from a randomised, open-label, phase 3 trial. *The Lancet Oncology* 2017; 18: 732-742.
13. Colleoni M, Giobbie-Hurder A, Regan MM, et al. Analyses adjusting for selective crossover show improved overall survival with adjuvant letrozole compared with tamoxifen in the BIG 1-98 study. *J Clin Oncol* 2011; 29: 1117.
14. ISCRTN. ISCRTN Registry, https://www.isrctn.com/search?q=&filters=trialStatus%3AOngoing%2CconditionCategory%3ACancer&searchType=basic-search (accessed 29th July 2020).
15. NICE. Guidance and Advice List, https://www.nice.org.uk/guidance/published?type=ta&fromdate=01%20January%202019&todate=31%20January%202020&title=cancer (accessed 29th July 2020).
16. NICE. Cost comparison. Addendum to the Guide to the methods of technology appraisal https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/methods-guide-addendum-cost-comparison.pdf (accessed 29th July 2020).
17. Linstone HA and Turoff M. *The delphi method*. Addison-Wesley Reading, MA, 1975.
18. Delbecq AL, Van de Ven AH and Gustafson DH. Group techniques for program planning: A guide to nominal group and Delphi processes. 1975.
19. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA appropriateness method user's manual*. 2001. Rand Corp Santa Monica CA.
20. Goulão B, Morris TP, Blazeby J, et al. Involving patients and the public In sTatistIcal Analysis pLans (INITIAL): A delphi survey. *PLoS ONE* 2023; 18: e0292257. DOI: 10.1371/journal.pone.0292257.
21. Ferrara L, Otto M, Aapro M, et al. How to improve efficiency in cancer care: Dimensions, methods, and areas of evaluation. *Journal of Cancer Policy* 2022; 34: 100355. DOI: https://doi.org/10.1016/j.jcpo.2022.100355.
22. Correia RH, Dash D, Jones A, et al. Primary care quality for older adults: Practice-based quality measures derived from a RAND/UCLA appropriateness method study. *PLoS ONE* 2024; 19: e0297505. DOI: 10.1371/journal.pone.0297505.
23. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (2017, accessed 1st March 2023).
24. FDA. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical (2021, accessed 17th June 2025).
25. Collett D. *Modelling survival data in medical research*. Chapman and Hall/CRC, 2023.
26. Morris TP, Jarvis CI, Cragg W, et al. Proposals on Kaplan–Meier plots in medical research and a survey of stakeholder views: KMunicate. *BMJ Open* 2019; 9: e030215. DOI: 10.1136/bmjopen-2019-030215.
27. Pearl J and Mackensie D. The Ladder of Causation. *The book of why: The new science of cause and effect*. Penguin, 2018.
28. Greenland S, Pearl J and Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999; 10: 37-48.

29. Sullivan TR, White IR, Salter AB, et al. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res* 2018; 27: 2610-2626. 20161219. DOI: 10.1177/0962280216683570.
30. Pate A, Riley RD, Collins GS, et al. Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Statistical Methods in Medical Research* 2023; 32: 555-571. DOI: 10.1177/09622802231151220.
31. Morden JP, Lambert PC, Latimer N, et al. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Medical Research Methodology* 2011; 11: 4. DOI: 10.1186/1471-2288-11-4.
32. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting for treatment switching in randomised controlled trials - A simulation study and a simplified two-stage method. *Stat Methods Med Res* 2017; 26: 724-751. 20141121. DOI: 10.1177/0962280214557578.
33. Latimer NR, Abrams KR, Lambert PC, et al. Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study. *Statistical Methods in Medical Research* 2018; 27: 765-784. DOI: 10.1177/0962280216642264.
34. Morris TP, White IR and Crowther MJ. Using simulation studies to evaluate statistical methods. 2019; 38: 2074-2102. DOI: 10.1002/sim.8086.
35. White IR, Pham TM, Quartagno M and Morris TP. How to check a simulation study. *International Journal of Epidemiology* 2023; 53. DOI: 10.1093/ije/dyad134.
36. NIHR. Public involvement in research, https://www.nihr.ac.uk/get-involved/public-involvement (accessed 13th January 2026).
37. HRA. Putting people first - embedding public involvement in health and social care research, https://www.hra.nhs.uk/planning-and-improving-research/best-practice/public-involvement/putting-people-first-embedding-public-involvement-health-and-social-care-research/ (2025, accessed 17th June 2025).

Chapter 2 How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review

Kara-Louise Royle[1], David Meads[2], Jennifer K. Visser-Rogers[3], Ian R White[4], David A Cairns[1]

1. Leeds Cancer Research UK Clinical Trials Unit, Leeds Institute of Clinical Trials Research, University of Leeds, UK

2. Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

3. PHASTAR, UK

4. MRC Clinical Trials Unit at UCL, London, UK

Study 1	
Journal	Trials
Submission Status	Published in November 2023
Reference	Royle KL, Meads D, Visser-Rogers JK, White IR, Cairns DA. How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review. Trials. 2023 Nov 6;24(1):708 https://doi.org/10.1186/s13063-023-07730-1

2.1 Abstract

2.1.1 Background

Overall survival is the “gold standard” endpoint in cancer clinical trials. It plays a key role in determining the clinical- and cost-effectiveness of a new intervention and whether it is recommended for use in standard of care. The assessment of overall survival usually requires trial participants to be followed up for a long period of time. In this time, they may stop receiving the trial intervention and receive subsequent anti-cancer treatments, which also aim to extend survival, during trial follow-up. This can potentially change the interpretation of overall survival in the context of the clinical trial. This review aimed to determine how overall survival has been assessed in cancer clinical trials and whether subsequent anti-cancer treatments are considered.

2.1.2 Methods

Two searches were conducted using MEDLINE within OVID© on the 9th of November 2021. The first sought to identify papers publishing overall survival results from randomised controlled trials in eight reputable journals and the second to identify

papers mentioning or considering subsequent treatments. Papers published since 2010 were included if presenting or discussing overall survival in the context of treating cancer.

2.1.3 Results

One hundred and thirty-four papers were included. The majority of these were presenting clinical trial results (98, 73%). Of these, 45 (46%) reported overall survival as a (co-) primary endpoint. A lower proportion of papers including overall survival as a (co)-primary endpoint compared to a secondary endpoint were published in recent years. The primary analysis of overall survival varied across the papers. Fifty-nine (60%) mentioned subsequent treatments. Seven papers performed additional analysis, primarily when patients in the control arm received the experimental treatment during trial follow-up (treatment switching).

2.1.4 Discussion

Overall survival has steadily moved from being the primary to a secondary endpoint. However, it is still of interest with papers presenting overall survival results with the caveat of subsequent treatments, but little or no investigation into their effect. This review shows that there is a methodological gap for what researchers should do when trial participants receive anti-cancer treatment during trial follow-up. Future research will identify the stakeholder opinions, on how this methodological gap should be addressed.

2.1.5 Funding

NIHR301510

2.2 Key Words

Systematic review, overall survival, randomised controlled trial, cancer, subsequent treatment lines

2.3 Introduction

Overall survival (OS), disease-free survival, progression-free survival and cause-specific survival are all used to assess the clinical effectiveness of experimental treatments in cancer clinical trials. However, due to subjectivity and variability in the definitions of response, progression, and cause of death, OS is the only endpoint that has a definite event which cannot be questioned. As a result of this, OS is often referred to as the “gold standard” endpoint for assessing the clinical effectiveness of experimental interventions for the treatment of cancer. In England, OS is used in

conjunction with other evidence in appraisals by NICE (National Institute for Health and Care Excellence) to determine whether interventions should be recommended for use in the NHS. A recent review of Health Technology Assessment (HTA) systems in the European Union, UK, Canada and Australia, found that the model in England which uses clinical and cost-effectiveness data in their review alongside other evidence is similar to eight other HTA bodies including France and Sweden. In addition, there were 46 other HTA bodies across multiple countries including Italy, The Netherlands and Spain which also use both clinical and cost-effectiveness data in their appraisals.[1]

To ensure consistency between clinical trials, guidance exists on their analysis and interpretation. The International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) topic E9 is entitled “Statistical Principles for Clinical Trials” and includes, for example, advice on how to analyse a trial when stratification factors are used during randomisation.[2] In addition, for time to event endpoints like OS, a stakeholder consultation on presenting results using Kaplan-Meier plots was published in 2019.[3] Finally, OS itself has the published definitive definition as time from randomisation to death from any cause.[4]

The assessment of OS requires trial participants to be followed up for what can be a considerable length of time. This poses a problem when it comes to interpreting the trial results, as participants may discontinue their trial intervention prior to death. For example, participants may discontinue at disease progression or earlier due to withdrawal from trial treatment (as a result of toxicity or patient choice), or an intervention stopping rule. Trial participants may then go on to receive multiple subsequent lines of anti-cancer treatment during trial follow-up. These subsequent treatments will also aim to control a patient’s cancer and therefore may improve OS. This improvement may reduce the number of OS events observed in the trial and in turn reduce the power available to detect a difference in OS between the experimental and control interventions. Consequentially, using England as an example, this could lead to a beneficial intervention not being recommended for use in the NHS, resulting in sub-optimal patient outcomes.

This issue with interpretation is well known and is referenced in clinical trial recommendations.[4] Methodological solutions exist to account for scenarios where participants in the control arm switch to receive the experimental intervention, by design or in a later line of therapy (treatment switching). These methods were summarised as part of a technical support document in 2014 and include the rank preserving structural failure time model (RPSFTM) and the Inverse Probability Censoring Weights (IPCW) method.[5]

It is currently unknown how widely the published guidance on the analysis of OS is followed and how researchers at present perceive subsequent treatment lines as a whole compared to the specific scenario of treatment switching. Therefore, the aim of this systematic review was to determine how OS is currently analysed in practice and whether subsequent lines of treatment are mentioned in the main manuscript or supplementary material of cancer clinical trials publications.

2.4 Methods

2.4.1 Search strategy

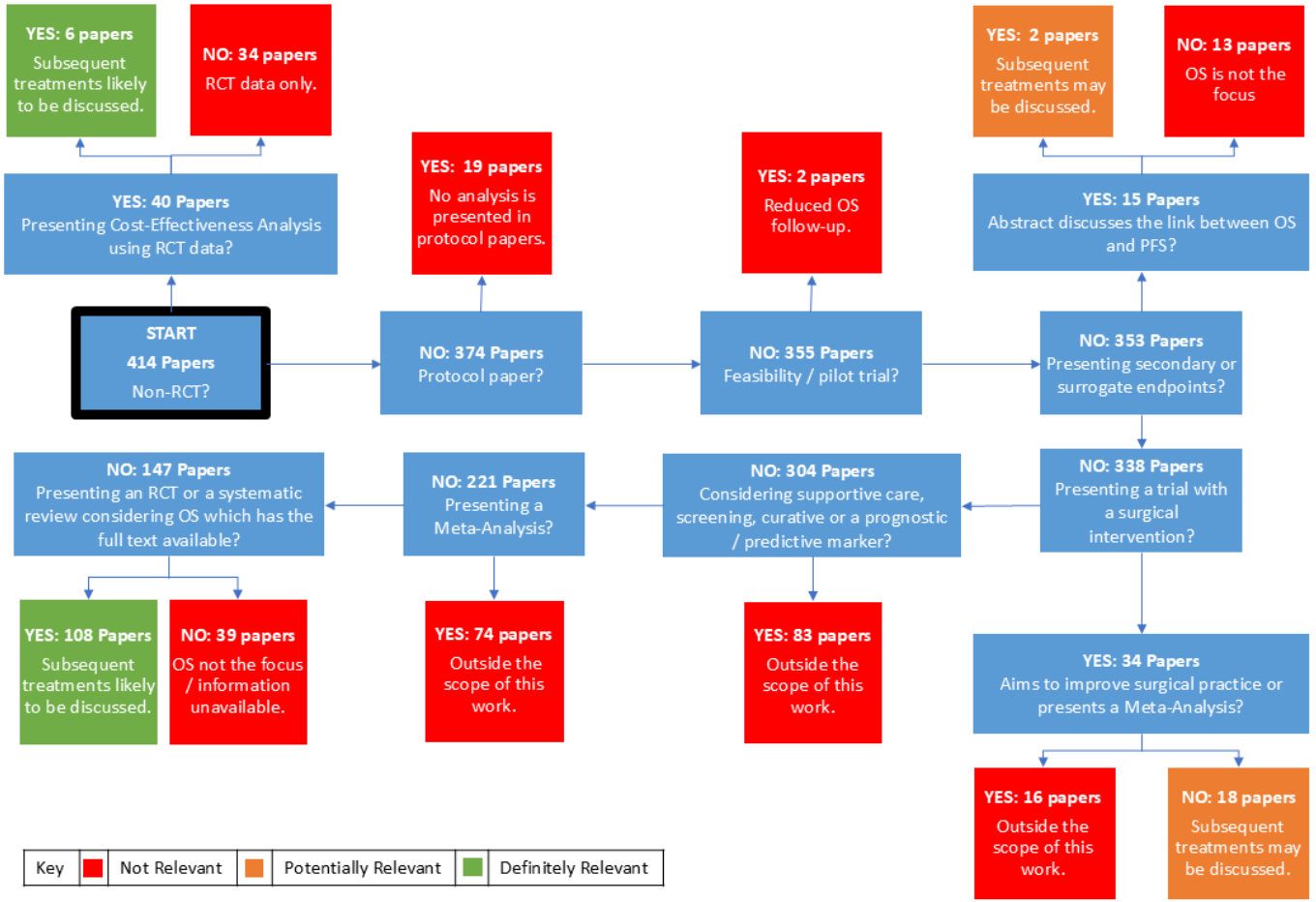
The electronic database MEDLINE using OVID® was searched on 9 November 2021. To ensure that the search was not biased towards trials that did or did not report the limitations surrounding the analysis of OS in terms of subsequent treatments, two approaches were taken (Table 2.1). The first approach aimed to gain a broad view of the analysis of OS in cancer clinical trials. Due to the breadth of publications available, the search was restricted to several major journals focusing on general medicine and cancer: Lancet, Lancet Oncology, JCO, PLOS One, PLOS Medicine, BMJ, JAMA and New England Journal of Medicine. The second approach aimed to target papers which were specifically discussing the issue around subsequent treatments in conjunction with the analysis of OS in cancer clinical trials. Both searches were restricted to the English language and to papers published since 1st January 2010. The results of the two searches were combined with the Boolean operator OR and all abstracts were extracted into Microsoft Excel (Microsoft, Redmond, Washington, USA) and Endnote™ (Clarivate, Philadelphia, Pennsylvania, USA).

Table 2.1: Literature Search Terms

Approach 1 – Overall Survival				
Concept	"Overall Survival".tw.	Randomi?ed Controlled Trial.tw.	Cancer.tw.	
Synonyms	OS mortality	RCT	Carcinoma Neoplasm Sarcoma	
MESH	"Survival Analysis"	Randomized Controlled Trials as Topic	Neoplasms	
Approach 2 – Subsequent Treatment Lines				
Concept	"Overall Survival".tw.	Randomi?ed Controlled Trial.tw.	"Subsequent Treatments"	Cancer.tw.
Synonyms	OS mortality	RCT	(Line* adj3 therap*) (Line* adj3 treatment*) (Subsequent adj3 therap*) (Subsequent adj3 treatment*) (Subsequent adj3 medication*)	Carcinoma Neoplasm Sarcoma
MESH	"Survival Analysis"	Randomized Controlled Trials as Topic		Neoplasms

*Overview of the literature search. Each row for each approach (concept, synonyms, MESH term) was combined with the Boolean operator OR and each column was combined with the Boolean operator AND. Terms which were required to be in the title or abstract were suffixed with .tw., full phrases were specified with quotation marks, terms which could have multiple endings were suffixed with a * and terms which could have English or American spellings were noted with a ? so both were included. Any terms where phrases could be given in several ways were specified using the adj3 term which looks for the terms within 3 words of each other. MESH subjects were considered as focussed terms only.*
OS: Overall Survival; RCT: Randomised Controlled Trial.

Figure 2.1: Screening Process for abstracts



Flow diagram showing the process for abstracts to be included in the review.

2.4.2 Initial screening process

The results were screened based on their abstracts. The screening process was conducted by K-LR, reviewed by DAC and any uncertainty discussed. The screening process categorised the papers into: red (not relevant), amber (potentially relevant), and green (relevant) (Figure 2.1). Any paper which did not mention OS in the abstract was classed as red. In addition, if the abstract or full publication was not available it was also classed as red.

2.4.3 Reading process

All amber and green papers which remained were read in full by K-LR and categorised as: presenting clinical effectiveness results of a clinical trial, presenting a cost-effectiveness analysis, or a methodology or review paper discussing multiple treatment lines. Category-specific information was extracted from each publication (Table 2.2) and summarised using descriptive statistics. Note that for the review, crossover by design was included in the definition of a subsequent treatment line. Any uncertainty in the categorisation or interpretation of the information being extracted was discussed and agreed with DAC. The papers presenting cost-effectiveness were also read by DM given his experience as a Health Economist performing and reviewing cost-effective analysis and any disagreements discussed. The Kappa Coefficient of agreement was also calculated. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards were followed for reporting (Appendix Table B.1 and Appendix Table B.2).[6]

Table 2.2: Information Extracted from Each Article Type

Article Type	Information Extracted

Clinical Effectiveness Paper	<u>Overall Survival (OS)</u>
	• Endpoint level (Primary, Secondary, Exploratory, Unclear)
	• Definition
	• The analysis used for the interpretation of the primary endpoint
	• Whether a Kaplan-Meier Plot for OS was included
	  ○ Whether an at-risk table was included with the plot
	  ○ Whether the at-risk table included the number censored / number of events
	  ○ Whether confidence intervals were included on the plot
	• Whether a logrank test for OS was conducted
	  ○ Whether the test was adjusted or unadjusted for other factors
	• Whether a Cox proportional hazards model for OS was included
	  ○ Whether the results were adjusted or unadjusted for other factors
	<u>Subsequent Treatment[1]</u>
	• Whether the number or percent of trial participants receiving treatment in follow-up is included in the results or appendix
	• Whether a summary of the number of treatment lines and / or types of treatments trial participants received
	• Whether subsequent treatments are discussed in the discussion
	  ○ If yes – what are their thoughts?
	• Whether they did some form of analysis to account for subsequent treatments
	  ○ If yes – what analysis?
	[1]Note that subsequent treatments included cross-over by design

Cost-Effectiveness Paper	<u>Cost-Effectiveness</u>
	• What OS results fed into the Cost-Effectiveness analysis and how were they determined
	<u>Subsequent Treatment</u>
	• Whether subsequent treatment costs were incorporated into the Cost-Effectiveness analysis
	• Whether subsequent treatments are discussed in the discussion
	  ○ If yes – what are their thoughts?
	• Whether they did some form of analysis to account for subsequent treatments
	  ○ If yes – what analysis?

Methodology / Review Paper	<u>General Information</u> • Treatment Overview or Methodology Paper? <u>Subsequent Treatment</u> • Whether subsequent treatments are discussed.

Article Type	Information Extracted
	○ If yes – what are their thoughts?

2.5 Results

2.5.1 Overview and sample size

A total of 415 results were identified from the literature search. Following the screening process 281 (68%) were classed as red, 26 (6%) amber and 108 (26%) green (Figure 2.1). No changes were made to the screening process following DAC review. Therefore, a total of 134 papers were included in the review. Following the reading process these were identified to be 98 (73%) clinical effectiveness papers, 27 (20%) methodology or review papers and 9 (7%) cost-effectiveness papers.

2.5.2 Clinical effectiveness papers

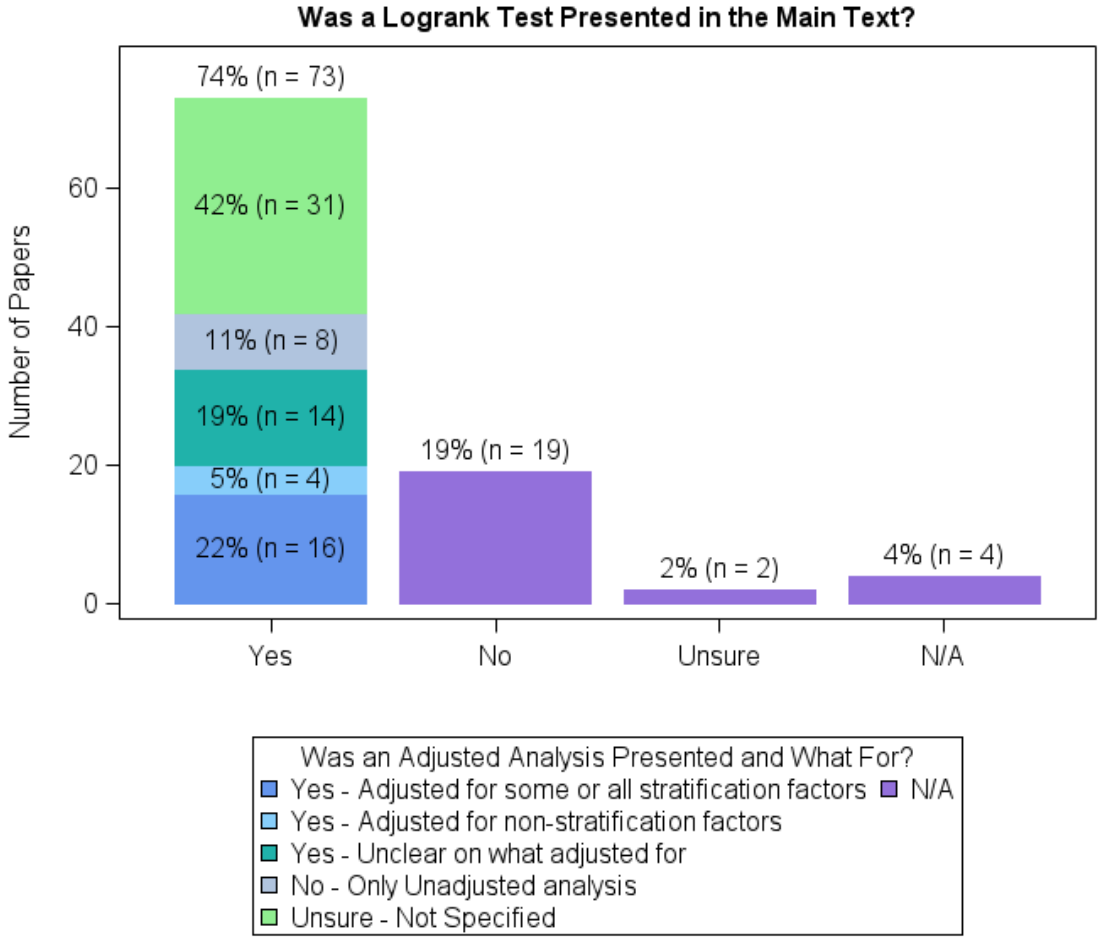
Overall Survival. The 98 categorised clinical effectiveness papers along with citations are listed in Appendix Table B.3. In total 45 (46%) of the clinical effectiveness papers reported OS to be the primary endpoint of the trial. Forty-eight papers (49%) reported OS as a secondary endpoint. Five papers (5%) stated that OS was an exploratory endpoint only.

OS was defined to be time from randomisation to death from any cause in 87 (89%) papers. Two papers (2%) defined the starting point to be diagnosis, three papers (3%) intervention initiation, one paper enrolment (1%) and one paper pre-inclusion (1%), but it was unclear what these meant in relation to the point of randomisation. Four papers (4%) did not explicitly define OS within the text, two of which were not presenting OS results.

Most papers (86/98, 88%) presented a Kaplan-Meier curve for OS in the main manuscript. Of these, the majority included the numbers at risk (79/86, 92%). However, only a small proportion included the number of events or number censored in their at-risk table (13/79, 16%). None of the papers provided confidence intervals on their Kaplan-Meier curve. Eight papers, with mature OS data, did not present a Kaplan-Meier curve for OS in their main manuscript. Four papers included them in the appendix [7-10] and four omitted a Kaplan-Meier curve entirely [11-14]. A similar proportion of papers presented a logrank test (73/98, 74%) and/or a Cox proportional hazards model (71/98, 72%). However, the level of adjustment (also referred to as stratification) of the tests varied across papers. Of the 73 papers which included a logrank test, 31 (42%) were not clear as to whether they reported adjusted or unadjusted results (Figure 2.2).

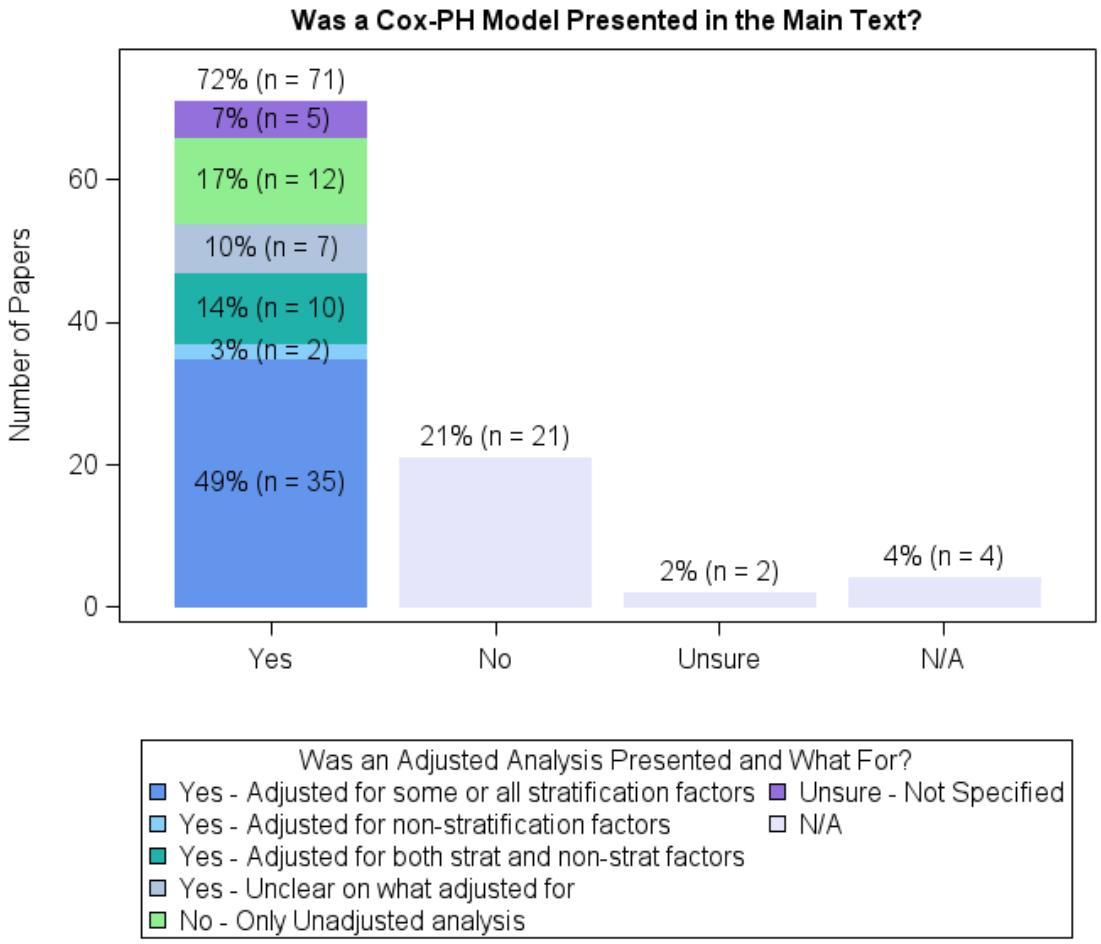
In contrast, around half of papers which presented results of the Cox model reported results which clearly had been adjusted for some or all of the trial's stratification factors (Figure 2.3, 35/71, 49%).

Figure 2.2: Was a Logrank Test Presented in the Main Text?



Stacked bar chart showing whether a logrank test was presented in the main text of the effectiveness papers (quantified by the percentages at the top of the bar) and whether it was adjusted or unadjusted for additional factors (quantified by the percentages within the bar).

Figure 2.3: Was a Cox-PH Model Presented in the Main Text?



Stacked bar chart showing whether a Cox-Proportional Hazards model was presented in the main text of the effectiveness papers (quantified by the percentages at the top of the bar) and whether it was adjusted or unadjusted for additional factors (quantified by the percentages within the bar).

A summary of the primary analysis methods used for OS are shown by endpoint level in Table 2.3, by year in Appendix Table B.4 and by cancer type in Appendix Table B.5. The most popular primary analysis method was the combination of a p-value from a logrank test and the hazard ratio from a Cox proportional hazards model. In total 36 of the 98 papers (37%) papers interpreted OS results using the p-value from a stratified or unstratified logrank test and the hazard ratio from an adjusted or unadjusted Cox model. Of the 98 papers, 15 (15%) used a hazard ratio and p-value from an adjusted Cox model to interpret OS. Of these 15, 12 (80%), were when OS was a primary or co-primary endpoint. When considered by year (Appendix Table B.4) of publication the proportion of papers reporting OS as a (co-)primary endpoint fluctuates over the years. Since 2018 OS was more often a secondary rather than (co-)primary endpoint (Appendix Table B.4). Considering cancer type (Appendix Table B.5) Upper gastrointestinal / Hepato-biliary cancers had the highest proportion of papers presenting OS as a primary endpoint.

Table 2.3: Summary of the Clinical Effectiveness papers by endpoint type

	Primary or Co-Primary (n=45)	Secondary (n=48)	Exploratory (n=5)	Total (n=98)
Was a Kaplan-Meier curve for OS presented in the main text?				
Yes	40 (88.9%)	41 (85.4%)	5 (100.0%)	86 (87.8%)
No	2 (4.4%)	6 (12.5%)	0 (0.0%)	8 (8.2%)
N/A[1]	3 (6.7%)	1 (2.1%)	0 (0.0%)	4 (4.1%)
Were the numbers at risk included in the Kaplan-Meier curve?				
Yes	36 (80.0%)	38 (79.2%)	5 (100.0%)	79 (80.6%)
No	4 (8.9%)	3 (6.3%)	0 (0.0%)	7 (7.1%)
N/A	5 (11.1%)	7 (14.6%)	0 (0.0%)	12 (12.2%)
Were the number of patients censored or number of events included in the at risk table?				
Yes	9 (20.0%)	3 (6.3%)	1 (20.0%)	13 (13.3%)
No	27 (60.0%)	35 (72.9%)	4 (80.0%)	66 (67.3%)
N/A	9 (20.0%)	10 (20.8%)	0 (0.0%)	19 (19.4%)
Were confidence intervals included in the Kaplan-Meier curve?				
No	40 (88.9%)	41 (85.4%)	5 (100.0%)	86 (87.8%)
N/A	5 (11.1%)	7 (14.6%)	0 (0.0%)	12 (12.2%)
Primary analysis				
Hazard Ratio and 95% CI from an unadjusted Cox model	1 (2.2%)	0 (0.0%)	0 (0.0%)	1 (1.0%)
Hazard ratio and 95% CI from an adjusted Cox model	1 (2.2%)	1 (2.1%)	0 (0.0%)	2 (2.0%)
Hazard ratio and P-value from a logrank test	8 (17.8%)	11 (22.9%)	0 (0.0%)	19 (19.4%)
Hazard ratio and P-value from a stratified logrank test.	3 (6.7%)	3 (6.3%)	0 (0.0%)	6 (6.1%)
Hazard ratio and p-value from an adjusted Cox model	12 (26.7%)	2 (4.2%)	1 (20.0%)	15 (15.3%)
Hazard ratio and p-value from an unadjusted Cox model	2 (4.4%)	3 (6.3%)	0 (0.0%)	5 (5.1%)
Hazard ratio from a Cox model and p-value from a logrank test	3 (6.7%)	7 (14.6%)	0 (0.0%)	10 (10.2%)
Hazard ratio from an adjusted Cox model and p-value from a logrank test.	2 (4.4%)	4 (8.3%)	0 (0.0%)	6 (6.1%)
Hazard ratio from an adjusted Cox model and p-value from a stratified logrank test.	5 (11.1%)	5 (10.4%)	3 (60.0%)	13 (13.3%)
Hazard ratio from an unadjusted Cox model and p-value from a logrank test.	1 (2.2%)	4 (8.3%)	0 (0.0%)	5 (5.1%)
Hazard ratio from an unadjusted Cox model and p-value from a stratified logrank test.	0 (0.0%)	2 (4.2%)	0 (0.0%)	2 (2.0%)
N/A[1]	3 (6.7%)	1 (2.1%)	0 (0.0%)	4 (4.1%)
Other	4 (8.9%)	5 (10.4%)	1 (20.0%)	10 (10.2%)
Were subsequent treatments mentioned in the paper?				
Yes	33 (73.3%)	23 (47.9%)	3 (60.0%)	59 (60.2%)
No	12 (26.7%)	25 (52.1%)	2 (40.0%)	39 (39.8%)
Was the percentage or number of participants who received later lines included in the main text or appendix?				
Yes	31 (68.9%)	18 (37.5%)	1 (20.0%)	50 (51.0%)
No	1 (2.2%)	5 (10.4%)	2 (40.0%)	8 (8.2%)
N/A	13 (28.9%)	25 (52.1%)	2 (40.0%)	40 (40.8%)
Was a breakdown or summary of the number of subsequent treatments (Two, Three, Four) included in the main text or appendix?				

	Primary or Co-Primary (n=45)	Secondary (n=48)	Exploratory (n=5)	Total (n=98)
Yes	3 (6.7%)	1 (2.1%)	0 (0.0%)	4 (4.1%)
No	28 (62.2%)	22 (45.8%)	3 (60.0%)	53 (54.1%)
N/A[2]	14 (31.1%)	25 (52.1%)	2 (40.0%)	41 (41.8%)
Was a breakdown or summary of the type of subsequent treatments (Treatment 1, Treatment 2, Treatment 3) included in the main text or appendix?				
Yes	26 (57.8%)	18 (37.5%)	1 (20.0%)	45 (45.9%)
No	5 (11.1%)	5 (10.4%)	2 (40.0%)	12 (12.2%)
N/A[2]	14 (31.1%)	25 (52.1%)	2 (40.0%)	41 (41.8%)
Was additional analysis conducted to account for subsequent treatment lines?				
Yes	2 (4.4%)	3 (6.3%)	2 (40.0%)	7 (7.1%)
No	30 (66.7%)	20 (41.7%)	1 (20.0%)	51 (52.0%)
N/A[2]	14 (31.1%)	25 (52.1%)	2 (40.0%)	41 (41.8%)
Were subsequent treatment lines mentioned in the discussion?				
Yes	22 (48.9%)	15 (31.3%)	2 (40.0%)	39 (39.8%)
No	11 (24.4%)	8 (16.7%)	1 (20.0%)	20 (20.4%)
N/A	12 (26.7%)	25 (52.1%)	2 (40.0%)	39 (39.8%)
How were subsequent treatments mentioned in the discussion?				
The uptake of subsequent treatments is given as a reason for the discrepancy between OS and surrogate endpoints.	0 (0.0%)	2 (4.2%)	0 (0.0%)	2 (2.0%)
The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	4 (8.9%)	1 (2.1%)	1 (20.0%)	6 (6.1%)
The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	6 (13.3%)	7 (14.6%)	1 (20.0%)	14 (14.3%)
The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	4 (8.9%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	2 (4.4%)	1 (2.1%)	0 (0.0%)	3 (3.1%)
Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	2 (4.4%)	2 (4.2%)	0 (0.0%)	4 (4.1%)
None or limited options of subsequent treatment lines for patients are given as a reason as to why subsequent treatment lines will not have affected the OS results.	0 (0.0%)	2 (4.2%)	0 (0.0%)	2 (2.0%)
Other	4 (8.9%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
N/A	23 (51.1%)	33 (68.8%)	3 (60.0%)	59 (60.2%)

[1] Not-Applicable here refers to papers where OS is a pre-specified endpoint, but the data is not mature at the time of the publication. [15-18]
[2] Note one paper stated no patients received subsequent treatments so is classed as N/A for these summaries.
N/A: Not Applicable; OS: Overall Survival.

Subsequent Treatment Lines. In total 59 of the 98 (60%) papers mentioned subsequent treatment lines in the main manuscript or supplementary material. Of those that mentioned subsequent treatment lines, the majority (50/59, 85%) provided the number or percentage of participants who received a later treatment line during trial follow-up. Four of the 59 (7%) papers provided a breakdown of the number of later lines a participant received. However, 45 of the 59 (76%) provided a summary of the types of treatments participants received during follow-up. Finally, seven of the 59 (12%) papers performed additional analysis addressing subsequent treatments. The most common reason for the additional analysis was due to participants switching to receive the experimental intervention.[10, 19-21]. Miles et al[19] included time-dependent covariates in the Cox proportional hazards model to adjust for trial participants crossing over from placebo to active treatment. Dimopoulos et al[22] conducted a post-hoc landmark analysis which measured OS from time of progression. Hodi et al[23] included a sensitivity analysis for OS which censored participants at the point of subsequent treatment. Middleton et al[20] conducted a post-hoc landmark analysis which compared survival of patients on the active arm who received standard chemotherapy after progression with those who were still alive on the standard chemotherapy arm at this timepoint. Goldhirsch et al[10] conducted analysis which cut-off follow-up at specific timepoints to report how the effect changed over time after patients switched from observation to active in follow-up. Penichoux et al[12] included time dependent covariates in various Cox proportional hazards models to account for progression, toxicity and introduction of third line treatment, they also modelled progression and toxicity events as repeating events using shared gamma frailty models where death was classed as the definitive event. Finally, Gianni et al[21] conducted an analysis which censored patients at the point of cross-over from observation to active treatment. In the discussion section, 39 of the 59 (66%) papers mentioned subsequent treatment lines (Table 2.3). The most common discussion point for 14 of the 39 (36%) papers was "*The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.*" However, four papers also stated that "*The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.*"

2.5.3 Cost-effectiveness papers

In total nine papers whose primary purpose was to report cost-effectiveness analysis were identified during the literature review (Appendix Table B.6). There were no disagreements between K-LR and DM following review. Therefore, the Kappa coefficient was one for each summary.

The majority of the papers (6/9, 67%) used a partitioned survival analysis model where the OS estimate was extrapolated from a Kaplan-Meier plot (5/9, 56%). Five of the nine (56%) papers included subsequent costs and four of the nine (44%) papers performed additional analysis to account for subsequent lines of therapy. One paper used the Rank-Preserving Structural Failure Time Model (RPSFTM)[24], another used the two-stage method[25] and the other two adjusted for the cross-over in the trial, but the method used was not specified.[26, 27]

2.5.4 Methodology papers

In total 27 methodology / review papers were identified (Appendix Table B.7). Of the papers which mentioned subsequent treatments (14/27, 52%), the majority (9/14, 64%) classed them as a limitation for the interpretation of OS in clinical trials.[28-36] One paper (1/14, 7%) stated that they were not a limitation for interpretation if they reflected the care patients would receive after the trial[37] and four papers (4/14, 29%) did not mention subsequent treatments in this context.[38-41]

2.6 Discussion

This review identified 134 papers which had reported or discussed OS in the last decade with the aim to determine how it was currently analysed and if subsequent treatment lines were being accounted for and/or discussed.

The main finding in terms of the analysis of OS is that whilst it is generally defined consistently as an endpoint, and is commonly visualised with Kaplan-Meier curve, the primary analysis is not as consistent as might have been expected. The authors expected that the majority of papers would report the results of a Cox proportional hazards model adjusted for at least the trial's stratification factors, where both the hazard ratio and the p-value from the adjusted model would be used to interpret OS. This would be in line with guidance in ICH E9 which states, "*If one or more factors are used to stratify the design, it is appropriate to account for those factors in the analysis.*"[2] However, this was not the case. The majority of authors used a combination of a Cox model and a logrank test to interpret their primary analysis and make their

conclusions. Whilst this may be statistically sound, it could be considered inconsistent as the confidence interval around the hazard ratio for the Cox model does not directly relate to the p-value from the logrank test as it would under the laws of duality if the p-value from the Cox model was used. It was reassuring to see that where a Cox model was included in the text, as a secondary analysis for OS, the majority of papers did report that it was adjusted for some or all of the stratification factors. This was less clear when a logrank test was presented. This is unexpected given that ICH E9 was first published in 1998, before the exclusion publication date for this literature review. In order to encourage consistency and transparency in reporting researchers in the future should consider planning and describing their analysis using the estimand framework proposed within ICHE9(R1).[42] These results are consistent with the overview of reporting survival analysis endpoints by Batson et al[43] which was identified as a methodology / review paper in the search (ORN 209). In the paper, Batson and colleagues concluded that the majority of papers identified in their review reported the results of a Cox model and a logrank test with varying levels of adjustment. However, it is important to note that they were focusing on time to event endpoints broadly rather than just OS. It is also interesting that whilst a recent survey found that respondents consider adding a confidence interval to a Kaplan-Meier plot useful[3] no papers included in this review included confidence intervals.

The findings around subsequent treatment lines were less surprising. The majority of papers, which mentioned subsequent treatment lines, summarised the types of treatments being received, this complements a recent review by Oliver et al[44]. However, only seven papers, five of which were primary or interim results papers[10, 20-23], implemented some form of analysis to account for anti-cancer treatment received during trial follow-up. The majority of these only considered when participants switched to the alternative trial intervention and only two used methods akin to that described by Latimer et al[5] (censoring at point of switch[21], sequencing models[12]). However, both were published prior to these recommendations. The work conducted by Penichoux et al[12] using a shared frailty model to model progression and toxicity events as repeating events seems novel and does not seem to have received a lot of attention with only four papers citing this work. Two of the citations are not relevant to this review. One is reporting the results of a single-arm study[45] and one a conference which is not available in the English language[46].One of the remaining citations by Malka et al[47] is an editorial piece which references Penichoux's work as an example of how to use random-effects to address patients receiving the same intervention at different treatment lines. The final paper to cite Penichoux's work is relevant to this research. Petracci et al[48] used recurrent event analysis to assess the overall effectiveness of an

intervention, in terms of progression-free survival, when it could have been used in first or second line by trial design. In the paper, the authors propose OS as an area for future work. There is a question as to whether this is restricted to the context of trials comparing specific sequences of treatment or whether it could be applied to account for subsequent lines of therapy that are not defined by trial design. Latimer et al[5] suggest that sequencing models are a naïve approach given the need for data on each treatment line is hard to come by. However, there is a question as to whether this is still the case given the utilisation of routine data and patients awareness of how their data can be used through strategies such as DATA-CAN. In terms of the papers which discussed subsequent treatment lines, the most common discussion point was that uptake of subsequent treatments may have reduced the OS effect, affected the results or used to caveat the results. Interestingly the lack of subsequent treatment lines was also given as a reason for no OS effect. Whilst we agree with Oliver et al that each paper reporting OS should summarise subsequent anti-cancer treatment[44], we believe that descriptive statistics alone are not sufficient to support the interpretation of OS when patients receive anti-cancer therapy during trial follow-up. Instead, methodological work is required to develop estimation techniques to disentangle, if present, the effects of the trial experimental treatment and subsequent treatment lines.

Conversely, whilst only nine cost-effectiveness papers were identified they did seem to have a consistent approach to their primary analysis. In addition, where it was clear, the majority of papers did include subsequent treatment costs with a good proportion going further and including additional analysis which adjusted the OS estimate for subsequent treatment lines. Given the contrast to the clinical effectiveness papers, there is a question as to whether adjusting for subsequent lines is being left to be considered by those that evaluate the intervention for HTA review, which often requires cost-effectiveness to be fully analysed, rather than at the publication of the primary clinical effectiveness results.

The majority of the identified methodology/review papers were discussing treatment as a whole and did not consider subsequent lines in the context being considered here. As expected, a number of the methodology papers mentioned subsequent lines as a reason for assessing surrogate endpoints in their field[28, 30, 35] as this is in line with existing guidance.[4] This may also explain the shift in OS from a primary to secondary endpoint as seen in the analysis of the clinical effectiveness papers. In addition, one of the papers published in 2011 stated that subsequent lines are only an issue for the interpretation of clinical trial results if they fall outside of standard of care (SoC).[37] Whilst this may be true, cancer research has accelerated in the last decade with more varied treatment options being available to patients. For example, in England, there are

calls for guidance to be published quickly, and appraisals to be conducted efficiently, so that patients can get access to new treatments faster as part of the NHS, where, once approved, treatments should be available to patients within three months.[49] Currently there are 377 trials in cancer listed as “ongoing” on the ISCRTN database.[50] In addition, between 1st Jan 2021 and 1st Jan 2022 there were 35 technology appraisal guidelines published by NICE to advise on the treatment of cancer.[51] This is more than four times the eight appraisals which were published in between 1st Jan 2010 and 1st Jan 2011.[52] This rapidly evolving clinical landscape means that the SoC is changing during the trial timeline and will continue to change after the trial closes and the treatment is incorporated into standard practice. Therefore, if OS is analysed without consideration of the subsequent lines, it is possible that the results will soon be invalid after publication. Methodological research is needed now to better account for subsequent lines so that the true effect of the new experimental intervention in trials is understood and considered when interventions are reviewed for use in SoC. This would work towards ensuring that interventions are properly evaluated so that beneficial interventions are not rejected due to subsequent lines diminishing any effect, and the time and resources it takes for a phase III confirmatory trial to report OS results are not wasted.

This review provides evidence that whilst OS is less frequently the primary endpoint, it is still of interest, and even when treatment switching is not the case, subsequent treatments are being considered as a limitation. This is the same across the different perspectives from the clinical effectiveness papers to the methodological and cost-effectiveness papers. At present, the approach to the assessment of both OS and subsequent treatment lines is not consistent. Whilst there will never be a one size fits all to the analysis of clinical trials, there is a need for an approach to subsequent treatment lines to aid interpretation of clinical trial results as more interventions are developed and recommended for use in standard or care.

There are some limitations to this systematic review. As there was only one researcher conducting the review, the more general search to gauge practice was restricted to only include top United Kingdom and international journals. It may be that the exploratory analysis which we are interested in here is reported in other, less-prominent journals. However, this was accounted for in the two approaches where the first search was not restricted by journal. In addition, articles prior to 2010 were excluded from the search. However, this is considered to be less of an issue as the technology support document advocating accounting for subsequent lines in the form of treatment switching was published in 2014.[5] Furthermore, whilst all uncertainties were discussed with DAC, having an additional researcher conduct and read all of the

results may have also helped to overcome some of the inconsistencies in interpretation of the terminology. For example, in some cases it was unclear whether stratification of a Cox model referred to applying different baseline hazards for different strata, or including the variable in the model as an independent variable (the latter more commonly referred to as adjustment). However, logrank tests were also reported to be adjusted when they can only be stratified. Finally, only papers which were written in English and available through free-text or institutional access were included. Papers written in other languages or behind a paywall may also have considered subsequent treatment lines and how they impact on the analysis of OS.

2.7 Conclusion

Over the last decade OS has changed from being the primary endpoint of cancer clinical trials to more commonly a secondary endpoint. Despite this, subsequent treatment lines in relation to OS are being consistently mentioned in articles, confirming that whilst OS may not be the primary endpoint, the analysis and results are still of interest. The majority of articles concluded that subsequent lines were a limitation of the interpretation of their OS analysis. However, analysis accounting for subsequent lines was generally only conducted in the event of treatment switching. The implemented methods were often naïve, and researchers did not utilise the more sophisticated and less well-known methods of IPCW or the RPSFTM. In fact, RPSFTM was only observed in one paper aiming to determine cost-effectiveness rather than clinical effectiveness. There is a question of whether researchers are aware of these more advanced methods (a possible educational gap) or whether they are just not being published (a reporting gap). This will be explored through the next stage of this research, in which stakeholders (patients and the public, statisticians and other data analysts, healthcare professionals, payers and industry partners) will be invited to complete a questionnaire on their views of OS and its analysis. Within the questionnaire, statisticians and other data analysts (e.g., health economists) will be asked to answer whether they have heard of and whether they have applied the various methods to analyse OS in cancer clinical trials. The overarching aim of the questionnaire will be to agree a consensus across the stakeholder groups in terms of the direction of this research and how statistical methodology should be extended or developed to account for subsequent treatment lines in the analysis of cancer clinical trials.

2.8 Declarations

2.8.1 Ethics approval and consent to participate

Not applicable

2.8.2 Consent for publication

Not applicable

2.8.3 Availability of data and material

The material from this literature review is available upon request.

2.8.4 Competing Interests

The authors declare that they have no competing interests.

2.8.5 Funding

This project is funded by the NIHR Doctoral Fellowship Programme [NIHR301510]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also supported by Core Clinical Trials Unit Infrastructure from Cancer Research UK [C7852/A25447]. IRW was supported by the Medical Research Council [Programme MC_UU_00004/07].

2.8.6 Author contributions

Under supervision from DM, JKV-R, IRW and DAC, K-LR conceived the idea for the review, planned the project, performed the screening, extracted the data, and drafted the manuscript. All authors read and approved the final manuscript.

2.8.7 Acknowledgements

Not applicable

2.9 Chapter 2 References

1. Fontrier AM, Visintin E and Kanavos P. Similarities and Differences in Health Technology Assessment Systems and Implications for Coverage Decisions: Evidence from 32 Countries. *Pharmacoecon Open* 2022; 6: 315-328. 20211129. DOI: 10.1007/s41669-021-00311-5.
2. EMA. ICH Topic E9 Statistical Principles for Clinical Trials, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf (1998, accessed 6th September 2022).
3. Morris TP, Jarvis CI, Cragg W, et al. Proposals on Kaplan–Meier plots in medical research and a survey of stakeholder views: KMunicate. *BMJ Open* 2019; 9: e030215. DOI: 10.1136/bmjopen-2019-030215.

4. FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry, https://www.fda.gov/media/71195/download (2018, accessed 6th September 2022).
5. Latimer NR and Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
6. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews* 2021; 10: 1-11.
7. Chopra S, Gupta S, Kannan S, et al. Late Toxicity After Adjuvant Conventional Radiation Versus Image-Guided Intensity-Modulated Radiotherapy for Cervical Cancer (PARCER): A Randomized Controlled Trial. *J Clin Oncol* 2021; 39: 3682-3692. DOI: https://dx.doi.org/10.1200/JCO.20.02530.
8. Carrie C, Hasbini A, de Laroche G, et al. Salvage radiotherapy with or without short-term hormone therapy for rising prostate-specific antigen concentration after radical prostatectomy (GETUG-AFU 16): a randomised, multicentre, open-label phase 3 trial. *Lancet Oncol* 2016; 17: 747-756. Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S1470-2045(16)00111-X.
9. Kunkler IH, Williams LJ, Jack WJ, et al. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. *Lancet Oncol* 2015; 16: 266-273. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)71221-5.
10. Goldhirsch A, Gelber RD, Piccart-Gebhart MJ, et al. 2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. *Lancet* 2013; 382: 1021-1028. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(13)61094-6.
11. Vora A, Goulden N, Mitchell C, et al. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. *Lancet Oncol* 2014; 15: 809-818. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)70243-8.
12. Penichoux J, Michiels S, Bouche O, et al. Taking into account successive treatment lines in the analysis of a colorectal cancer randomised trial. *Eur J Cancer* 2013; 49: 1882-1888. Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.ejca.2013.02.006.
13. Cuzick J, Sestak I, Baum M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol* 2010; 11: 1135-1141. Clinical Trial, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(10)70257-6.
14. Zietman AL, Bae K, Slater JD, et al. Randomized trial comparing conventional-dose with high-dose conformal radiation therapy in early-stage adenocarcinoma of the prostate: long-term results from proton radiation oncology group/american college of radiology 95-09. *J Clin Oncol* 2010; 28: 1106-1111. Clinical Trial, Phase I, Clinical Trial, Phase II, Comparative Study, Randomized Controlled Trial, Research Support, N.I.H., Extramural. DOI: https://dx.doi.org/10.1200/JCO.2009.25.8475.
15. Clamp AR, James EC, McNeish IA, et al. Weekly dose-dense chemotherapy in first-line epithelial ovarian, fallopian tube, or primary peritoneal carcinoma treatment (ICON8): primary progression free survival analysis results from a GCIG phase 3 randomised controlled trial. *Lancet* 2019; 394: 2084-2095. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(19)32259-7.
16. Minnaar CA, Kotzen JA, Ayeni OA, et al. The effect of modulated electro-hyperthermia on local disease control in HIV-positive and -negative cervical cancer

women in South Africa: Early results from a phase III randomised controlled trial. *PLoS ONE* 2019; 14: e0217894. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0217894.
17. Park K, Tan EH, O'Byrne K, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. *Lancet Oncol* 2016; 17: 577-589. Clinical Trial, Phase II, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S1470-2045(16)30033-X.
18. James ND, Sydes MR, Mason MD, et al. Celecoxib plus hormone therapy versus hormone therapy alone for hormone-sensitive prostate cancer: first results from the STAMPEDE multiarm, multistage, randomised controlled trial. *Lancet Oncol* 2012; 13: 549-558. Clinical Trial, Phase II, Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(12)70088-8.
19. Miles D, Im YH, Fung A, et al. Effect of docetaxel duration on clinical outcomes: exploratory analysis of CLEOPATRA, a phase III randomized controlled trial. *Ann Oncol* 2017; 28: 2761-2767. Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1093/annonc/mdx406.
20. Middleton G, Silcocks P, Cox T, et al. Gemcitabine and capecitabine with or without telomerase peptide vaccine GV1001 in patients with locally advanced or metastatic pancreatic cancer (TeloVac): an open-label, randomised, phase 3 trial. *Lancet Oncol* 2014; 15: 829-840. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)70236-0.
21. Gianni L, Dafni U, Gelber RD, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol* 2011; 12: 236-244. Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(11)70033-X.
22. Dimopoulos MA, Goldschmidt H, Niesvizky R, et al. Carfilzomib or bortezomib in relapsed or refractory multiple myeloma (ENDEAVOR): an interim overall survival analysis of an open-label, randomised, phase 3 trial. *Lancet Oncol* 2017; 18: 1327-1337. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(17)30578-8.
23. Hodi FS, Chesney J, Pavlick AC, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol* 2016; 17: 1558-1568. Clinical Trial, Phase II Multicenter Study Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S1470-2045(16)30366-7.
24. Srivastava T, Prabhu VS, Li H, et al. Cost-effectiveness of Pembrolizumab as Second-line Therapy for the Treatment of Locally Advanced or Metastatic Urothelial Carcinoma in Sweden. *Eur Urol Oncol* 2020; 3: 663-670. Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.euo.2018.09.012.
25. Bhadhuri A, Insinga R, Guggisberg P, et al. Cost effectiveness of pembrolizumab vs chemotherapy as first-line treatment for metastatic NSCLC that expresses high levels of PD-L1 in Switzerland. *Swiss Med Wkly* 2019; 149: w20170. Comparative Study. DOI: https://dx.doi.org/10.4414/smw.2019.20170.
26. Miquel-Cases A, Retel VP, Lederer B, et al. Exploratory Cost-Effectiveness Analysis of Response-Guided Neoadjuvant Chemotherapy for Hormone Positive Breast Cancer Patients. *PLoS ONE* 2016; 11: e0154386. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0154386.
27. Hoyle M, Peters J, Crathorne L, et al. Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for

KRAS wild-type patients with metastatic colorectal cancer. *Value Health* 2013; 16: 288-296. Comparative Study Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.jval.2012.11.001.
28. Nomura S, Hirakawa A and Hamada C. Sample size determination for the current strategy in oncology Phase 3 trials that tests progression-free survival and overall survival in a two-stage design framework. *J Biopharm Stat* 2018; 28: 589-611. Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1080/10543406.2017.1372775.
29. Osanto S and van der Hulle T. Cabozantinib in the treatment of advanced renal cell carcinoma in adults following prior vascular endothelial growth factor targeted therapy: clinical trial evidence and experience. *Ther Adv Urol* 2018; 10: 109-123. Review. DOI: https://dx.doi.org/10.1177/1756287217748867.
30. Zer A, Prince RM, Amir E and Abdul Razak A. Evolution of Randomized Trials in Advanced/Metastatic Soft Tissue Sarcoma: End Point Selection, Surrogacy, and Quality of Reporting. *J Clin Oncol* 2016; 34: 1469-1475. Review Systematic Review. DOI: https://dx.doi.org/10.1200/JCO.2015.64.3437.
31. Hussain M, Tangen C, Higano C, et al. Evaluating Intermittent Androgen-Deprivation Therapy Phase III Clinical Trials: The Devil Is in the Details. *J Clin Oncol* 2016; 34: 280-285. Clinical Trial, Phase III Randomized Controlled Trial. DOI: https://dx.doi.org/10.1200/JCO.2015.62.8065.
32. Garcia-Albeniz X, Maurel J and Hernan MA. Why post-progression survival and post-relapse survival are not appropriate measures of efficacy in cancer randomized clinical trials. *Int J Cancer* 2015; 136: 2444-2447. Research Support, N.I.H., Extramural. DOI: https://dx.doi.org/10.1002/ijc.29278.
33. Bai X, Liu J, Li L and Faries D. Adaptive truncated weighting for improving marginal structural model estimation of treatment effects informally censored by subsequent therapy. *Pharm Stat* 2015; 14: 448-454. DOI: https://dx.doi.org/10.1002/pst.1719.
34. Matulonis UA, Oza AM, Ho TW and Ledermann JA. Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. *Cancer* 2015; 121: 1737-1746. Research Support, Non-U.S. Gov't Review. DOI: https://dx.doi.org/10.1002/cncr.29082.
35. Bria E, Massari F, Maines F, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma. Correlation with overall survival, benchmarking and power analysis. *Crit Rev Oncol Hematol* 2015; 93: 50-59. Research Support, Non-U.S. Gov't Review. DOI: https://dx.doi.org/10.1016/j.critrevonc.2014.08.001.
36. Fojo T, Amiri-Kordestani L and Bates SE. Potential pitfalls of crossover and thoughts on iniparib in triple-negative breast cancer. *J Natl Cancer Inst* 2011; 103: 1738-1740. Research Support, N.I.H., Intramural Review. DOI: https://dx.doi.org/10.1093/jnci/djr386.
37. Korn EL, Freidlin B and Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol* 2011; 29: 2439-2442. Review. DOI: https://dx.doi.org/10.1200/JCO.2011.34.6056.
38. Yeh C and Bates SE. Two decades of research toward the treatment of locally advanced and metastatic pancreatic cancer: Remarkable effort and limited gain. *Semin Oncol* 2021; 48: 34-46. DOI: https://dx.doi.org/10.1053/j.seminoncol.2021.01.001.
39. Cranmer H, Shields GE and 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: 1176-1185. DOI: https://dx.doi.org/10.1080/13696998.2020.1796360.
40. Chen H, Horita N, Ito K, et al. Hazard ratio of progression-free survival is an excellent predictor of overall survival in phase III randomized controlled trials evaluating the first-line chemotherapy for extensive-disease small-cell lung cancer. *Transl* 2020; 9: 1333-1342. DOI: https://dx.doi.org/10.21037/tlcr-20-377.

41. De Maeseneer DJ, Delafontaine B and Rottey S. Checkpoint inhibition: new treatment options in urologic cancer. *Acta Clin Belg* 2017; 72: 24-28. Review. DOI: https://dx.doi.org/10.1080/17843286.2016.1260890.
42. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (2017, accessed 1st March 2023).
43. Batson S, Greenall G and Hudson P. Review of the Reporting of Survival Analyses within Randomised Controlled Trials and the Implications for Meta-Analysis. *PLoS ONE* 2016; 11: e0154870. Review Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0154870.
44. Olivier T, Haslam A and Prasad V. Post-progression treatment in cancer randomized trials: a cross-sectional study of trials leading to FDA approval and published trials between 2018 and 2020. *BMC Cancer* 2023; 23: 448. DOI: 10.1186/s12885-023-10917-z.
45. Texier M, Rotolo F, Ducreux M, et al. Evaluation of Treatment Effect with Paired Failure Times in a Single-Arm Phase II Trial in Oncology. *Computational and Mathematical Methods in Medicine* 2018; 2018: 1672176. DOI: 10.1155/2018/1672176.
46. Sureda BM, Izquierdo ALY and Lorenzo JP. INTEGRACIÓN DE LA INNOVACIÓN TERAPÉUTICA PARA AUMENTAR LA SUPERVIVENCIA EN EL CÁNCER COLORRECTAL METÁSTASICO.
47. Malka D, Rotolo F and Boige V. First-line treatment in metastatic colorectal cancer: Important or crucial? *Eur J Cancer* 2017; 84: 363-366.
48. Petracci E, Scarpi E, Passardi A, et al. Effectiveness of bevacizumab in first- and second-line treatment for metastatic colorectal cancer: ITACa randomized trial. *Ther Adv Med Oncol* 2020; 12: 1758835920937427. 20200723. DOI: 10.1177/1758835920937427.
49. NICE. Cost comparison. Addendum to the Guide to the methods of technology appraisal https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/methods-guide-addendum-cost-comparison.pdf (accessed 29th July 2020).
50. ISCRTN. ISCRTN Registry, https://www.isrctn.com/search?q=&filters=trialStatus%3Aongoing%2CconditionCategory%3ACancer (accessed 6th September 2022).
51. NICE. Guidance, NICE advice and quality standards: 2021 - 2022, https://www.nice.org.uk/guidance/published?from=2021-01-01&to=2022-01-01&ndt=Guidance&ngt=Technology%20appraisal%20guidance&ps=9999 (accessed 6th September 2022).
52. NICE. Guidance, NICE advice and quality standards: 2010 - 2011, https://www.nice.org.uk/guidance/published?ps=2500&from=2010-01-01&to=2011-01-01&ndt=Guidance&ngt=Technology+appraisal+guidance (accessed 6th September 2022).

Chapter 3 How should overall survival be analysed in randomised clinical trials in cancer if participants receive subsequent treatment lines? A stakeholder consultation

Kara-Louise Royle[1], Peter Wheatstone[2], David Meads[3], Jennifer K. Visser-Rogers[4], Ian R White[5], David A Cairns[1]

1. Leeds Cancer Research UK Clinical Trials Unit, Leeds Institute of Clinical Trials Research, University of Leeds, UK
2. Patient Partner
3. Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
4. Coronado Research, Kent, England
5. MRC Clinical Trials Unit at UCL, London, UK

Study 2	
Journal	Trials
Submission Status	Published in October 2025
Reference	Royle, KL., Wheatstone, P., Meads, D. *et al.* How should overall survival be analysed in randomised clinical trials in cancer if participants receive subsequent treatment lines? A stakeholder consultation. *Trials* **26**, 434 (2025). https://doi.org/10.1186/s13063-025-09148-3

3.1 Abstract

3.1.1 Background

Overall survival is used to assess clinical effectiveness in cancer clinical trials. However, in practice it may be influenced by intercurrent events post-randomisation. The decisions made on how to address these intercurrent events change the interpretation of the results.

An example is when participants stop their trial intervention and start subsequent anti-cancer interventions (treatment lines) during trial follow-up. At present there is no evidence on the views of all stakeholders about this intercurrent event or consensus on how it should be addressed. The aim of this work was to understand the perspectives of all stakeholders and to obtain consensus through a qualitative study to guide future methodological work.

3.1.2 Methods

A modified Rand/UCLA appropriateness method was implemented. Stakeholder views were collected using an online questionnaire and discussed at a focus group. The

questionnaire included items on, the different methods for addressing an intercurrent event, data collection following an intercurrent event, statistical assumptions, and data presentation. Analysis was descriptive incorporating a conventional content approach. Consensus was defined a priori.

3.1.3 Results

103 stakeholders (30 Statisticians or Other Data Analysts, 6 Payers or Industry Partners, 22 Healthcare Professionals and 45 Patient, Carer or Members of the Public) completed the questionnaire between 3/8/2022 and 30/9/2022. 79% of respondents thought it important to consider the potential effect of subsequent treatment lines.

Consensus was reached on most questionnaire items. Stakeholders agreed that statistical assumptions were applicable only in "*Some Scenarios*" and that results should be presented using both a visual and summary measure. The focus group discussed different methods for addressing an intercurrent event and items around data collection where consensus was unclear. Seven participants attended (two patients/carers, one healthcare professional, three statisticians and one payer) with K-LR and PW. Attendees agreed that the treatment policy approach should be considered in future work as it was the most realistic, and that data collection was acceptable with informed consent.

3.1.4 Discussion

This work demonstrates that all stakeholder groups are interested in how subsequent treatment lines may impact overall survival and provides evidence on what future methodological work in the area should consider. The next step of this work will investigate whether it is possible to estimate the overall survival treatment effect in a hypothetical scenario where participants who received second-line therapy all received the same second-line therapy. This will aim to complement the existing treatment policy approach and quantify the impact of subsequent treatments.

3.1.5 Funding

NIHR301510

3.2 Key Words

Qualitative analysis, online questionnaire, focus group, overall survival, randomised controlled trial, cancer, subsequent treatment lines.

3.3 Introduction

In a confirmatory randomised controlled trial, the endpoint overall survival (OS) is defined as the time from randomisation to death from any cause.[1] It is a definitive endpoint used to assess the clinical effectiveness of experimental interventions for the treatment of cancer.

On analysing OS, like any endpoint, statisticians have to make decisions on how to address post-randomisation events which could affect the interpretation or occurrence of the outcome. These are termed intercurrent events and are defined as part of the estimand framework.[2] The estimand framework is a structured way to pre-specify how a clinical trial will answer a research question. It enables researchers to explicitly document the population of interest, the interventions being considered in the trial, the outcome of interest to answer the research question, how the outcome will be summarised at the population level, and consideration of any intercurrent events which may affect the interpretation of the outcome. There are five different approaches for addressing intercurrent events. The decision to use any one of these approaches affects the question being asked and therefore, how the results should be interpreted.

The results from the analysis of OS are used, alongside other evidence, in Health Technology Assessments (HTA) to determine whether interventions should be recommended for use in standard of care.[3] If approved, these results, alongside other factors, are then used by patients, carers and clinicians to decide on the best treatment option. Hence the decisions made by the statistician during analysis can impact the choices made by each stakeholder group. It follows that stakeholders should be consulted when statisticians are considering how to address intercurrent events.

An example of an intercurrent event, is when trial participants discontinue their trial intervention prior to death and receive subsequent anti-cancer treatment during trial follow-up. This intercurrent event is well known and is acknowledged as a disadvantage of OS in clinical trial recommendations.[1] At present methodological solutions only exist in the very specific and limited scenario of treatment switching, where patients in the control arm receive the experimental intervention during a later line of treatment.[4] However, a recent literature review showed that the intercurrent event, as a whole, is of interest. The review included 98 clinical effectiveness papers and 59 (60%) of them mentioned subsequent treatments. Of these, seven performed additional analysis for subsequent treatments. The review concluded that the majority of researchers are applying the treatment policy approach to the intercurrent event i.e. it is considered part of the treatment strategy and, to the author’s knowledge, there is

currently no methodological solution to determine how much of an effect subsequent treatment has on the interpretation of OS.[5]

The objectives of this work were to understand to different stakeholder perspectives and experience on the topic and seek consensus through a qualitative study, incorporating an online questionnaire and focus group. Consensus was sought in four different areas to guide the development of a methodological solution to analyse OS when trial participants receive subsequent anti-cancer treatment during trial follow-up, to ensure the solution was informed by stakeholder views.

The consensus exercise primarily aimed to identify the question stakeholders were the most interested in, when they knew that trial participants may have stopped their trial treatment and received subsequent treatment lines during trial follow-up. This was facilitated by considering the different methods available to address intercurrent events (Table 3.1). Alongside this, the consensus exercise aimed to establish agreement on: what and how data should be collected to answer the question of interest; which statistical assumptions may be appropriate in answering the question; and how the analysis should be presented. Finally, the online questionnaire also sought to collect information on the current practice for the analysis of OS from statisticians and other data analysts.

Table 3.1:Methods of Dealing with an Intercurrent Event

Method of dealing with an intercurrent event	Approach	Impact	Question*
Treatment Policy	Intercurrent event is considered part of the treatment strategy.		How does the new treatment extend survival compared to the control treatment; even though some participants stopped their trial treatment prior to death?
Hypothetical	The treatment effect for the hypothetical scenario where the intercurrent event is assumed to not have occurred is estimated.	Changes the interpretation of the results.	How would have the new treatment extended survival compared to the control treatment, if no one stopped their trial treatment prior to death?
Principal Stratum	The treatment effect in which the intercurrent event would (or would not) occur is estimated.	Subsets the population to only consider those who did (or did not) experience the intercurrent event.	How does the new treatment extend survival compared to the control treatment, in participants who only received their trial treatment prior to death?
Composite	The intercurrent event is incorporated into the endpoint definition.	Changes the definition of the endpoint.	How much longer did participants stay on the experimental treatment compared to the control treatment?
While on treatment / alive	The endpoint prior to treatment discontinuation is of interest	Changes the definition of the endpoint.	

** Note these Questions are as written in the questionnaire for Healthcare professionals, Payers and Industry Partners, and Statisticians and Data Analysts. The version for Patient, Carer or Member of the public can be seen at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM3_ESM.pdf page 17.*

3.4 Methods

The qualitative study utilised a modified Rand/UCLA appropriateness method[6], consisting of both an online questionnaire and a focus group (Appendix Figure C.1). As a result of the combined approach, both the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [7] and COREQ (Consolidated criteria for reporting qualitative research) [8] reporting guidelines have been used throughout the manuscript (Appendix Table C.1 and Appendix Table C.2).

3.4.1 Online Questionnaire

3.4.1.1 Development and Testing

The target population for the questionnaire were the stakeholders of clinical trials research. The stakeholders were categorised into: patients, carers and the public, health care professionals, statisticians and other data analysts, and payers and industry partners (Table 3.2).

Table 3.2:The Definitions of the Stakeholder Groups

Stakeholder Group	Definition
A Member of the Public*	For those who have not. • Treated patients with cancer • Analysed a cancer clinical trial • Discussed whether a treatment should be recommended for use on the NHS • Been diagnosed with cancer or cared for a family or close friend with cancer But are interested to learn more about cancer clinical trials and/or have a say in how they work and how their results are explained to your family, friends, and the public. If you are not sure which group to pick, everyone is a member of the public
A Patient or Carer*	For those who have cancer or have cared for someone with cancer and are interested to learn more about cancer clinical trials and/or have a say in how they work and how the results are explained to family, friends, and the public.
A Health Professional	For those with a medical background and who have experience in treating or discussing cancer treatment options with patients.
An Industry Partner#	For those who are employed by the pharmaceutical industry and work in the development of cancer treatments.
A Payer#	For those with experience in evaluating whether a new cancer treatment or intervention should be recommended for use in standard practice. For example, a member of a technology appraisal committee.
A Statistician or Other Data Analyst	For those with a statistical or mathematical background with experience in analysing cancer clinical trial data. This also includes Health Economists and Data Scientists.

**# indicate groups which were combined to make a total of four stakeholder groups*

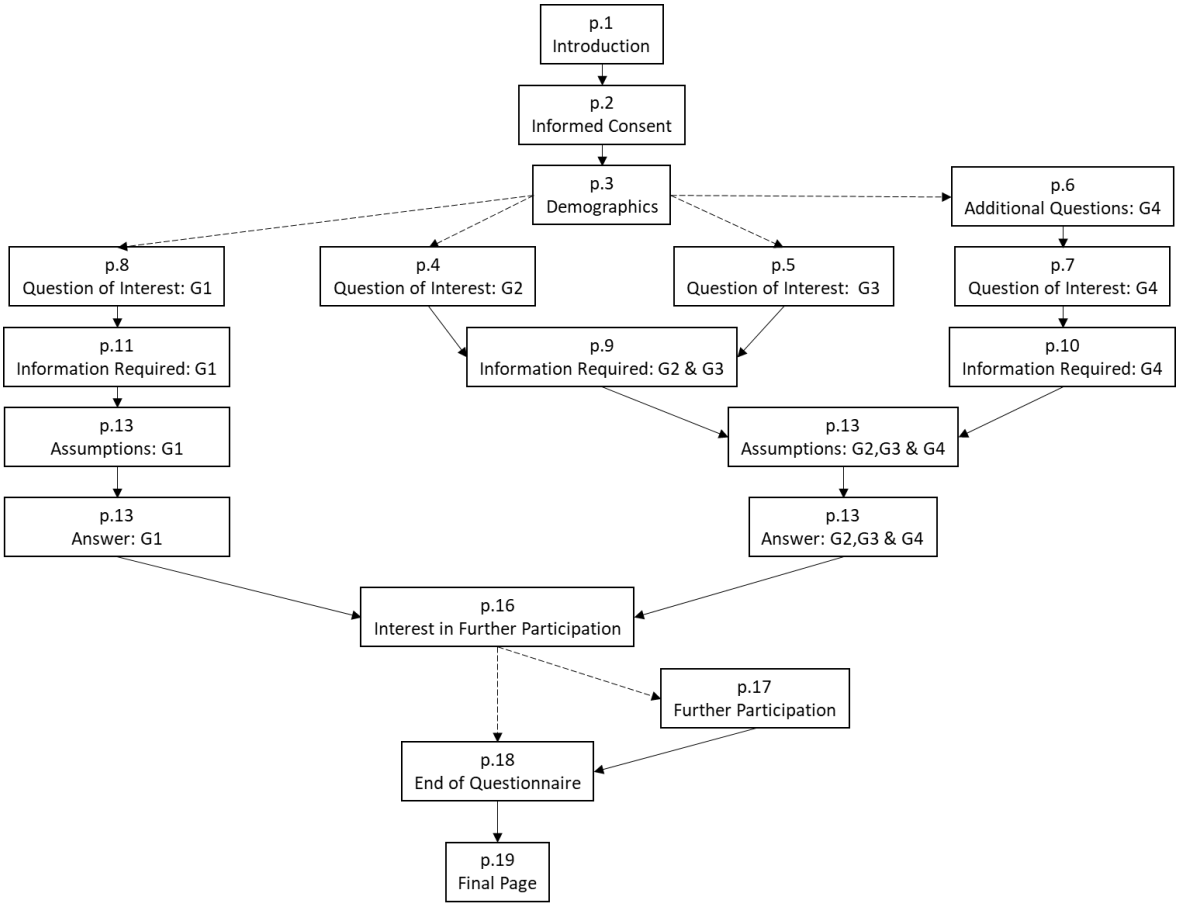
The questionnaire was co-designed with five patients (including PW) over two sessions. This considered the patient information sheet (PIS), the content and format of the questions, and questionnaire structure on Online Surveys (JISC, Bristol, UK). The draft questionnaire was reviewed and tested by the project's stakeholder advisory group (SAG). The SAG consists of at least one member of each stakeholder group. They meet and discuss key decisions about the study to ensure all stakeholder perspectives are considered.

In order to inform future methodological research, it was agreed that the questionnaire should include five core sections in addition to consent, demographics, request for optional further participation and any final comments.

1. Question of Interest: This section was included to inform what general question the future methodological research should investigate with particular attention on how to address participants stopping trial treatment and potentially commencing a subsequent treatment.
2. Information Required: This section was included to consider what data could be collected on participants following discontinuation of trial treatment and how it should be obtained as existing methodology e.g. two stage method requires additional information at the point of switching therapy to be used in model development. If any data required for a method was not acceptable, then the method would not be investigated in future research.
3. Assumptions: This section was included to understand perceptions on statistical assumptions, as all methods require assumptions and if any assumptions were not acceptable the methods requiring them wouldn't be considered in future work.
4. Answer: This section was included to be aware of how stakeholders like results presented to them so that any developed methodology could be presented in a way which would be understood.
5. Additional Questions: This section was included to gather information on how to disseminate the work to statisticians through statistical software and understand whether work would be required to change practice for example in terms of data collection.

The final questionnaire had 19 pages. However, this included optional pages and each of the four stakeholder groups had their own pathway through the questionnaire (Figure 3.1) and its nine sections (Table 3.3). Therefore, the total number of possible questions for an individual was 83. However, many of the questions were optional opportunities for the respondent to explain their answers. In addition, aside from consent, there was only one mandatory question. This asked respondents to self-identify with a stakeholder group (Table 3.2). This was mandatory as it determined the respondent's path through the questionnaire.

Figure 3.1: Questionnaire Structure



Flow diagram of questionnaire structure. Solid lines indicate the next page, dashed lines indicate a custom route. The custom routes from p.3 were determined by the stakeholder group the respondent self-defined themselves as (G1: Patients, carers and the public, G2: Healthcare Professionals, G3: Payers and Industry Partners, G4: Statisticians and other Data Analysts). The custom route from p.16 was determined by the respondent expressing an interest in further participation.

The participant information sheet (PIS) was the first page of the questionnaire and informed consent was obtained from respondents on the second page (Figure 3.1, Table 3.3). The PIS included links to YouTube videos summarising the study and the purpose of the questionnaire[9] and provided a study email address for queries. The PIS also included statements on data sharing and incentives for respondents. The questionnaire software was GDPR compliant and there was no monetary incentive for taking part. At the end of the questionnaire respondents could optionally consent to further participation. Respondents had the option to save their answers, complete it at a later date and move back through the questionnaire, but they could not view all their answers on a single page prior to submission.

The full questionnaire, including the PIS, can be viewed at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM3_ESM.pdf. Ethical approval for the project was obtained from an NHS Research Ethics Committee (22/YH/0155).

Table 3.3: Summary of the Online Questionnaire

Section	Summary	Pages[α]	Number of questions
Introduction and Informed Consent	An overview of the study with a frequently asked question section covering data protection risks and benefits of taking part. The informed consent had 7 points to agree to.	P2	7
Demographics	Twelve questions considering respondent demographics, baseline opinions on the topic and how they found the questionnaire.	P3	12
Additional Questions	Questions for statisticians and data analysts only, asking about current practices. This included what statistical software they use, what data has been collected during follow-up on their past trials and their previous analysis experience.	P6	12
Question of Interest	Ranking exercise of the question the respondent was most interested in with the option to explain and suggest additional questions. The pre-specified questions related to the five approaches to dealing with an intercurrent event (Table 3.1).	P4, P5, P7, P8	4
Information Required	List of six data items, with definitions, which could be collected on patients for respondents to give their opinions on whether they should be collected. There was also the option for respondents to explain their answers, suggest additional data to be collected and score how data should be collected and recorded. The pre-specified data items included date and cause of death, anti-cancer treatment, patient and disease characteristics.	P9, P10, P11	13*
Assumptions	List of seven statistical assumptions, with definitions, which could be applied for respondents to give their opinions on whether they were appropriate to be applied in all, some, or no scenarios. There was also the option for respondents to explain their answers after scoring each assumption. The pre-specified statistical assumptions included non-informative censoring, proportional hazards, common treatment effect, no unmeasured confounding, no time dependent covariates and no competing risks.	P12, P13	15
Answer	List of six ways, with definitions, data can be presented for respondents to give their opinion on whether they are helpful. There was also the option for respondents to explain their answers after scoring each option. The pre-specified	P14, P15	13[¥]

	data presentation options included survival curves, median time, mean survival, hazard ratios, confidence intervals and p-values.		
Further Participation	Option to hear about and consent to contact for additional participation in the focus groups and study updates.	P16, P17	6
End of Questionnaire	Opportunity for final comments.	P18	1

[α] The page number is as included on Figure 3.1
** For Healthcare professionals an extra question was included for them to explain their answers. For the other stakeholder groups this section only included 12 questions.*
¥ For non-patient/public groups this section only included 9 questions. Following the advice of the co-design group additional questions were included for the patient/public group to have the opportunity to explain their answers.
The complete questionnaire can be viewed at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM3_ESM.pdf which includes all definitions and explanations.

3.4.1.2 Recruitment

The questionnaire was distributed electronically using targeted emails. The questionnaire was an open survey i.e. registration was not required; participants self-selected and accessed the questionnaire through a link. The targeted emails were sent to all stakeholder groups and included charity groups, funding bodies and clinical trial units (Appendix Table C.3). In addition to the targeted emails, social media was utilised to promote the study and associated questionnaire. At two-weekly intervals during recruitment, threads introducing different topics were shared on X (previously called Twitter). These included introductions to clinical trials (Appendix Figure C.2), statistics (Appendix Figure C.3), and survival analysis (Appendix Figure C.4). Two summary videos about the study were also made for X.[10, 11] Recruitment was for eight weeks. A formal sample size calculation was not applicable for the questionnaire. However, a target of 30 respondents per stakeholder group was pre-specified. This equated to 120 respondents, which balanced the desire to collect a wide variety of views with conducting a qualitative analysis within the project timescales.

3.4.1.3 Analysis

The online survey software did not use Cookies, check Internet Protocol (IP) addresses or provide a log file for analysis to identify potential duplicate responses. Therefore, any unidentifiable duplicates have been considered as separate responses. Furthermore, whilst the total number of visits to each page of the questionnaire was recorded, there was no way to determine the number of unique visits. Instead, the total number of completed (submitted) questionnaires was compared to the total number of respondents who provided informed consent.

At the end of recruitment, the questionnaire responses were downloaded and read into SAS 9.4 (SAS Institute, Cary, NC). Analysis was descriptive. Themes were identified from the free-text fields using a conventional content analysis approach.[12] Prior to the analysis, consensus guidelines, based on the RAND/UCLA method, were agreed between research team (K-LR, DM, JV-R, IRW and DAC) (Table 3.4). Questionnaires with missing responses were included in the analysis, but missing answers did not feed into the consensus guidelines. Statistical correction was not applied to adjust for a non-representative sample.

Table 3.4: Questionnaire Decision Guidelines

Guidelines	Question Type	Guidelines
Question Decision Guideline	Ranked (1-4)	Include - Median = 1 Uncertain - Median = 2-3 or disagreement Exclude - Median = 4 Disagreement: ≥ 1/3 in 1 and ≥1/3 in 4
Data Decision Guideline	Likert (1-5)	Include - Median = 1-2 Uncertain - Median = 3 or disagreement Exclude - Median = 4-5 Disagreement: ≥ 1/3 in 1 and ≥1/3 in 5
Choice Guideline	Multiple Choice (Choose 1/2)	The one with the highest percentage of respondents would be considered for further discussion.
Assumptions Guideline	Likert (1-3 + Unsure)	Include - Median = 1 Uncertain - Median = 2 or disagreement Exclude - Median = 3 Disagreement: ≥ 1/3 in 1 and ≥1/3 in 3 OR 50% in unsure
Answer Guideline	Likert (1-5)	Include - Median = 1-2 Uncertain - Median = 3 or disagreement Exclude - Median = 4-5 Disagreement: ≥ 1/3 in 1 and ≥1/3 in 5

3.4.2 Focus Group

3.4.2.1 Design

Following the analysis of the online questionnaire, the results were discussed and interpreted within the research team and a proposal was presented to the SAG (including PW) as to which aspects needed further discussion (available at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM4_ESM.pdf). Once agreed, the focus group materials (a pre-focus group information pack and a post-focus group questionnaire) were finalised in collaboration with the research team and SAG. The pre-focus group information pack, included an overview of the questionnaire results, which aspects reached consensus from the pre-defined rules, and a plan of what would be asked and discussed at the focus group (available at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM5_ESM.pdf). The post-focus group questionnaire formalised the recommendations made within the focus group by asking attendees to re-rate the points discussed during the focus group and allowed for feedback outside the wider group (available at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM6_ESM.pdf). Ethical approval for focus group materials was obtained via a substantial amendment.

Focus group invitees were randomly selected from the questionnaire respondents who consented to be contacted for further participation. The focus group aimed to have 12 attendees equally balanced by stakeholder group and sex. Therefore, consenting respondents were stratified by stakeholder group and gender. Invitees were then randomly selected from each stratum. The random sampling was conducted using the procedure surveyselect in SAS 9.4. Any questionnaire respondents who declined were replaced by resampling from the relevant stratum.

3.4.2.2 Format

The focus group was a hybrid meeting with K-LR chairing (a female PhD student with an MSc in Statistics). Prior to the focus group, K-LR had attended focus groups with different stakeholder groups and discussed with colleagues their experience of leading a focus group. PW facilitated the focus group in person with K-LR. Attendees joined online using Microsoft Teams (Microsoft, Redmond, Washington, USA). Breakout rooms were utilised to allow for within and across stakeholder discussion. K-LR did not input into the discussion to reduce the possibility of bias. During the focus group, attendees could write notes using Padlet (San Francisco, California, U.S. Singapore). The focus group was scheduled for three hours. As the attendees had completed the questionnaire they knew about the researcher (K-LR), and the reasons for the research from the PIS. Following the focus group, attendees were asked to complete the post-focus group questionnaire online.

3.4.2.3 Analysis

The focus group recording and notes were summarised. The post-focus group questionnaire was analysed descriptively by K-LR where free text was considered in full. The meeting summaries and answers to the questionnaire were used to finalise consensus, where possible, and provide guidance for the next step of the project.

3.5 Results

3.5.1 Online Questionnaire

Questionnaire recruitment took place between 3rd August 2022 and 30th September 2022. There were 1808 visits to the introduction page, 314 gave informed consent and 103 stakeholders completed the questionnaire, resulting in a completion rate of 33% (patient, carer, or member of the public: 45/103 (44%), health professional: 22/103 (21%), payer or industry partner: 6/103 (6%), statistician or other data analyst: 30/103 (29%)).

Results are presented as per the questionnaire sections in Table 3.3.

3.5.1.1 Demographics

The respondents predominantly found the questionnaire from a mailing list 35/103 (34%), had a median age of 52, were mostly female 63/103 (61%), from the UK and NI 84/103 (81%) and had some prior experience of clinical trials 68/103 (66%). When asked about their opinion on the topic 81/103 respondents (78.6%) thought that subsequent treatments should be taken into consideration when analysing OS (Table 3.5).

Table 3.5: Summary of Questionnaire Respondents

	Patient, Carer or Member of the Public (n=45)	Healthcare Professional (n=22)	Payer or Industry Partner (n=6)	Statistician or Other Data Analyst (n=30)	Total (n=103)
How old are you?					
Median (IQR)	60 (54,66)	48 (38,55)	43 (37,48)	40 (30,48)	52 (40,60)
Missing	1	0	0	3	4
What is your ethnic group?					
English/Welsh/Scottish/Northern Irish/British	41 (91.1%)	17 (77.3%)	6 (100.0%)	20 (66.7%)	84 (81.6%)
Irish	0 (0.0%)	2 (9.1%)	0 (0.0%)	2 (6.7%)	4 (3.9%)
African	0 (0.0%)	1 (4.5%)	0 (0.0%)	0 (0.0%)	1 (1.0%)
Caribbean	1 (2.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.0%)
Other, please describe	3 (6.7%)	1 (4.5%)	0 (0.0%)	2 (6.7%)	6 (5.8%)
Prefer not to say	0 (0.0%)	1 (4.5%)	0 (0.0%)	3 (10.0%)	4 (3.9%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (10.0%)	3 (2.9%)
What sex are you?					
Male	10 (22.2%)	9 (40.9%)	2 (33.3%)	12 (40.0%)	33 (32.0%)
Female	33 (73.3%)	12 (54.5%)	4 (66.7%)	14 (46.7%)	63 (61.2%)
Prefer not to say	1 (2.2%)	0 (0.0%)	0 (0.0%)	1 (3.3%)	2 (1.9%)
Missing	1 (2.2%)	1 (4.5%)	0 (0.0%)	3 (10.0%)	5 (4.9%)
Where in the UK are you based?					
Scotland	4 (8.9%)	1 (4.5%)	0 (0.0%)	2 (6.7%)	7 (6.8%)
Northern Ireland	1 (2.2%)	0 (0.0%)	0 (0.0%)	2 (6.7%)	3 (2.9%)
North West England	4 (8.9%)	1 (4.5%)	1 (16.7%)	2 (6.7%)	8 (7.8%)
North East England	1 (2.2%)	1 (4.5%)	0 (0.0%)	0 (0.0%)	2 (1.9%)
Yorkshire and The Humber	18 (40.0%)	12 (54.5%)	2 (33.3%)	10 (33.3%)	42 (40.8%)
East Midlands	4 (8.9%)	2 (9.1%)	0 (0.0%)	0 (0.0%)	6 (5.8%)
West Midlands	1 (2.2%)	0 (0.0%)	0 (0.0%)	3 (10.0%)	4 (3.9%)
London	1 (2.2%)	3 (13.6%)	3 (50.0%)	5 (16.7%)	12 (11.7%)
South East England	6 (13.3%)	0 (0.0%)	0 (0.0%)	2 (6.7%)	8 (7.8%)
South West England	3 (6.7%)	1 (4.5%)	0 (0.0%)	1 (3.3%)	5 (4.9%)
Other	2 (4.4%)	0 (0.0%)	0 (0.0%)	2 (6.7%)	4 (3.9%)
Prefer not to say	0 (0.0%)	1 (4.5%)	0 (0.0%)	0 (0.0%)	1 (1.0%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.3%)	1 (1.0%)
Where did you find out about this questionnaire - Mailing List					
Selected	10 (22.2%)	11 (50.0%)	2 (33.3%)	12 (40.0%)	35 (34.0%)
Where did you find out about this questionnaire - Social Media					
Selected	21 (46.7%)	3 (13.6%)	0 (0.0%)	0 (0.0%)	24 (23.3%)
Where did you find out about this questionnaire - Word of Mouth					
Selected	4 (8.9%)	0 (0.0%)	1 (16.7%)	6 (20.0%)	11 (10.7%)
Where did you find out about this questionnaire - Other					
Selected	11 (24.4%)	8 (36.4%)	3 (50.0%)	11 (36.7%)	33 (32.0%)
Are You?					

	Patient, Carer or Member of the Public (n=45)	Healthcare Professional (n=22)	Payer or Industry Partner (n=6)	Statistician or Other Data Analyst (n=30)	Total (n=103)
A Member of the Public	13 (28.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	13 (12.6%)
A Patient or Carer	32 (71.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	32 (31.1%)
A Health Professional	0 (0.0%)	22 (100.0%)	0 (0.0%)	0 (0.0%)	22 (21.4%)
An Industry Partner	0 (0.0%)	0 (0.0%)	2 (33.3%)	0 (0.0%)	2 (1.9%)
A Payer	0 (0.0%)	0 (0.0%)	4 (66.7%)	0 (0.0%)	4 (3.9%)
A Statistician or Other Data Analyst	0 (0.0%)	0 (0.0%)	0 (0.0%)	30 (100.0%)	30 (29.1%)
Have you had any prior experience of clinical trials?					
Yes	20 (44.4%)	16 (72.7%)	4 (66.7%)	28 (93.3%)	68 (66.0%)
No	25 (55.6%)	6 (27.3%)	2 (33.3%)	2 (6.7%)	35 (34.0%)
Do you think it is important to consider the effect that treatment given after trial treatment has on overall survival?					
Yes	40 (88.9%)	16 (72.7%)	6 (100.0%)	19 (63.3%)	81 (78.6%)
No	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.3%)	1 (1.0%)
Depends	3 (6.7%)	4 (18.2%)	0 (0.0%)	8 (26.7%)	15 (14.6%)
Unsure	2 (4.4%)	2 (9.1%)	0 (0.0%)	2 (6.7%)	6 (5.8%)

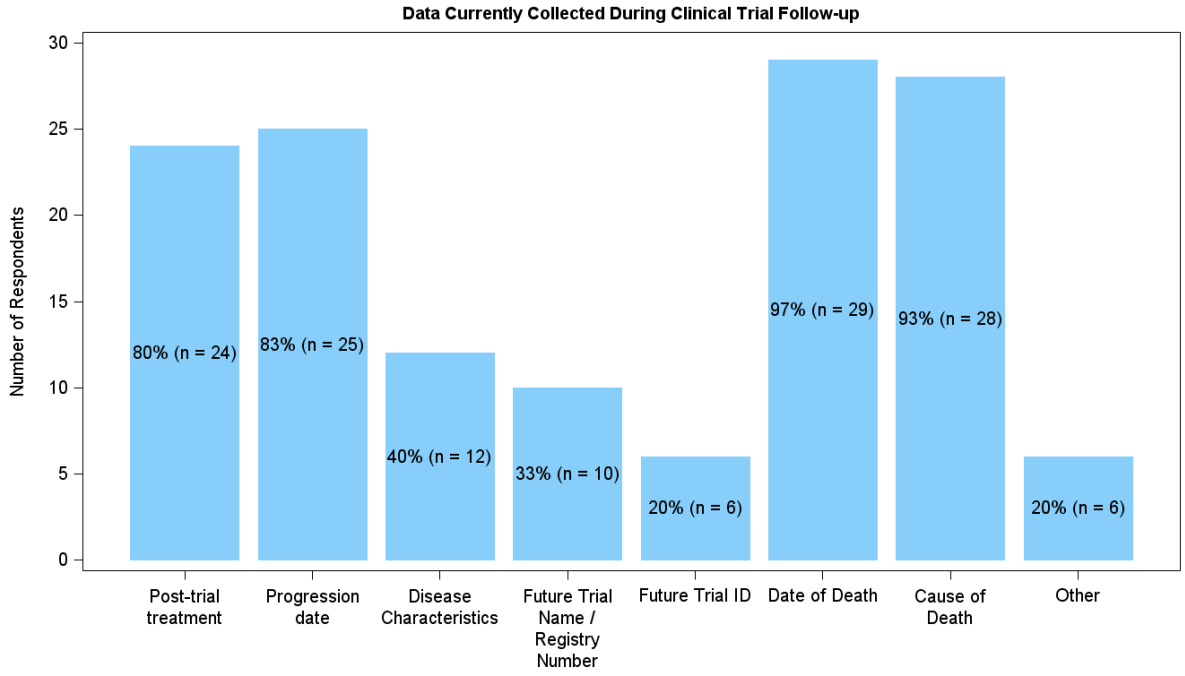
3.5.1.2 Additional Questions

In total 30 respondents identified themselves as a statistician or other data analyst and answered the additional questions about their previous experiences analysing OS.

The majority of the 30 respondents used the statistical programming languages STATA or SAS (22/30, 73.3%) to conduct statistical analysis. Seven (23.3%) used the language R and one respondent did not complete the question.

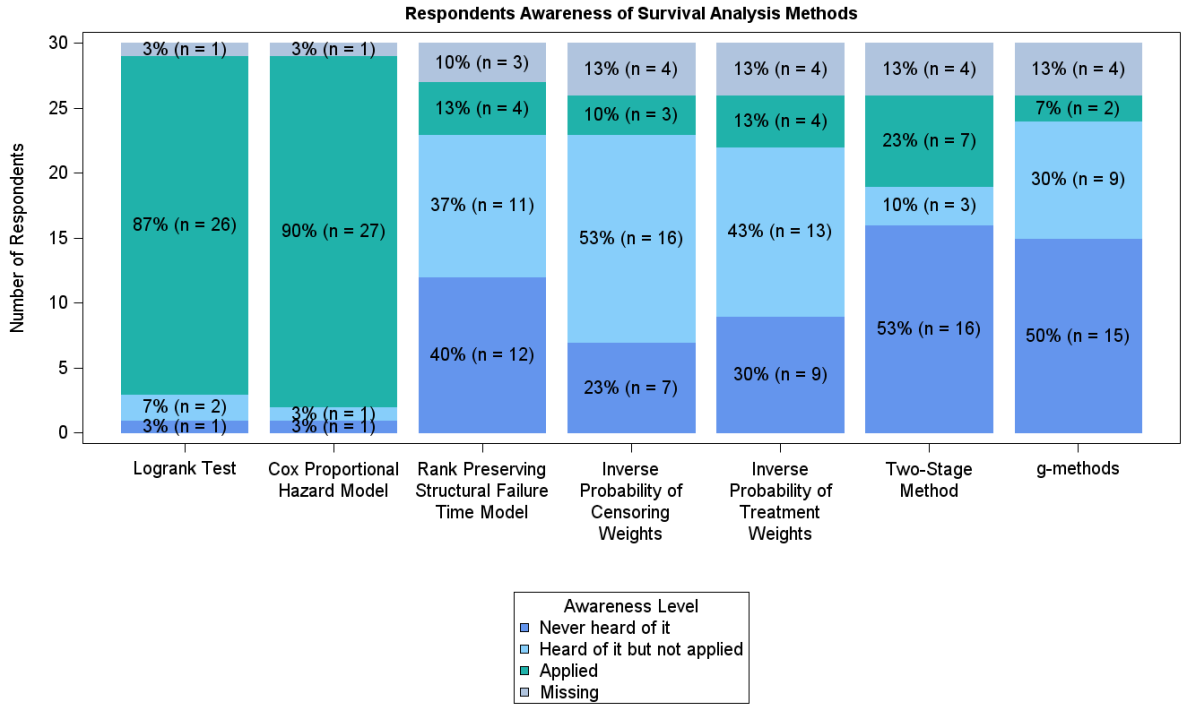
In terms of data collection, post-trial treatment, progression date, and date and cause of death were reported as being collected during trial follow-up by a very high proportion of respondents (Figure 3.2). However, disease characteristics and participation in future trials along with trial identification number were not selected as often. When asked for other data which is collected, respondents answered with quality of life (QoL) and other safety information. It was also noted that "*This is really hard to answer as it really depends on the nature of the index trial/length of follow-up*" and "*could be any or all of these - there isn't one standard*" highlighting that there is not a common approach to data collection across trials.

Figure 3.2: Bar Chart Showing the Number of Statisticians or Other Data Analysts who Collect Information During Trial Follow-up



On consideration of the statistical methods used to analyse OS, often referred to as survival analysis, most respondents, as expected, had applied the logrank test and the Cox Proportional Hazards model which were the most widely used survival analysis methods used in a literature review on the topic.[5] Respondents had heard of, but not applied, IPCW (inverse probability of censoring weighting) and IPTW (inverse proportional treatment weighting) and had not heard of the RPSFTM (rank preserving structural failure time model), the two-stage method or g-methods (Figure 3.3). This is not surprising given that these methods are more complex and applied in specific scenarios like treatment switching. In fact, only (5 / 30) 16.7% of respondents had adjusted OS analysis to account for subsequent lines, with the two-stage method, landmark analysis and RPSFTM being specifically mentioned. Eight respondents took up the opportunity to elaborate on their answers. One respondent commented "*Also used semiparametric methods such as spline-based APC models, accelerated failure time, proportional odds and frailty models*". These again are considered more advanced survival analysis methods and were rarely reported as being implemented in the literature review.

Figure 3.3: Bar Chart Showing Statisticians and Other Data Analysts' Awareness of Methodology

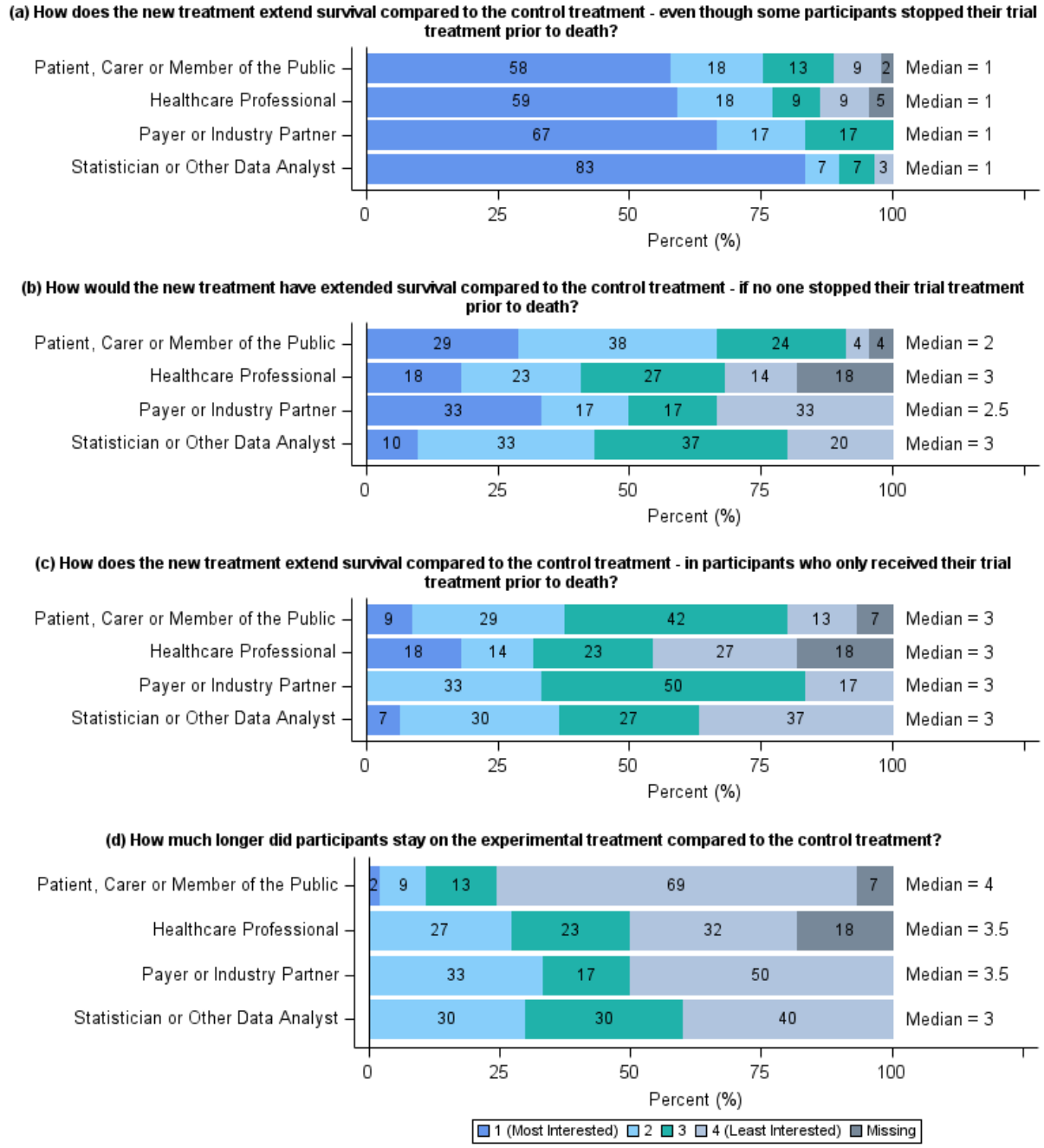


The Logrank Test and the Cox Proportional Hazard Model are statistical methods which are widely used in the literature. The remaining methods are less widely used in the literature.

3.5.1.3 Question of Interest

Following the pre-specified guidelines (Table 3.4) consensus was reached that the question of interest to each stakeholder group was "*How does the new treatment extend survival compared to the control treatment - even though some participants stopped their trial treatment prior to death?*" (Figure 3.4). This relates to the treatment policy method of addressing an intercurrent event (Table 3.1). The main themes from the additional questions posed from respondents were, details on the subsequent treatments, consequences of being in the trial, other effects of the trial treatment and how patient characteristics influenced the treatment effect. At the SAG, it was agreed that the decision for the research to consider the first question should be discussed at the focus group.

Figure 3.4: Summary of Question of Interest Results



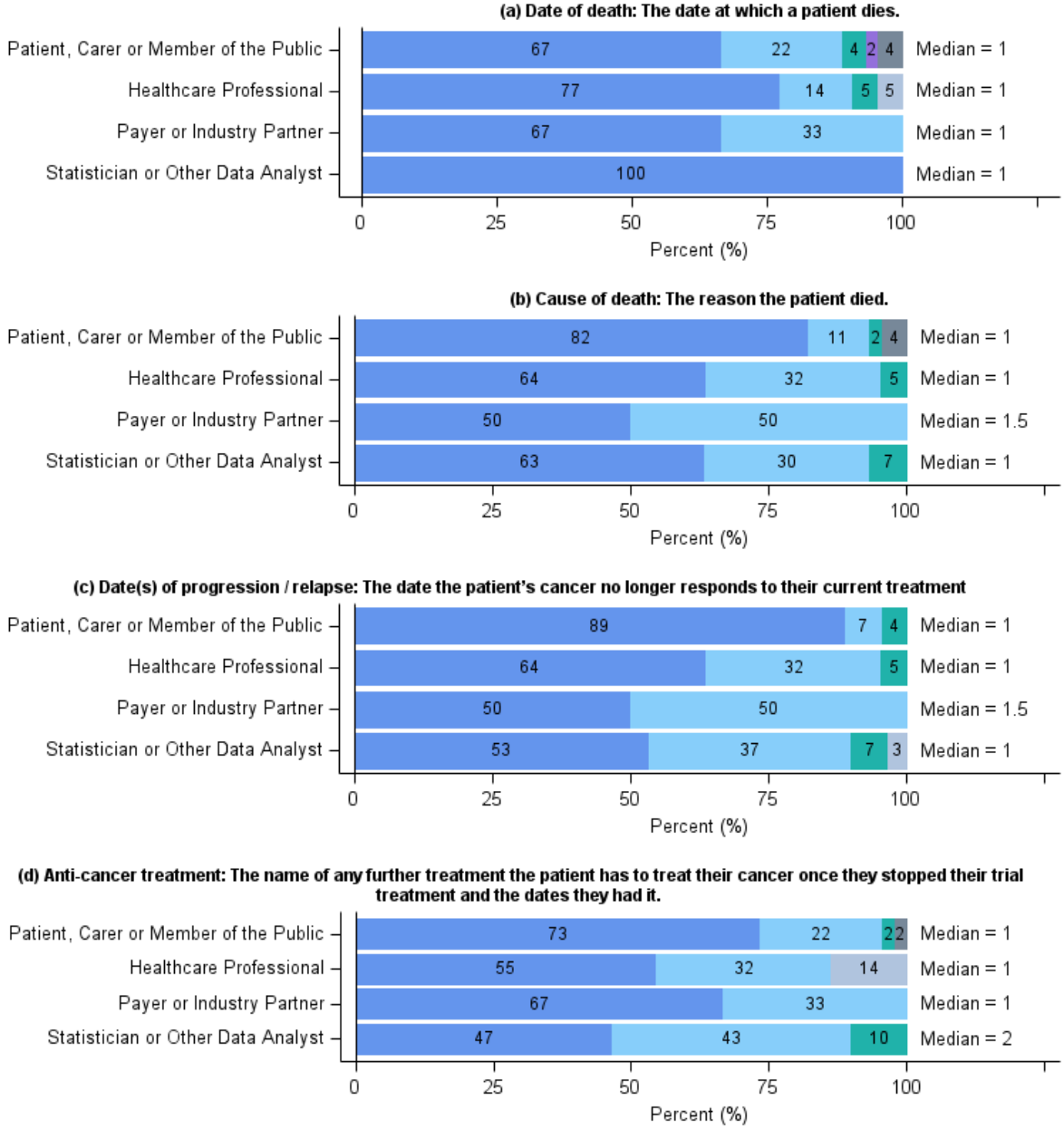
Horizontal stacked bar charts showing the responses to the question "Please read the five questions below and rank them in order of what would be the most important for you to know the answer to (1) to the least important (4) if you or a friend were being treated for cancer. Please remember this is asking what you think, there are no right or wrong answers." from the online questionnaire separated by stakeholder group.

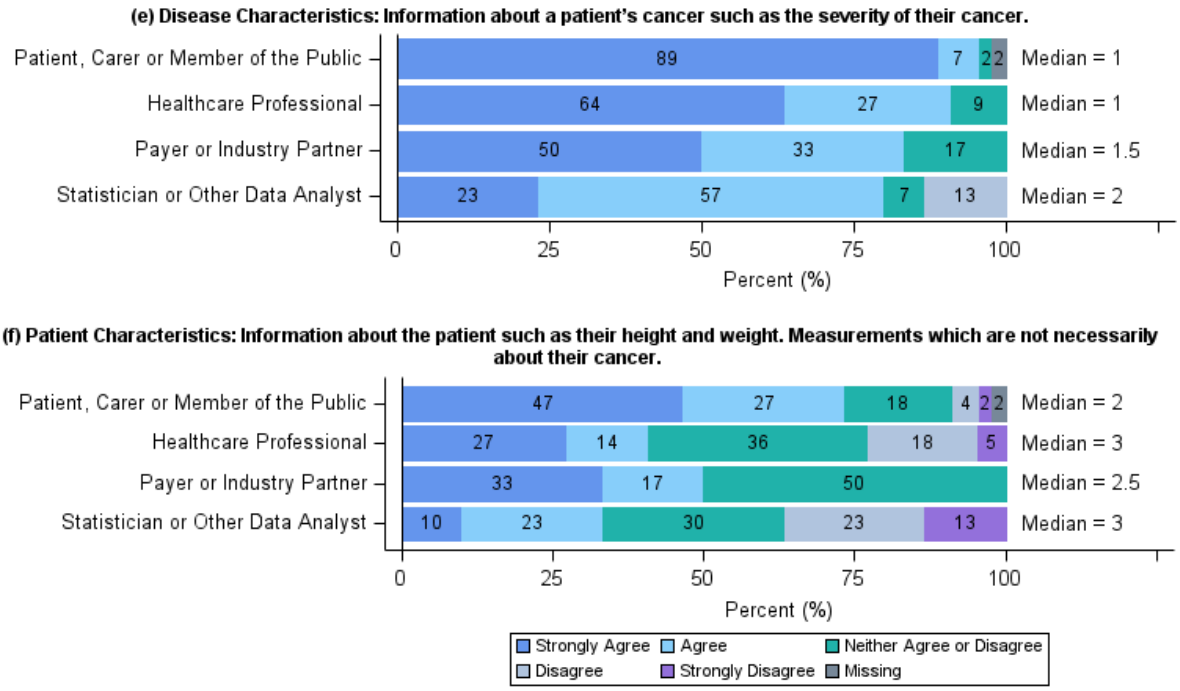
3.5.1.4 Information Required

Consensus was reached that date of death, cause of death, date(s) of progression/relapse, subsequent anti-cancer treatment and disease characteristics could be collected on trial participants once they have stopped their trial treatment (Figure 3.5). However, there was disagreement around patient characteristics. Therefore, the collection of patient characteristics along with the collection of co-morbidities, concurrent treatment, QoL / psychological data, toxicity and participation in

future trials which were suggested by questionnaire respondents, were agreed to be discussed at the focus group.

Figure 3.5: Summary of Data Collection Responses



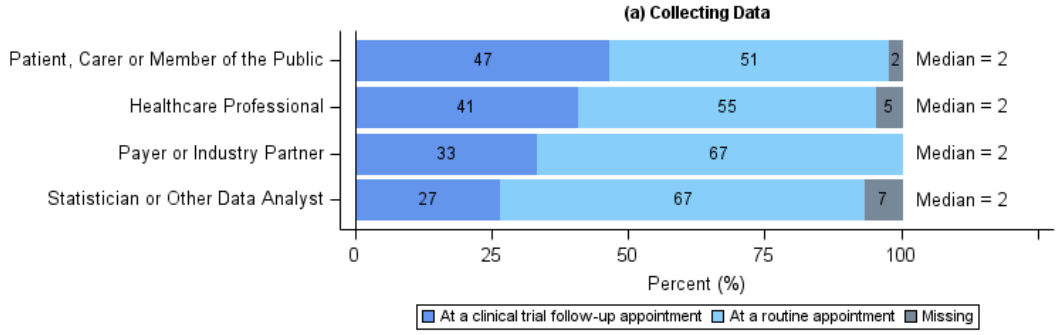


Horizontal stacked bar charts showing the responses to the question "Please consider each piece of information listed below and decide whether you agree it is practical and appropriate to collect this information about clinical trial participants, after they have stopped receiving their trial treatment." from the online questionnaire separated by stakeholder group.

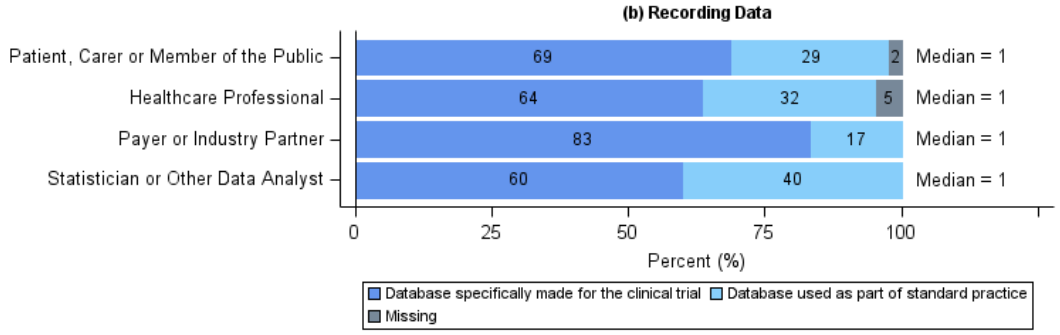
When asked how data should be collected and recorded on participants once they have stopped their trial treatment, the majority of stakeholders thought that data should be collected at a routine appointment (Figure 3.6). However, reasons were provided for both routine and trial follow-up appointments. The main reasons provided in favour of a trial follow-up appointment were: trial focus, the respondents' explanation mentions that a trial follow-up appointment is conducted by the research team who will keep focus on trial and ensure data collection is completed; phone preference, the respondents' explanation mentions that a phone appointment is preferable to a face-to-face appointment to reduce burden on patient; and personal experience, the respondents' explanation includes reference to their own opinion and experience of trials. On the other hand, the main reasons provided in favour of a routine appointment were: patient burden, the respondents' explanation includes reducing the burden and requirement on patients to attend hospital visits in order to improve compliance; context, the respondents' explanation notes that the regularity of follow-up appointments would influence their decision on a trial-by-trial basis; and logistics, the respondents' explanation includes a consideration of how it would work in practice. Similarly, whilst most stakeholders felt that data should be recorded on a database which was made specifically for the trial (Figure 3.6), reasons were provided for both options. The main reasons provided in favour of a trial database were: standardised data collection, the reasoning includes that the required data is collected reliably and consistently in a controlled way to allow for it to be as accurate as possible to ensure a robust analysis

and routine data access, the reasoning includes a comment around the barriers to accessing data from routine data sources. Alternatively, the main reasons provided in favour of a standard database were: site burden, the respondents' explanation includes a comment about reducing the pressure on site staff to input data; duplication, the respondents' explanation includes a comment around reducing duplication across databases; and data sharing, the respondents' explanation favours data sharing between general practice and researchers. Following discussion with the SAG it was agreed that the collection and recording of data once trial treatment has stopped should be discussed with the focus group and extended to consider what is currently possible, and what would be ideal if the resources were available.

Figure 3.6: Summary of Collecting Data and Recording Information Results



Horizontal stacked bar chart showing the responses to the question "There are two different ways this information can be collected from patients who have stopped their trial treatment. Which do you think is the most appropriate?" from the online questionnaire separated by stakeholder group.



Horizontal stacked bar chart showing the responses to the question "There are also two different ways this information can be recorded and returned to the trial researchers. Which do you think is the most appropriate?" from the online questionnaire separated by stakeholder group.

3.5.1.5 Assumptions and Answer

In terms of statistical assumptions, the majority were scored as being appropriate in "*Some Scenarios*" (Appendix Figure C.5). The results were discussed with the SAG and decided not to be taken to the focus group. Similarly, as most ways to present the information were scored as "*Very Helpful*" or "*Helpful*" (Appendix Figure C.6), it was agreed that data presentation would not be discussed further. Instead, any future work will be presented to stakeholders for review and updated accordingly.

3.5.1.6 Summary

To summarise, following the online questionnaire there were three discussion points for the focus group:

1. The decision to consider the treatment policy approach of addressing the intercurrent event with the question "*How does the new treatment extend survival compared to the control treatment - even though some participants stopped their trial treatment prior to death?*"
2. The data items that are collected once trial treatment has finished.
3. How data is collected and recorded once trial treatment has finished.

These three discussion points were included with the full results of the questionnaire in a pre-meeting discussion pack (available at https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM5_ESM.pdf). The first point was presented with the researcher's opinion as consensus had been reached within the initial questionnaire based on the pre-defined criteria (Table 3.4). The remaining were presented as open-ended questions. The schedule of the focus group was also included in the pre-meeting pack and included the explicit questions the focus group would be discussing (https://static-content.springer.com/esm/art%3A10.1186%2Fs13063-025-09148-3/MediaObjects/13063_2025_9148_MOESM5_ESM.pdf).

3.5.2 Focus Group

Just under half of respondents (50/103, 48.5%) were interested in learning about further participation in the study. This resulted in 33 (32%) of respondents consenting to being contacted for attendance at discussions with other respondents in a different stakeholder group. The 33 respondents predominantly identified as a patient, carer, or member of the public (17/33, 51.5%). The other respondents identified as healthcare professional (4/33, 12.1%), payer or industry partner (2/33, 6.1%) and statistician or other data analyst: (10/33, 30.3%). In total 19 respondents were invited (Appendix

Table C.4). Of the eight who declined, all were not available on the day of the focus group. One of the respondents who declined sent a colleague in their place. For this attendee, consent was obtained via email as they had not completed the original questionnaire. In total the focus group had seven attendees, balanced for stakeholder group but not for sex (Appendix Table C.4).

3.5.2.1 Discussion 1: The question of interest

The focus group considered the researcher's decision to consider the treatment policy approach by focusing on the question "*How does the new treatment extend survival compared to the control treatment - even though some participants stopped their trial treatment prior to death?"*. All stakeholder groups agreed that this was a logical approach and that it reflected real life. A patient member commented "*When you were talking about reality, that's a word that resonated. I suppose being slightly suspicious in general life as a patient you would wonder what had happened to the other people, the special people who we weren't being reported back on."* However, patients did comment that QoL would still be of interest but accepted that was not the focus of the study.

These thoughts in terms of logic and interest in QoL were re-iterated by the five attendees (three statisticians and two patients) who completed the post-focus group questionnaire. The three statisticians responded: "*I think this is a choice which makes sense both to clinical experts and to patients. It [does] not preclude pre-specified analyses of narrower questions."*; "*I think [this] is the best focus as it focuses on the pragmatic realistic question that is of most interest to patients and clinicians*"; and "*I think this is the most common question of primary interest, so it makes sense to focus on this. Although context might dictate that considering the other options is more or less relevant, I would imagine this question would have greatest interest.*" The two patients responded: "*I'm happy with this as it reflects the real world not some artificial scenario*" and "*Quality of life should be taken into account. Survival might be longer but at what cost to quality of life?*".

3.5.2.2 Discussion 2: Data Items

The focus group considered the collection of: patient characteristics, whether participants received subsequent anti-cancer treatment as part of a clinical trial, QoL / psychological data and toxicity data, once the trial participant has stopped taking their trial treatment. All stakeholder groups agreed that collecting data from trial participants is acceptable when: there is a clear rationale, patients have been informed during the consent process; and patients are involved in the decision to collect the data. In the post-focus group questionnaire all items reached consensus to be included overall

(median 1-2,Table 3.4) and the majority reached consensus within each stakeholder group (Appendix Figure C.7). Only patient characteristics did not meet the pre-defined consensus guidelines within the Statistician or Other Data analyst group (Appendix Figure C.7). This reiterates the point made during the focus group discussion from this group that there needs to be a clear rationale for data collection.

3.5.2.3 Discussion 3: Data Collection and Recording

When thinking about how data is collected, perhaps months or years after the participant has stopped taking their trial treatment, the perspectives from each stakeholder group focused on reducing participant and site burden and encouraging the use of remote data capture over the telephone where possible. In the focus group all groups agreed that routinely collected data would be ideal, but understood the statisticians concerns around data access requirements changing during the trial and that the data needed for the trial may not be collected routinely potentially resulting in research waste. A statistician summarised the issue as "*That's what it comes down to, the whole thing is a waste if we don't get the right data*".

In the post-focus group questionnaire consensus was reached within each stakeholder group for collecting information in the current research world with Patients, Carers or Members of the Public preferring a trial follow-up appointment and Statisticians or Other Data Analysts preferring a routine follow-up appointment. In an ideal research world both stakeholder groups preferred a trial follow-up appointment to collect data (Appendix Figure C.8). This reflected some of the discussions within each stakeholder group during the focus group. The Patient, Carer and Member of Public group raised the context around the disease. They noted that if routine appointments do not happen as part of standard of care then a trial appointment was necessary. In addition, they commented that if the information is only needed for trial purposes rather than routine care that the trial team should collect it to reduce the burden on sites. The Statisticians and Other Data Analysts also commented that it would make sense to use the routine appointments, but that it depended on the standard of care visit schedule and whether it reflects the follow-up schedule required for the trial outcomes.

In terms of recording information, consensus was reached across the stakeholder groups for recording information in the current research world, with all groups preferring a database specifically made for the trial (Appendix Figure C.9). However, in the ideal research world whilst Statisticians or Other Data Analysts preferred recording information using a database which is completed normally as part of standard practice. Patients, Carers or Members of the Public were evenly split. However, if considered together the consensus is to use a database completed normally as part of standard

practice. This reiterates the discussion in the focus group, where patients did not mind where the data was stored if it was accessible and minimised patient and clinician burden, and statisticians wanted to use routine data if it collected the correct information and was accessible.

Summary

Following the summary of the discussions and analysis of the post-focus group meeting questionnaire, it was concluded that sufficient information and perspectives had been obtained to inform future methodology work but that wider conversations on data storage and recording would be needed in the future. Therefore, no further rounds (pre-focus group questionnaire, focus group and post-focus group questionnaire) were conducted.

3.6 Discussion

The overarching aim of this qualitative study was to obtain consensus from stakeholder groups, to guide the development of a methodological solution to analyse OS when trial participants receive subsequent treatment lines during trial follow-up, which is an intercurrent event. This was achieved as consensus was reached on the majority of topics, and the future direction of the project was agreed at the focus group. As no additional topics, within the scope of the project, were raised from the questionnaire or the focus group, data saturation was concluded to have been reached.

Following both parts of the study, it was agreed that future work should consider the question: "*How does the new treatment extend survival compared to the control treatment - even though some participants stopped their trial treatment prior to death?*" Important caveats were that the question does not: Assume no one stopped their trial treatment prior to death; consider only those who only received their trial treatment prior to death or shorten overall survival to be the time on treatment. Considering the estimand framework, this question relates to the treatment policy method of dealing with intercurrent events. There are two possible approaches to this question. The first is to ignore subsequent treatment lines in the estimation of the trial treatment effect for OS. This is already done in practice[5] and is the definition of the treatment policy approach. As most stakeholders thought that the potential effect of subsequent treatment lines should be considered during the analysis of OS, an alternative approach will be explored in the next step of this work utilising both a simulation study and real data. The next step will consider the hypothetical scenario assuming all participants, where appropriate, went on to receive the same second-line of therapy. This will aim to provide a sensitivity analysis for the treatment policy approach, to allow researchers to begin to quantify the impact of subsequent treatments. This is thought to

be predominantly of interest in the scenario where the treatment landscape has changed during trial-follow-up and there are more effective options available for patients. This decision could be criticised in that stakeholders have not been explicitly asked to comment on this direction. However, from the results of the general questionnaire all stakeholders agreed that the subsequent treatment lines should be considered, and this question is considered a natural next step by the investigators. It is also unknown whether an answer to the question can be obtained. Following an investigation via a simulation study and application to real data, stakeholders will be invited to comment again and future directions, including alternative estimands, agreed collaboratively. The aim of this stakeholder consultation was to understand current opinions, ensure that this research area was of interest to all and guide the future research which have been met.

The future work will be guided using the results from the qualitative study presented here. Only data which was scored "*Appropriate*" with a clear rationale to collect will be simulated, and different scenarios considering different levels and routes of data collection during post-trial treatment will be used to compare how the different methods perform in different contexts before applying them existing trial data. The consideration of different contexts is also important when applying the methods themselves and acknowledging the statistical assumptions they use. All stakeholders agreed that all assumptions were applicable in "*Some Scenarios*" and it was noted that one method or approach may not fit all trial designs. The use of alternative data generating mechanisms will help to determine which scenarios the assumptions are applicable in, to help future researchers decide which method they should use when considering the context of their trial.

Once complete the best way to describe and present the future work to the different stakeholder groups will be discussed and agreed with the SAG and consenting questionnaire respondents. Using the questionnaire results, a visual representation, a summary statistic such as a median and a hazard ratio which will be supported by at least a confidence interval will aim to be presented as a starting point for discussion.

Following the responses from the Statisticians and Other Data Analysts, to support the impact of any developed methodology it will be first created in Stata (18, StataCorp LLC, College Station, TX) and then translated into SAS (SAS Institute, Cary NC) and R (R Core Team, Vienna, Austria), where appropriate. In addition, as many of the statisticians and other data analysts had not heard of the more advanced methods currently available to tackle the specific event of treatment-switching (Figure 3.3), a variety of different methods of dissemination will be utilised. This will be planned with the SAG and will include pre-conference workshops, conference abstracts, journal

articles and social media to improve uptake of the work. It is important to note that the questionnaire and focus group also identified QoL as a key area for future research. Whilst QoL is outside the scope of this project, patients, carers, and the public were interested to know how the long-term QoL compared to long-term survival so that the information could be used in combination for patients with differing priorities to make an informed decision about their treatment. Another avenue for future research, which is outside the scope of this project, is trial design methodology. This work is focusing on the analysis methods. However, trialists may want to consider how trials could be designed to address the problem of subsequent treatment lines upfront. An example of an area to consider could be the implementation of trials with a SMART (Sequential, Multiple Assignment, Randomized Trial) design.[13] Furthermore, the discussions around how information is collected and recorded during trial follow-up is an interesting are to pursue further. However, the current movement in this area to store health systems data in secure/trusted research environments needs to be completed first before data availability and access can be considered. At present the only approach to consistently collect this data is through trial specific case report forms which requires consideration at trial outset.

In terms of the general process, the study follows a modified RAND/UCLA appropriateness method (RAM) which may be seen as a limitation as it deviated from the traditional RAM approach. The traditional RAM invites a small group of experts to rate the a list of items before and after a meeting to discuss their views, where the items to be rated are identified through a review of the literature.[6] In the implemented RAM the items to be rated initially were identified by K-LR in discussion with the co-design group as areas which needed consensus to guide the development of statistical methodology. The items which were then discussed and re-rated in the focus group were determined by those which had not reached consensus following analysis of the questionnaire using traditional RAM rules (Table 3.4) and ratification by the SAG. This modification was chosen to obtain a wider range of expert views to the traditional RAM. However, not every stakeholder group was well represented. Only six payers and industry partners completed the questionnaire. This is not seen as a major limitation as review panels include representatives from each stakeholder group. In addition, there was not a pre-focus group rating as all attendees had intended to be questionnaire respondents, so their pre-discussion views should have already been recorded. Another deviation from the traditional RAND was that the experts self-selected themselves to be invited to take part in the focus group in the initial questionnaire. This approach was chosen to reduce bias by not only those in the immediate research teams network being invited to take part.

Another limitation is the terminology used in the questionnaire. As it was directed at multiple stakeholder groups, the attempt to balance the terminology in the questionnaire made it too complex for some patients, but too simplistic for some statisticians. On reflection, this could have been resolved by having separate questionnaire for every stakeholder group rather than separate questionnaire pathways. In addition, at the suggestion of the co-design group, the ethnic groups used within the questionnaire were taken from the categorizations used by the Office of National Statistics.[14] However, it was commented that "*The ethnic groups are really odd*". Therefore, the questionnaire may have benefited from further testing. Nonetheless even with these changes or additional testing of the questionnaire, it is still likely that not every respondent would have been happy with every aspect of the questionnaire.

Considering technical limitations, hosting the questionnaire online meant that it was subject to the provider restrictions. This was a particular limitation for the format and the fact Cookies were not collected to stop duplicate responses. Given the length of the questionnaire it is unlikely anyone completed it twice and the alternative to require respondents to log in would have added an unnecessary recruitment barrier.

In terms of the focus group, only seven out of the intended 12 attended on the day despite 19 being invited. It may be that in the time between the closure of the questionnaire (September 2022) and the focus group (March 2023) invited respondents may have changed emails or forgotten about completing the survey. A few attendees also dropped out in the days prior to the focus group, which made it difficult to invite another in their place. In the future we would recommend inviting more than the required sample size to account for last minute dropouts. This may have also led to the imbalance of stakeholders within the focus group. Whilst every stakeholder group was represented, there were more statisticians (three) and patients (two) than payers and industry partners (one) and healthcare professionals (one). An alternative format would have been to offer multiple dates and times for focus groups so that more people could attend, but this has its implications in terms of time and cost. However, it is thought that hosting it online using Microsoft Teams was a strength, as attendees did not have to travel to input into the discussions.

The main strength of this study is the use of patient and public involvement (PPI) in conjunction with participatory research tools to determine what should be considered during the development of statistical methodology. This approach is not novel but is rare with only a few examples existing in the literature, such as the work by Flight using qualitative methods to identify stakeholder views on the use of health economics in the design and analysis of trials[15]. Given the impact methodological decisions can have on

the outcome it seems counterintuitive to not include the patients and the public perspective and there is a movement at present to improve their general involvement in the numerical aspects of trials.[16, 17] From this study, both the patient members on the co-design group and focus group attendees were enthusiastic to be included with an attendee commenting that they had "*enjoyed the session*" and another one noting that at the beginning of the focus group they wondered if they would be able to understand it, but they did. This feedback, along with the fact that more than a third of questionnaire respondents were patients and the public demonstrates that the patient and public stakeholder group are interested in being included in methodological and numerical discussions, and that it is down to the researcher to consider, with PPI representatives, which aspects are appropriate and make them accessible.

3.7 Conclusion

This work is an example of successfully implementing qualitative research in the process of developing statistical methods. By seeking consensus from stakeholder groups, a range of perspectives have been formally obtained that can be used as an evidence base to guide the development of a methodological solution to analyse OS when trial participants receive subsequent treatment lines during trial follow-up. If these findings are applied, we can be confident that any developed methodology is relevant and aligns to the interests of each stakeholder group.

3.8 Declarations

3.8.1 Ethics approval and consent to participate

Ethical approval has been obtained from the Yorkshire & The Humber – Bradford Leeds Research Ethics Committee (22/YH/0155). Respondents provided informed consent as part of the online questionnaire. Informed consent was required prior to starting the questionnaire. Respondents were able to withdraw their consent at any time.

3.8.2 Consent for publication

Consent for any quotes to be published from both the questionnaire and focus group was sought as part of the informed consent process.

3.8.3 Availability of data and material

Consent to share individual level data was not sought in this project. Therefore, only reasonable requests of aggregate data and anonymised quotes will be considered.

3.8.4 Competing Interests

The authors declare that they have no competing interests.

3.8.5 Funding

This project is funded by the NIHR Doctoral Fellowship Programme [NIHR301510]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also supported by Core Clinical Trials Unit Infrastructure from Cancer Research UK [C7852/A25447]. IRW was supported by the Medical Research Council [Programme MC_UU_00004/07].

3.8.6 Author contributions

Under supervision from DM, JKV-R, IRW and DAC, K-LR devised and conducted the qualitative analysis, and drafted the manuscript. PW helped to recruit to and was an active member of both the co-design and stakeholder group. He also facilitated the focus group. All authors read and approved the final manuscript.

3.8.7 Acknowledgements

We acknowledge the contributions of the co-design and stakeholder advisory group in designing and testing the questionnaire. We thank the respondents who completed the questionnaire and those who attended the focus group.

3.9 Chapter 3 References

1. FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry, https://www.fda.gov/media/71195/download (2018, accessed 6th September 2022).
2. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (2017, accessed 1st March 2023).
3. Fontrier AM, Visintin E and Kanavos P. Similarities and Differences in Health Technology Assessment Systems and Implications for Coverage Decisions: Evidence from 32 Countries. *Pharmacoecon Open* 2022; 6: 315-328. 20211129. DOI: 10.1007/s41669-021-00311-5.
4. Latimer NR and Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
5. Royle K-L, Meads D, Visser-Rogers JK, et al. How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review. *Trials* 2023; 24: 708.
6. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA appropriateness method user's manual*. 2001. Rand Corp Santa Monica CA.

7. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004; 6: e34. 20040929. DOI: 10.2196/jmir.6.3.e34.
8. Tong A, Sainsbury P and Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care* 2007; 19: 349-357. DOI: 10.1093/intqhc/mzm042.
9. SOLVE Study. SOLVE Summary Video, https://youtu.be/HKYbl8vcCaw (2022, accessed 8th February 2023).
10. SOLVE Study. What is the SOLVE Study?, https://youtu.be/dZsle9e-Npl (2022, accessed 8th February 2023).
11. SOLVE Study. How will the SOLVE study answer it's question?, https://youtu.be/AdfS1FeBhBs (2022, accessed 8th February 2023).
12. Hsieh H-F and Shannon SE. Three Approaches to Qualitative Content Analysis. *Qualitative health research* 2005; 15: 1277-1288. DOI: 10.1177/1049732305276687.
13. Kidwell KM and Almirall D. Sequential, Multiple Assignment, Randomized Trial Designs. *Jama* 2023; 329: 336-337. DOI: 10.1001/jama.2022.24324.
14. Office for National Statistics. Ethnic group, national identity and religion, https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion, accessed 17th October 2023).
15. Flight L, Julious S, Brennan A, et al. How can health economics be used in the design and analysis of adaptive clinical trials? A qualitative analysis. *Trials* 2020; 21: 252. DOI: 10.1186/s13063-020-4137-2.
16. Beatriz G, Camille P and Katie G. Patient and public involvement in numerical aspects of trials: a mixed methods theory-informed survey of trialists' current practices, barriers and facilitators. *BMJ Open* 2021; 11: e046977. DOI: 10.1136/bmjopen-2020-046977.
17. Goulao B, Bruhn H, Campbell M, et al. Patient and public involvement in numerical aspects of trials (PoINT): exploring patient and public partners experiences and identifying stakeholder priorities. *Trials* 2021; 22: 499. DOI: 10.1186/s13063-021-05451-x.

Chapter 4 What Would Happen if Everyone Received the Same Subsequent Therapy? A Hypothetical Estimand for Overall Survival in the Presence of Treatment Discontinuation

Kara-Louise Royle[1], David Meads[2], Jennifer K. Visser-Rogers[3], David A Cairns[1] and Ian R White[4]

1. Leeds Cancer Research UK Clinical Trials Unit, Leeds Institute of Clinical Trials Research, University of Leeds, UK
2. Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
3. Coronado Research, Kent, England
4. MRC Clinical Trials Unit at UCL, London, UK

Study 3	
Journal	Statistics in Medicine
Submission Status	Submitted January 2026
Reference	TBC

4.1 Abstract

Overall survival in cancer clinical trials is usually analysed using the treatment policy approach. This incorporates the effect of treatment switching, i.e. the uptake of subsequent anti-cancer therapy following discontinuation of the trial intervention, which may differ over time and by arm as the clinical landscape evolves. This approach could influence interpretation. Previous work has shown that the uptake of subsequent anti-cancer therapy is noted, but not accounted for in analysis, and that stakeholders are interested in considering how the effect of subsequent therapy can be quantified in the analysis of overall survival. We propose a hypothetical estimand which estimates the experimental trial intervention effect, if all participants received the same subsequent anti-cancer therapy following trial intervention discontinuation, where appropriate. We use an extensive simulation study to assess the ability of simple (intention-to-treat and per protocol) and advanced (two-stage estimation and inverse probability of censoring weighting) statistical methods to estimate the hypothetical estimand. The methods are then applied in a phase III, randomised, controlled multi-centre trial in renal cancer, to assess the practicality of the proposed approach. We show that the two-stage estimation methods are the most consistent across the simulation study and most appropriate to apply in practice. Accordingly, we recommend that researchers collect suitable information on subsequent anti-cancer therapy in their clinical trials, and apply two-stage estimation methodology as a sensitivity analysis for overall survival to fully comprehend the impact of subsequent anti-cancer therapy.

4.2 Key Words

Simulation study, overall survival, randomised controlled trial, cancer, subsequent treatment lines, treatment discontinuation, treatment switching

4.3 Introduction

Overall survival (OS) is a definitive endpoint for clinical effectiveness in cancer clinical trials. It is defined as time from randomisation to death from any cause. However, the analysis of OS can require trial participants to be followed up for a long period of time. In this time trial participants may discontinue their trial intervention and begin, potentially multiple, different subsequent anti-cancer therapies. These subsequent anti-cancer therapies may impact OS, and the choice may be impacted by the participant's response to the original trial intervention, especially if new subsequent therapies are made available over time. Therefore, the uptake of them may impact the analysis of OS. This is an example of an intercurrent event for OS in the estimand framework[1], and can impact of the interpretation of OS if trialists are unsure whether an effect, or lack of effect, between trial interventions are due to an ineffective experimental intervention, or the uptake of subsequent therapies.

This challenge in the interpretation of OS is well known[2] and is one reason for the use of surrogate endpoints, for example, progression-free survival (PFS), disease-free survival (DFS) or relapse-free survival (RFS), as the primary endpoint of clinical trials, and the relegation of OS to a secondary endpoint. Nevertheless, even as secondary endpoint, OS is still of interest and is used alongside other evidence in health technology appraisals (HTA) of experimental interventions.[3] Therefore, it is important that this intercurrent event is dealt with appropriately to ensure that incorrect conclusions are not made in appraisals due to the inappropriate application of statistical methodology.

A recent literature review on the analysis of OS[4] found that treatment discontinuation and the uptake of subsequent anti-cancer therapy is usually considered using the "Treatment Policy" approach to intercurrent events, i.e. it is considered to be part of the treatment strategy. The review also showed that the subsequent anti-cancer therapy that trial participants go on to receive is often descriptively summarised in trial reports. If the distribution of the different types of subsequent anti-cancer therapy that participants receive are similar across randomisation arms, researchers assume that there is little impact on their OS conclusions. If it is not similar, they generally accept it as a limitation and conduct no additional modelling. However, a stakeholder consultation on the topic concluded that whilst stakeholders are interested in the endpoint as a whole i.e. the treatment policy approach, they believe that the impact of

subsequent therapy should be considered during the analysis of OS. The stakeholders did not agree with the approaches assuming that no one stopped their trial treatment prior to death (a hypothetical approach), consider only those who only received their trial treatment prior to death (a principal stratum approach) or shorten OS to consider time on treatment instead (a composite / while on treatment approach). [5]

A potential approach to consider the impact of subsequent therapy, whilst taking into consideration the stakeholders' perspectives, is to consider an alternative hypothetical estimand. We define this as the treatment effect if everyone who did discontinue before death, went on to receive the same subsequent anti-cancer therapy. This is a plausible scenario which would no longer ignore subsequent anti-cancer therapy and would enable researchers to infer what would have happened to participants when discussing their results. This also extends existing work from the National Institute for Health and Care Excellence (NICE) Decision Support Unit which at outset, focused solely on treatment switching in terms of participants in the control arm subsequently receiving the experimental intervention.[6] This definition of treatment switching is not in the scope of the work presented here, as only one randomised group are considered taking up subsequent therapy. How the work presented here relates to the recent advancements in the definition of treatment switching are discussed in Section 4.7.

We sought to determine whether the proposed alternative estimand was a viable option. The objectives were (i) to evaluate whether the estimand could be estimated using existing statistical methods in a simulation study, and (ii) to evaluate the practicability of the estimand in a real data case study.

The article is structured as follows. In Section 4.4 the simulation study is described following the ADEMP framework.[7] The results of the simulation study are presented in Section 4.5. The methods are applied to a real data case study in Section 4.6. Finally in Section 4.7, the results are discussed and recommendations for researchers and areas for future research are considered.

4.4 Overview of Simulation Study

4.4.1 Aims

The aims of the simulation study were:

1. To explore whether the alternative estimand can be estimated with minimal bias, adequate precision and suitable coverage in a base scenario.

2. To identify the most stable method to estimate the alternative estimand i.e. the method which provided the most consistent results across a range of plausible and extreme scenarios.

4.4.2 Data-Generating Mechanism

4.4.2.1 Overview

A parametric data-generating mechanism (DGM) was implemented to simulate a dataset similar to a phase III cancer clinical trial. The data generating process is shown in Figure 4.1. Following the simulation of baseline variables, participants were randomised to one of two treatment arms. Follow-up variables were then simulated including discontinuation, subsequent treatment and survival. A detailed description of the process is given below.

4.4.2.2 Data Generating Process

The data was generated as follows. A full description of all the variables and parameters is provided in Appendix Table D.1.

1. Baseline: Three observed (bo_1, bo_2, bo_3) and two unobserved (bu_1, bu_2) baseline variables were generated as independent instances of standard distributions:
    a. $bo_1 \sim Bernoulli(bop_1)$
    b. $bo_2 \sim Bernoulli(bop_2)$
    c. $bo_3 \sim Binomial(2, bop_3)$
    d. $bu_1 \sim Bernoulli(bup_1)$
    e. $bu_2 \sim Normal(bum_2, bus_2^2)$
2. Randomisation: Stratified randomisation allocation was generated using randtreat[8, 9], stratified by bo_1, bo_2, bo_3.
3. Follow-up: Follow-up values of bo_2, bo_3 and bu_1, bu_2 at k equidistant timepoints were generated using either logistic or linear regression models as appropriate. The models included the baseline variables, randomisation allocation, with a diminishing effect, and the follow-up variables from all previous timepoints as covariates.
4. Discontinuation: Whether discontinuation occurred was generated at each follow-up timepoint using a logistic regression model which included the baseline variables, randomisation allocation and the follow-up variables at that timepoint as covariates.
5. Subsequent Treatment: In participants who discontinued, subsequent treatment (A or B) was generated using a logistic regression model. The model included

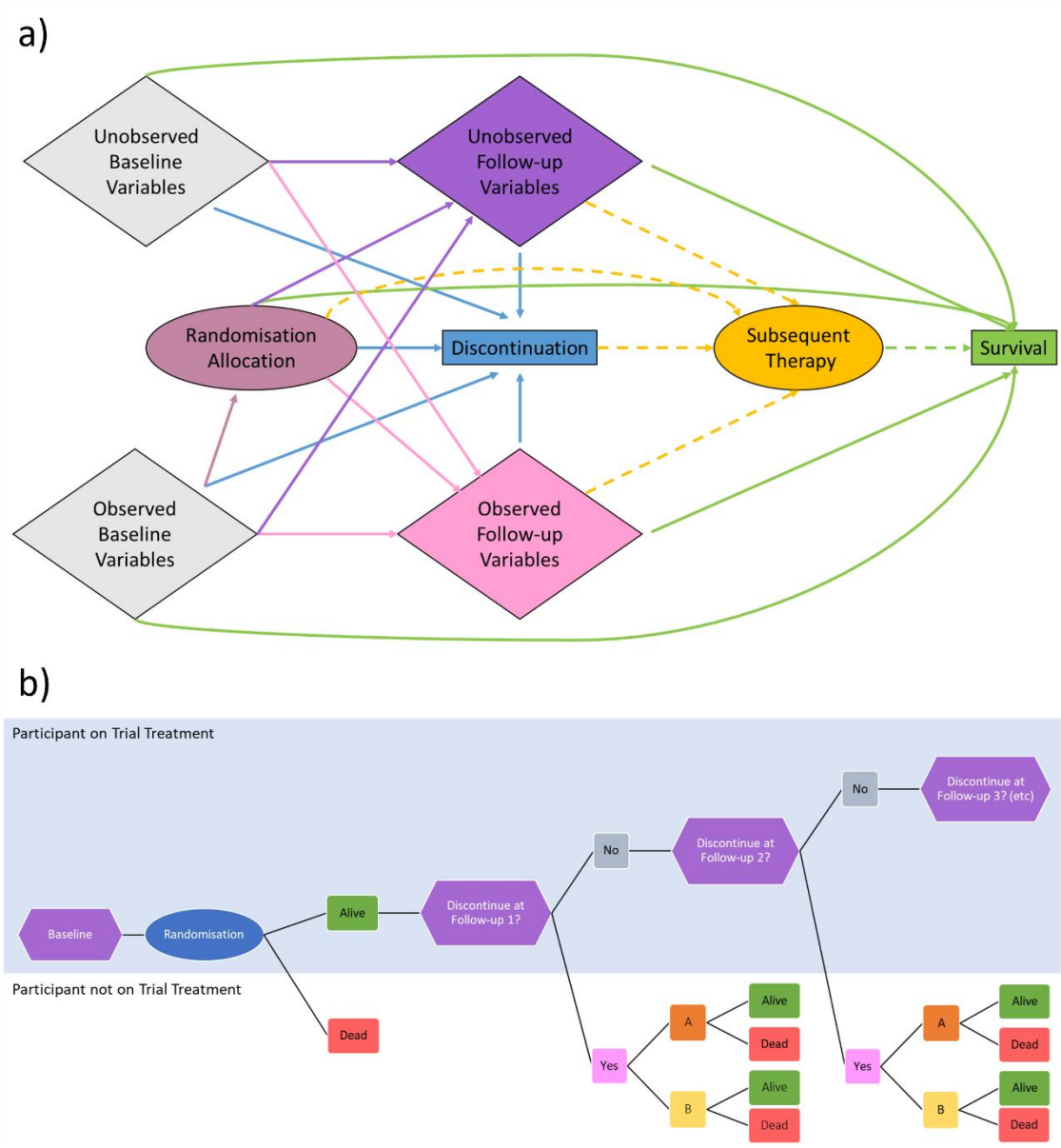
the follow-up variables at the timepoint the participant was simulated to discontinue as covariates.

6. Survival: Survival time was simulated up to a maximum time using survsim.[10] The baseline variables, randomisation allocation, follow-up variables and subsequent treatment were included as covariates in the log-hazard function. Indicator functions were included in the model to ensure that the follow-up variables and subsequent treatment only influenced the log-hazard function in the appropriate time intervals e.g. once the participant had begun subsequent treatment.
7. Reconciliation: If a participant was simulated to have died prior to the follow-up time, discontinuation was reset to 0 (i.e., did not discontinue) and subsequent anti-cancer therapy was reset to not applicable (i.e., did not receive a subsequent anti-cancer therapy).

4.4.2.3 Parameter Specification

The base scenario (Appendix Table D.1) represented a realistic trial scenario, based on a systematic literature review[3] where for trials with OS as a primary endpoint, median recruitment was 42 months, median follow-up 18 months, the median minimal clinically important difference equated to a hazard ratio (HR) of 0.725, and there was a median of three stratification factors. Note that 406 events were required for 90% power to detect a HR of 0.725 assuming a two-sided test at the 5% significance level. The different scenarios implemented for Aim 2 are summarised in Table 4.1.

Figure 4.1: Visual Representation of the Simulation Study



(a) DAG of DGM for Simulation Study
Colours indicated which variables are used to simulate other variables. Solid lines are completed for all participants, dashed lines indicate only simulated as required i.e. subsequent therapy only if participants discontinue. Diamonds indicate variables which are not influenced by other variables, ovals indicate variables which are influenced and influence other variables, rectangles are outcome variables which may impact other variables.
(b) Flow diagram of patient pathway
Potential pathway of participants through the simulation study, hexagons indicate study visits, squares and rectangles potential outcomes. The shaded area is during trial treatment / intervention.

Table 4.1: Simulation study scenario descriptions, rationale, and parameter changes from the base scenario

Scenario Name	Description	Rationale	Parameter	Base Value	New Value
BaseScenario	Moderate trial treatment, beneficial subsequent anti-cancer therapy.	To represent a simple clinical trial with drop out over four follow-up timepoints resulting in some subsequent anti-cancer therapy.	N/A	N/A	N/A
HarmTrialTrt	Harmful trial treatment, beneficial subsequent anti-cancer therapy	If the trial experimental treatment is harmful compared to the control treatment but the subsequent therapy is beneficial would the harmful effect of the experimental trial treatment be seen, or would it be masked by the effect of the subsequent therapy?	$\beta_{[S,trt]}$	0.725	1.25
NoEffectTrialTrt	No effect trial, beneficial subsequent anti-cancer therapy	If the trial experimental treatment has no effect compared to the control treatment but the subsequent therapy is beneficial would the no effect of the experimental trial treatment be seen, or would it be masked by the effect of the subsequent therapy?	$\beta_{[S,trt]}$	0.725	1
VeryBenTrialTrt	Really beneficial trial, beneficial subsequent anti-cancer therapy	If the trial experimental treatment is very beneficial compared to the control treatment but the subsequent therapy is also beneficial would the very beneficial effect of the experimental trial treatment be seen, or would it be masked by the effect of the subsequent therapy?	$\beta_{[S,trt]}$	0.725	0.5
ThreeMonthlyFup	3 monthly timepoints	To represent a clinical trial with a more plausible follow-up schedule.	k	4	19
			α_D	0.005	-2
EarlyDiscontinuation	The majority of participants who discontinue, discontinue as early as possible.	What is the impact when the effect of subsequent therapy is in play for longer than the trial treatment effect?	α_D	0.005	1.5, 0.01, -0.005, -1.5
LateDiscontinuation	The majority of participants who discontinue, discontinue as late as possible.	What is the impact when the effect of subsequent therapy is in play for less time than the trial treatment effect?	α_D	0.005	-3.5, -3, -2.5, 0.5

Scenario Name	Description	Rationale	Parameter	Base Value	New Value
MajorityExpTrtB	The majority of participants who discontinue on the experimental arm get treatment B	What is the impact when a lot of participants in one group are being excluded or censored?	$\beta_{[A,trt]}$	1.45	0.01
MajorityDiscontExp	Of those who discontinue the majority are in the experimental arm not the control arm	What is the impact when discontinuation is not balanced in favour of the experimental arm?	$\beta_{[D,trt]}$	0.65	4.65
			α_D	0.005	-1.995
MajorityDiscontA	70% of participants who discontinue receive treatment A	What is the impact when the majority of participants follow the treatment pathway the estimand is investigating?	$subtreatp$	0.3	0.7
VeryBenSubTrt	Very beneficial subsequent anti-cancer therapy	What is the impact when there is a very effective (slightly unrealistic) subsequent anti-cancer therapy	$\beta_{[S,A]}$	0.8	0.6

4.4.3 Estimand

The estimand of interest was the treatment effect for randomisation allocation if every participant that discontinued had subsequent anti-cancer therapy option A rather than B (Table 4.2).

Table 4.2: Estimand of Interest for the Simulation Study

Treatment Conditions	Control and Experimental Simulated Trial Treatment followed by subsequent treatment A if subsequent treatment was received
Population	All simulated individuals
Endpoint	Overall Survival: time from randomisation to death from any cause
Population Level Summaries	Hazard ratio of the experimental vs control trial treatments
Intercurrent Events	• Treatment discontinuation and uptake of subsequent anti-cancer therapy A: Treatment Policy approach • Treatment discontinuation and uptake of subsequent anti-cancer therapy B: Hypothetical approach: what would have happened if everyone starting subsequent anti-cancer therapy had received A rather than B

Since the estimand was not a parameter in the DGM described above, it was estimated using a numerical approach as in Williamson et al[11]. A second DGM was generated to create a trial of one million participants such that everyone who discontinued received subsequent anti-cancer therapy A. This dataset was then analysed using the intention-to-treat (ITT) approach, detailed in Section 4.4.4. The HR for randomisation allocation was defined as the true value of the estimand (true HR) to be used to assess the performance of the different methods within the simulation study.

4.4.4 Methods

Nine different methods, which are currently used or recommended for treatment switching[6], were implemented: ITT[12], Per-protocol[13] (PP) exclude (PPEx), PP censor (PPCens), inverse probability of censoring weights (IPCW) using both unstabilised (IPCWN) and stabilised weights (IPCWS)[14] , two-stage estimation (TSE) with and without re-censoring[15] using both the Weibull (TSEWNoR, TSEWR) and generalised gamma (TSEGNoR, TSEGR) accelerated failure time (AFT) models. In each method the substantive model was a Cox proportional hazards model[16] regressing randomisation allocation $(randtrt)$ on OS adjusted for the stratification factors of the trial (bo_1, bo_2, bo_3) (Equation 4.1, Appendix D.1.2). However, the preliminary steps differed between methods.

Equation 4.1: Cox Proportional Hazards Model

Let $h(t)$ represent the hazard of an individual dying at time t, then under the Cox proportional hazards model:

$$h(t) = \exp\left(\boldsymbol{\beta}_{[cox,b]}\boldsymbol{b}_o + \beta_{[cox,trt]}randtrt\right) h_0(t),$$

where $\boldsymbol{\beta}_{[cox,b]} = \left[\beta_{[cox,bo1]}, \beta_{[cox,bo2]}, \beta_{[cox,bo3]}\right]$ and $\beta_{[cox,trt]}$ are parameters to be estimated,
$\boldsymbol{b_o} = [bo_1, bo_2, bo_3]^T$ and $randtrt$ are variables and $h_0(t)$ is the baseline hazard function.

The ITT method made no changes to the population or simulated data and implemented the Cox model as described above. PPEx excluded all participants who discontinued at a follow-up timepoint and started subsequent anti-cancer therapy B. PPCens censored these participants at the point of discontinuation.

IPCW also censored all participants who discontinued at a follow-up timepoint and started subsequent anti-cancer therapy B. However, the Cox model was fitted using weights. For IPCWN the weights were derived using the probability of a participant not being censored at each follow-up timepoint. The weights were estimated by fitting a pooled logistic regression model in participants who discontinued at a follow-up timepoint (Equation 4.2). Participants were only included in the model if they discontinued in that interval as otherwise they could not have started one of the two subsequent treatments.

Equation 4.2: Logistic Regression Model to Predict the Probability of being Censored

Let $P_{discontinueB[k]}$ be the probability that a participant commenced subsequent anti-cancer therapy B in interval k, given they discontinued their trial intervention in interval k.

$$logit\left(P_{discontinueB[k]}\right) = \alpha + \boldsymbol{\beta}_{[IPCW,b]}\boldsymbol{b}_o + \beta_{[IPCW,trt]}randtrt + \boldsymbol{\beta}_{[IPCW,f]}\boldsymbol{f}_{ok} + \boldsymbol{\beta}_{[IPCW,time]}\boldsymbol{time}_k$$

where,
$\boldsymbol{\beta}_{[IPCW,b]} = \left[\beta_{[IPCW,bo1]}, \beta_{[IPCW,bo2]}, \beta_{[IPCW,bo3]}\right], \beta_{[IPCW,trt]}$, $\boldsymbol{\beta}_{IPCW_f} = \left[\beta_{[IPCW,fo2]}, \beta_{[IPCW,fo3]}\right]$ and $\boldsymbol{\beta}_{[IPCW,time]} = \left[\beta_{[IPCW,time]}, \beta_{[IPCW,time2]}, \beta_{[IPCW,time3]}\right]$ are to be estimated,
$\boldsymbol{b_o} = [bo_1, bo_2, bo_3]^T$ and $randtrt$ are variables, $\boldsymbol{f_{ok}} = [fo_2, fo_3, fo_3]^T$ are time varying covariates with different values in each time interval and $\boldsymbol{time}_k = \left[time_k, time_k^2, time_k^3\right]^T$ is the time interval k starts.

The fitted model was then used to predict the probability of a participant being censored within each interval and used to derive the denominator for the unstabilised weights (Equation 4.3). Note that due to the data structure, a participant could only be censored at the point they discontinued trial intervention, and this could only happen once. If a participant did not discontinue at the start of the interval the probability of being censored was set to zero. Therefore, by design, the probabilities were zero for intervals up to the point of discontinuation, changed for the interval in which the participant discontinued and then returned to zero afterwards. In terms of weights, this resulted in a constant weight of one for intervals prior to discontinuation and then a constant but different weight for intervals at the point of and post discontinuation.

Equation 4.3: Denominator for unstabilised weights

Let $P_{discontinueB[k]}$ be as per Equation 4.2 for $k = 1\ to\ K$ intervals. The denominator of the unstabilised weights for the k^{th} interval is given by:

$$\prod_{k=1}^{K}\left(1 - P_{discontinueB[k]}\right)$$

The IPCW method was also implemented using stabilized weights (IPCWS). This method calculated the weights using the same denominator as IPCWN. However, the numerator was obtained by fitting a pooled logistic regression model similar to Equation 4.2 in all intervals, omitting the baseline, follow-up, and treatment variables. The model was then used to predict the probability of a participant being censored within each interval, from which the product of the complements was calculated and used as the numerator for the stabilised weights. Note that unlike the denominator, the calculation for the numerator included all participants and no probabilities were set to zero.

The TSE method without re-censoring did not censor those who discontinued and received subsequent anti-cancer therapy B. Instead, the follow-up timepoint at which the participant stopped treatment was defined as a second baseline, and accelerated failure time (AFT) models were fitted to the remaining time in those who discontinued. The Weibull (TSEWNoR) and generalised gamma (TSEGNoR) AFT models were implemented by assuming the random error term followed an extreme value distribution and a log-gamma distribution, respectively (Equation 4.4). The Akaike information criterion (AIC) was used to assess goodness of fit between the two AFT models.

Equation 4.4: AFT Log-Linear Model

Let T_s be the simulated survival time, $T_1$the time from randomisation to the follow-up timepoint and T_2 the remaining time ($T_2 = T_s - T_1$).

Then the AFT log-linear parameterisation is given by:

$$\log(T_2) = \alpha + \boldsymbol{\beta}_{[TSE,b]}\boldsymbol{b}_o + \beta_{[TSE,Trt]}randtrt + \boldsymbol{\beta}_{[TSE,f]}\boldsymbol{f}_{od} + \beta_{[TSE,A]}\, subtrt + \sigma\epsilon$$

where

$\sigma, \alpha, \boldsymbol{\beta}_{[TSE,b]} = [\beta_{[TSE,bo1]}, \beta_{[TSE,bo2]}, \beta_{[TSE,bo3]}], \beta_{[TSE,Trt]},$
$\boldsymbol{\beta}_{[TSE,f]} = [\beta_{[TSE,fo2]}, \beta_{[TSE,fo3]}]$ and $\beta_{[TSE,A]}$ are to be estimated,
$\boldsymbol{b_o} = [bo_1, bo_2, bo_3]^T, randtrt, subtrt$ are variables, $\boldsymbol{f_{od}} = [fo_2, fo_3, fo_3]^T$ are the values of the follow-up variables at the point of discontinuation and ϵ is a random error term with an assumed distribution.

The estimate of the coefficient for subsequent anti-cancer therapy $\widehat{\beta_{[TSE,A]}}$, where subsequent anti-cancer therapy B was the reference level, was taken from the AFT model (Equation 4.4) and used to derive the counterfactual survival time (Equation 4.5). The Cox model was then fitted to the counterfactual times.

Equation 4.5: Counterfactual Survival Time within the TSE Method

Let T_s be the simulated survival time, $T_1$the time from randomisation to the follow-up timepoint and T_2 the remaining time ($T_2 = T_s - T_1$).

Then the counterfactual survival time T_{cf} is given by,

$$T_{cf} = \begin{cases} T_1 + \exp(\widehat{\beta_{[TSE,A]}})T_2, & \text{If received B} \\ T_s, & \text{Otherwise} \end{cases}$$

The final method, TSE with re-censoring (TSEWR, TSEGR), re-censored the counterfactuals derived in the previous TSE method (Equation 4.6) The Cox model was then fitted to the re-censored counterfactual times (T_{cf^*}) and events (E_*).

Equation 4.6: Re-censoring of the Counterfactual Survival Time

Let T_{cf} be the counterfactual survival time with time ratio $\widehat{\beta_{[TSE,A]}}$ estimated from the TSE method. Then if t_{max} is the maximum follow-up time, and *E* the event indicator which equals 1 if a participant experienced an event, the re-censored time T_{cf^*} and event E_* are given by:

$$T_{cf^*} = \min\left(T_{cf}, t_{max}\,,\, \widehat{\beta_{[TSE,A]}} \times t_{max}\right)$$

$$E_* \begin{cases} 0\,, & \text{If } T_{cf} > T_{cf^*} \\ E, & \text{Otherwise} \end{cases}$$

4.4.4.1 Assumptions

Both the IPCW and TSE methods assume no unmeasured confounding (NUC). However, the DGM included unobserved variables in the simulation of discontinuation and survival to represent unmeasured confounders (Figure 4.1, Appendix Table D.1). To investigate how unmeasured confounding influenced the results, all six methods (IPCWN, IPCWS, TSEWNoR, TSEGNoR, TSEWR, TSEGR) were repeated meeting the assumption of no unmeasured confounding (IPCWNNUC, IPCWSNUC, TSEWNUCNoR, TSEGNUCNoR, TSEWNUCR, TSENUCGR). In this investigation, the unobserved variables at both baseline and follow-up (Appendix Table D.1) were included in the derivation of weights for the IPCW methods, and the AFT models for the TSE methods.

The log-hazard survival function for the simulated survival times presented in Section 4.4.2 assumes that the treatment effect of the experimental treatment is constant, even after participants have discontinued their trial treatment. The impact the assumption has on the method performance was investigated through sensitivity analyses re-running all DGMs, except NoTrialTreatmentEffect, with a log-hazard survival function which negated this effect, by only including an experimental treatment effect in the log-hazard function prior to discontinuation.

4.4.5 Performance Measures

For each repetition of the simulation study, the number of participants included in the model and the number of events along with log HR, standard error (SE), and p-value for proportional hazards for randomisation allocation for each method was obtained. For the TSE methods this also included the AIC of the AFT models and the subsequent anti-cancer therapy effect ($\widehat{\beta_{[TSE,A]}}$). For the IPCW methods, summary statistics of the weights used in the model were also outputted.

The p-value for the proportional hazards test, conducted at the 5% significance level on the Schoenfeld residuals, was used to determine how often proportional hazards was violated and whether the Cox model was an appropriate substantive model. The AIC of the AFT model was used to decide which of the distributional forms (Gamma or Weibull) provided the best fit to the simulated data. The summary statistics of the weights were used to identify whether the pooled logistic regression model (Equation 4.2) was appropriate. The log HR and its associated SE were used to compare the different methods in terms of bias, coverage and power within each simulated dataset. An acceptable value of 0 for bias, 95% for coverage and 80% for power were pre-specified. The performance measures and the associated Monte Carlo (MC) SEs were calculated using the Stata command simsum.[17]

The estimates and performance measures of the simulated data were visualised using the Stata command siman.[18] Swarm and scatter plots were used to look for outliers across the replications. Bland-Altman plots were used to investigate which of the methods were the most similar to the ITT method. Lollipop plots were used to compare the average performance measure across the simulation study and its 95% MC confidence interval (CI) to the acceptable value. Finally, zip plots were used to visualise the coverage and determine whether this is due to bias or underestimation of SE (shown through an unsymmetrical plot).

4.4.6 Implementation of ADEMP

Prior to implementing the simulation study, a single dataset was generated under the specified DGM and summarised descriptively to ensure the data was as expected.[19] For the base scenario, to check for implementation errors, this included fitting a Cox model for the time from randomisation to the first follow-up timepoint including all observed and unobserved variables. Similarly, for the base scenario a small number of repetitions (ten) were run and analysed to confirm that each iteration was generating different data, the estimates were as expected, and it was confirmed that any repetition could be reproduced using the specified seed and RNG stream state.

In addition, the asymptotic performance of the methods as the sample size tends to infinity was investigated visually. A simulated dataset of one million participants was generated under the base scenario. The Kaplan-Meier survivor function for the ITT, TSE (Weibull, generalised gamma, with and without re-censoring) and the IPCW (stabilised and unstabilised weights) were calculated and plotted on a Kaplan-Meier curve along with the survival function for the true HR. The smoothed hazard estimates over time were also plotted.

Following confirmation that everything was as expected, 1000 repetitions were run for all scenarios. Note that whilst 1000 repetitions were pre-specified, this was planned to be increased as required following review of the MCSE.

All simulation study results can be replicated using the code on GitHub (https://github.com/KaraLouiseRoyle/SOLVE_Simulation). Statistical analysis was conducted in STATA (18, StataCorp LLC, College Station, TX) and SAS (9.4, SAS Institute, Cary, NC).

4.5 Simulation Study Results

All preliminary investigations concluded that the DGM was implemented as expected, and at least asymptotically the advanced methods corrected the observed bias in the ITT method (Appendix D.2.3).

Following the implementation of the base scenario it was concluded that all methods should be considered in the full simulation study and that 1000 repetitions were adequate.

4.5.1 Individual Method Implementation and Fit

Both IPCW and TSE methods required additional steps before the substantive model could be fitted. These are considered here.

There were no outliers in the weights for the IPCW methods within each scenario across the 1000 repetitions (Appendix Figure D.17) suggesting the (pooled) logistic regression model was appropriate.

Considering AFT model fit, the TSE method using the generalised gamma model (TSEGN, TSEGR) failed to converge on a number of repetitions (Appendix Table D.6). This is more noticeable in the LateDiscontinuation scenario when there is less information available for AFT model fitting. However, if model selection had been implemented, the Weibull model would have been selected for the majority of repetitions across all scenarios. Despite the non-convergence of some of the repetitions, the mean subsequent anti-cancer therapy effect (A vs B) derived from the AFT models was similar to the reciprocal of the inputted value implying that the model was appropriate (Appendix Table D.7).

4.5.2 Descriptive Cross-Method Performance

The number of events for each method for each scenario show that, as expected, in the scenarios which vary the impact of the treatment effect and/or methods where more censoring or exclusion of participants occurred, less events are observed or used in the

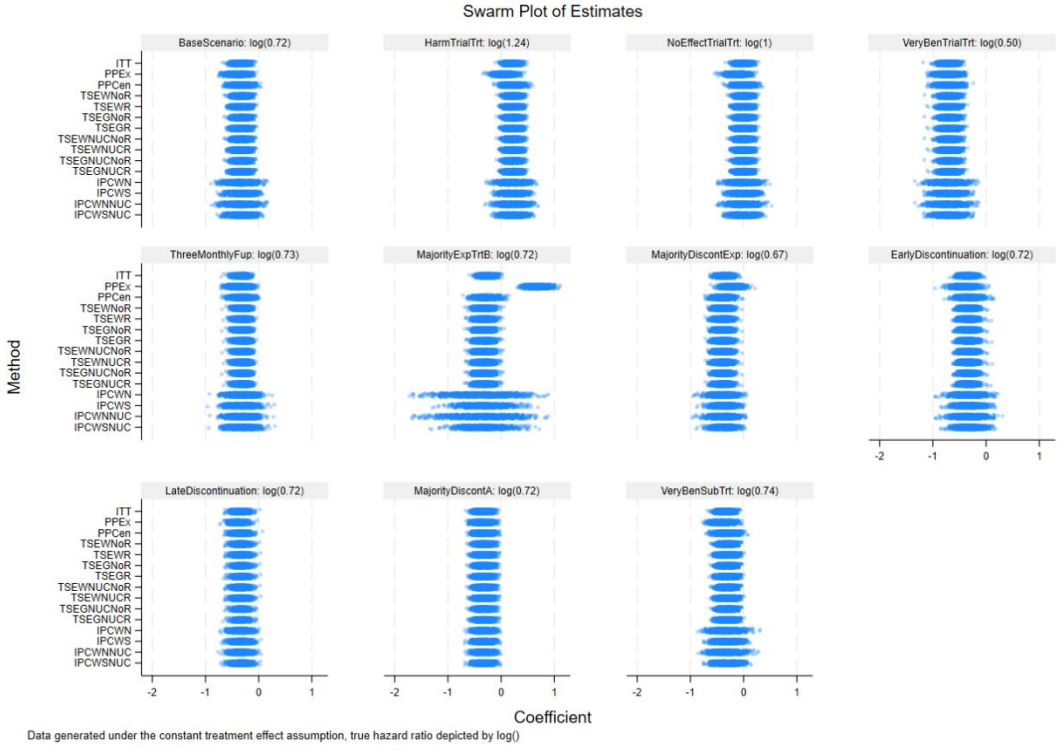
analysis (Table 4.3). In terms of proportional hazards, with the exception of PPEx and the IPCW methods, randomisation allocation did not violate proportional hazards above the 5% nominal level (Appendix Table D.8).

Table 4.3: Mean (SD) of the number of events used in the Cox Model averaged across the repetitions.

	ITT	PP Ex	PP Cen	TSE WNoR	TSE WR	TSE GNoR	TSE GR	TSEW NUCNoR	TSEW NUCR	TSEG NUCNoR	TSEG NUCR	IPCW N	IPCW S	IPCW NNUC	IPCW SNUC
BaseScenario	463 (12)	271 (13)	271 (13)	463 (12)	435 (21)	463 (12)	436 (21)	463 (12)	436 (21)	463 (12)	437 (20)	436 (26)	248 (16)	438 (26)	248 (16)
HarmTrialTrt	530 (11)	326 (13)	326 (13)	530 (11)	504 (19)	530 (11)	504 (19)	530 (11)	504 (19)	530 (11)	505 (19)	505 (24)	309 (17)	506 (25)	309 (17)
NoEffectTrialTrt	504 (12)	303 (13)	303 (13)	504 (12)	476 (20)	504 (12)	477 (20)	504 (12)	477 (20)	504 (11)	478 (20)	478 (25)	282 (17)	479 (26)	283 (17)
VeryBenTrialTrt	418 (12)	240 (13)	240 (13)	418 (12)	391 (21)	418 (13)	392 (20)	418 (12)	392 (21)	418 (12)	393 (20)	392 (26)	217 (15)	393 (26)	217 (15)
ThreeMonthlyFup	483 (12)	265 (13)	265 (13)	483 (12)	454 (21)	483 (12)	454 (21)	483 (12)	455 (21)	483 (12)	455 (21)	456 (26)	139 (12)	457 (27)	139 (12)
MajorityExpTrtB	463 (12)	279 (13)	279 (13)	463 (12)	435 (22)	463 (12)	436 (21)	463 (12)	435 (22)	463 (12)	436 (21)	439 (111)	248 (44)	442 (119)	249 (48)
MajorityDiscontExp	467 (12)	338 (13)	338 (13)	467 (12)	445 (18)	467 (12)	446 (18)	467 (12)	446 (18)	467 (12)	446 (17)	448 (22)	306 (16)	448 (23)	306 (17)
EarlyDiscontinuation	462 (13)	234 (13)	234 (13)	462 (13)	432 (22)	462 (13)	433 (22)	462 (13)	432 (22)	462 (13)	434 (22)	432 (28)	205 (16)	433 (28)	205 (16)
LateDiscontinuation	472 (12)	417 (13)	417 (13)	472 (12)	455 (21)	472 (12)	455 (24)	472 (12)	455 (22)	472 (12)	454 (24)	463 (19)	424 (17)	463 (20)	424 (17)
MajorityDiscontA	449 (13)	367 (13)	367 (13)	449 (13)	436 (15)	449 (12)	436 (14)	449 (13)	436 (15)	449 (12)	437 (15)	438 (15)	341 (14)	438 (15)	341 (14)
VeryBenSubTrt	450 (12)	258 (13)	258 (13)	450 (12)	391 (22)	450 (12)	393 (22)	450 (12)	392 (22)	450 (12)	393 (22)	392 (26)	231 (15)	393 (27)	231 (16)

Figure 4.2 shows all estimates for the 15 methods across the 11 scenarios. There were no obvious outliers in the estimates within each scenario across the 1000 repetitions (Figure 4.2, Appendix Figure D.18). However, in all scenarios the variation in the point estimates in the IPCW methods is larger than all other methods. This may be due to the fact the IPCW methods rely on the estimation of weights which are dependent on the data.

Figure 4.2: Swarm Plot of Estimates across the Scenarios Generated Under the Constant Treatment Effect Assumption

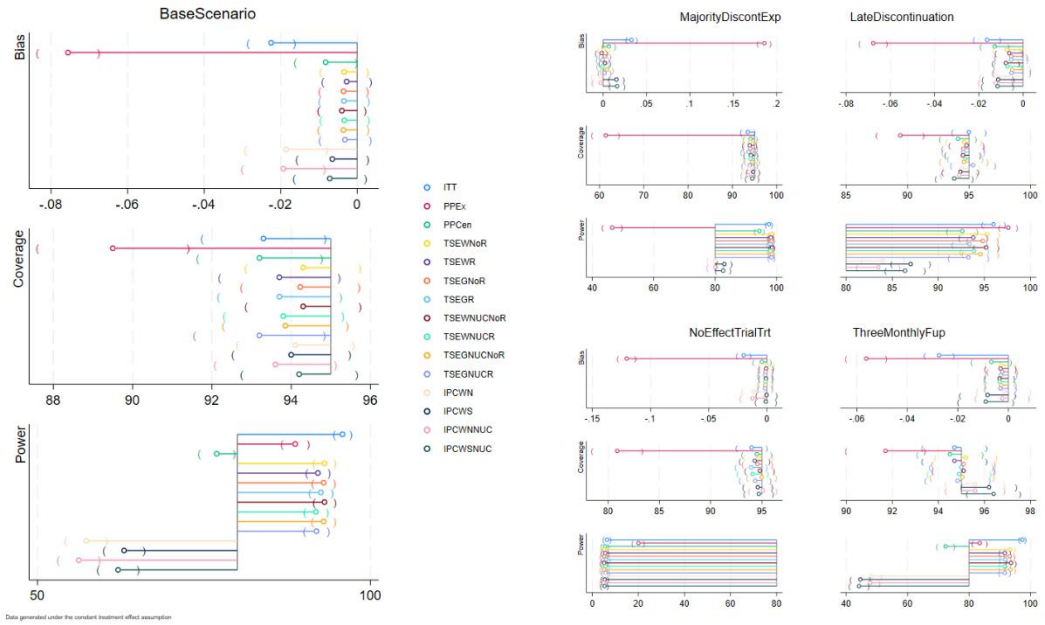


Overall, the TSE method estimates are the most similar to the ITT method estimates across all scenarios (Appendix Figure D.19). For the other methods, the similarities seem to depend on the amount of censoring. This can be seen through how all methods are similar to ITT in MajorityDiscontA and LateDiscontinuation when the intercurrent event has the least impact.

4.5.3 Overall Cross-Method Performance

The impact of the intercurrent event influences the performance of the methods in terms of bias, coverage and power (Figure 4.3, Appendix Figure D.20).

Figure 4.3: Lollipop Plots of Performance Measures from Simulations Generated under the Constant Treatment Effect Assumption



Across the lollipop plots, PPEx is shown to be biased and have poor coverage which gets worse as the impact of the intercurrent event increases or there are imbalances between the arms. This can be seen clearly when comparing the performance in the BaseScenario to MajorityExpTrtB or MajorityDiscontExp. This can also be seen in the zip plots (Appendix Figure D.21). Considering the other methods, ITT is biased across all scenarios with the magnitude dependent on the scenario. For example, comparing the BaseScenario to MajorityDiscontExp shows that imbalances in arm cause the bias to switch direction. The performance of PPCens fluctuates depending on the impact of the intercurrent event, but not to the same magnitude as PPEx. The IPCW methods fluctuate the most across the scenarios and have the widest zip plots which, as expected, is marginally corrected for when stabilised weights are used. This can be seen clearly in ThreeMonthlyFup where the IPCW unstabilised methods have the worst bias and coverage. The TSE methods are the most stable of the methods with very similar performance across the different scenarios. Only in the zip plots could the Weibull TSE methods be deemed better due to the generalised gamma TSE methods failing to converge in a number of scenarios, most notably in LateDiscontinuation. Finally, considering the different types of scenarios, some changes in performance were expected. For example, NoEffectTrialTrt has no power, which makes sense since there was not a significant effect to find.

4.5.4 Sensitivity Analysis

The previous sections focused on the performance of the IPCW and TSE methods assuming NUC. However, unmeasured confounders were present in the DGM (Figure 4.1). Adherence to the NUC assumption in the methods IPCWNNUC, IPCWSNUC, TSEWNUCNoR, TSEGNUCNoR, TSEWNUCR and TSENUCGR only noticeably impacted AFT model convergence, where notable changes in the proportion of repetitions not converging occur in MajorityDiscontExp and LateDiscontinuation (Appendix Table D.6). This may be due to the fact there is less data available in these scenarios for AFT model fitting, but more variables in the NUC methods. However, this limited impact may be by design as the inputted parameter effects of the unmeasured confounders were small.

Changing the DGM to relax the constant treatment effect assumption (Appendix D.4) had a noticeable impact on whether the proportional hazards assumption was violated (Appendix Table D.12) and method performance. In general, the lollipop plots showed a similar pattern to when this assumption was in place. However, if the method performed poorly with the assumption in place, it performed worse without it. This can primarily be seen through the magnitude of the biases which increased across the two approaches. In addition, the power was worse across all scenarios. However, as the estimand of interest is closer to one this was to be expected. The differences between the approaches are also clear in the zip plots (Appendix Figure D.27) where all methods apart from the TSE methods perform poorly in at least one scenario (ITT: HarmTrialTrt, MajorityExpTrtB, MajorityDiscontExp, PP: HarmTrialTrt, ThreeMonthlyFup, MajorityExpTrtB, MajorityDiscontExp, VeryBenTrialTrt, IPCW: VeryBenTrialTrt, VeryBenSubTrt, MajorityExpTrtB, EarlyDiscontinuation).

4.5.5 Summary

Across all of the scenarios, the performance of the TSE methods are the least impacted by the changes in the parameter values. In addition, the pre-specified acceptable values of bias (0) coverage (95%) and, when appropriate, power (80%) are within the MC CI across all scenarios. Therefore, provided the goodness of fit of the AFT model is considered, the simulation study results suggest that the TSE method should be applied in practice.

4.6 Case Study

4.6.1 Aim

The aim of the case study was to determine whether the estimand can be estimated in practice.

4.6.2 Trial Overview and Estimand

The STAR trial (ISRCTN:06473203) was a randomised multi-stage phase II/III trial of standard TKI (tyrosine kinase inhibitor) first-line therapy (sunitinib or pazopanib). Between 13th Jan 2012 and 12th Sept 2017 participants at 60 UK sites, with locally advanced and/or metastatic renal cancer, were randomised (1:1) to either a drug-free interval strategy (DFIS) or a conventional continuation strategy (CCS). Intervention was until RECIST[20] defined disease progression on treatment, or clinician or patient withdrawal, at which point subsequent anti-cancer therapy was determined by the treating clinician. Note that this could have included continuation of the first-line TKI. Randomisation was stratified by Memorial Sloan Kettering Cancer Center prognostic group risk factor[21] (favourable, intermediate, or poor), sex (male/female), trial site, age (<60 years / ≥60 years), disease status (metastatic / advanced), TKI (sunitinib/pazopanib), and previous nephrectomy (yes/no). The trial included OS as a co-primary endpoint and aimed to demonstrate non-inferiority between the two randomisation arms with a non-inferiority margin of 7.5% equating to a hazard ratio (HR) of 0.812.[22]

The trial collected patient and disease characteristics at baseline and six-weekly whilst participants received their trial intervention (CCS or DFIS). Following discontinuation of the trial intervention, subsequent anti-cancer therapy along with alive and progression status was collected at six months post discontinuation and annually thereafter.

At final analysis the ITT and PP analysis reported a HR (95% CI) of 0.97 (0.83,1.12) and 0.94 (0.80,1.09) respectively. Therefore, the trial met non-inferiority in the ITT but not the PP analysis. However, it should be noted the analysis was underpowered due to the observed number of events (ITT: 678, PP: 648) being less than the required 720 for 80% power.[23]

At trial conception, there were limited effective subsequent therapies. However, during the course of the trial, second line treatments were approved across the devolved nations of the United Kingdom by NICE and the Scottish Medicines Consortium. This included the immunotherapy nivolumab as a monotherapy.[24, 25] The final analysis publication inferred, in the discussion, that the advancements of subsequent therapies

impacted the results.[23] However, the monograph explicitly reported the information.[26] The authors stated that because a similar proportion of participants in each arm received immunotherapy post-trial intervention discontinuation, no additional analysis looking at the impact of subsequent anti-cancer therapies was conducted. Therefore, the research question for the case study was: "*What is the OS effect of DFIS compared to CCS if everyone who discontinued their trial intervention prior to death and continued to be treated received second-line immunotherapy?*". The associated estimand is presented in Table 4.4.

Table 4.4: Estimand for STAR Real Data Case Study

Treatment Conditions	CCS or DFIS followed by second line immunotherapy if second line was received.
Population	All individuals with a confirmed diagnosis of renal cell carcinoma and meet the STAR trial eligibility criteria
Endpoint	Overall Survival: time from randomisation to death from any cause
Population Level Summaries	Hazard ratio of CCS vs DFIS
Intercurrent Events	• Trial intervention (CCS or DFIS) discontinuation and started second-line with immunotherapy: Treatment Policy approach • Trial intervention (CCS or DFIS) discontinuation and started second-line but did not have immunotherapy: Hypothetical approach: what would have happened if everyone starting second-line therapy received immunotherapy • Trial intervention (CCS or DFIS) discontinuation and continuation of trial first-line TKI without starting a second-line prior to death: Treatment Policy approach • Trial intervention (CCS or DFIS) discontinuation and continuation of trial first-line TKI and started a second-line with immunotherapy prior to death: Treatment Policy approach • Trial intervention (CCS or DFIS) discontinuation and continuation of trial first-line TKI and started a second-line not with immunotherapy prior to death: Hypothetical approach: what would have happened if everyone starting second-line therapy received immunotherapy • Trial intervention (CCS or DFIS) discontinuation and palliative care: Treatment Policy approach • Any Third Line or Later: Hypothetical approach: what would have happened had no one received a third-line of therapy

4.6.3 Methods

OS was defined as time from randomisation to death from any cause. Participants who received a third line of therapy were censored at the end of their second line of therapy and any information lost summarised. Second line therapy was defined as the receipt

of any systemic anti-cancer therapy post-trial intervention discontinuation except for the continuation of the allocated trial TKI. The type of therapy and time to second line therapy were summarised.

The same methods were implemented as described in Section 4.4.4. Therefore, only STAR specific implementation requirements are described here. The PP methods excluded or censored participants who at the start of second-line therapy, began a second-line therapy without immunotherapy. The IPCW methods also censored the same participants. For IPCW the data was transformed into six-weekly intervals, to reflect the six-weekly on study visits. For each interval binary time dependent variables were derived indicating whether in the interval the participant; started second-line therapy, received second-line therapy without immunotherapy or died. The logistic regression model regressed receiving second-line therapy without immunotherapy on randomisation allocation, the stratification factors of the trial (excluding centre), ECOG performance status (<2, ≥2), calcium, haemoglobin, neutrophil and platelet counts compared to the upper limits of normal measured at the start of the interval, and the linear, quadratic and cubic forms of the interval start time. Note that participants were only included in the model if they started second-line therapy in that interval. The AFT models for the TSE method included randomisation allocation, the stratification factors of the trial (excluding centre), ECOG performance status (<2, ≥2), calcium, haemoglobin, neutrophil and platelet count compared to the upper limits of normal measured at the participants last on study visit, receipt of second-line immunotherapy (Yes, No) and time on trial treatment. Only, the best fitting AFT model in terms of AIC was carried forward.

For each method, the number of participants, OS events and censored observations were obtained along with the HR, 95% CI and p-value for randomisation allocation from a Cox proportional hazards model which was adjusted for the stratification factors for the trial excluding trial site. The p-value for the test of the proportional hazards assumption calculated from the supremum test for the randomisation allocation coefficient was also obtained. In addition, a Kaplan-Meier curve stratified by randomisation allocation (CCS, DFIS) was estimated for each method. For each method the results and conclusion were compared to the original analysis conducted by the authors. Missing data was addressed using the last observation carried forward approach.

4.6.4 Results

The STAR trial included 919 participants in the ITT population. At the end of trial follow-up 95.4% had discontinued trial intervention and 61.4% had recorded receiving systemic anti-cancer therapy during trial follow-up (Appendix Table D.13).

Once continuation of trial TKI had been considered, 492 (53.5%) started second-line therapy during trial follow-up. Median (IQR) time to second-line therapy from discontinuation was 53 (23, 167) days (CCS: 61 (26,197), DFIS: 46 (22, 121)) the most common second-line treatment type was a TKI (Table 4.5). In total, 115 (12.5%) participants received second-line immunotherapy.

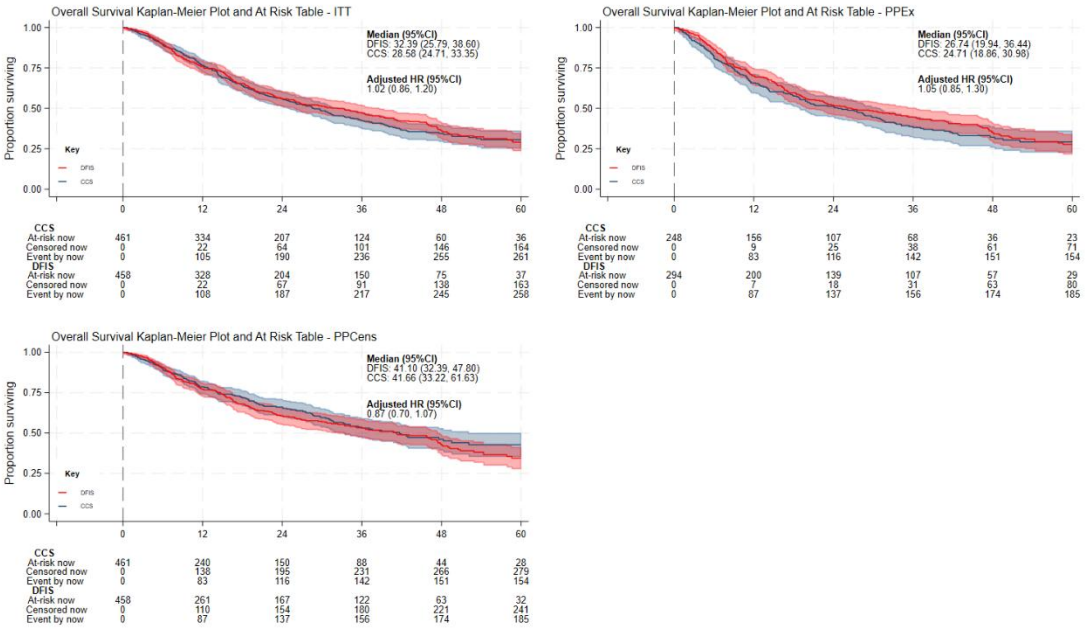
Table 4.5: STAR trial: Second Line Treatment Type

	CCS (n=461)	**DFIS (n=458)**	**Total (n=919)**
First-line TKI	47 (10.2%)	51 (11.1%)	98 (10.7%)
TKI	141 (30.6%)	93 (20.3%)	234 (25.5%)
Immunotherapy (alone or in combination)	56 (12.1%)	59 (12.9%)	115 (12.5%)
mTor (alone or in combination)	22 (4.8%)	20 (4.4%)	42 (4.6%)
Other cancer treatment or investigational product	3 (0.7%)	0 (0.0%)	3 (0.3%)
N/A	192 (41.6%)	235 (51.3%)	427 (46.5%)

In total 203 (22.1%) participants (CCS: 107 (23.2%), DFIS: 96 (21.0%)) were censored at the end of their second line of therapy resulting in a reduction of 144 OS events (CCS: 67, DFIS: 77). Median (IQR) follow-up for the ITT analysis was 44.8 (29.7,62.7) months (CCS: 42.3 (27.5, 60.1), DFIS: 46.5 (30.5,63.2)).

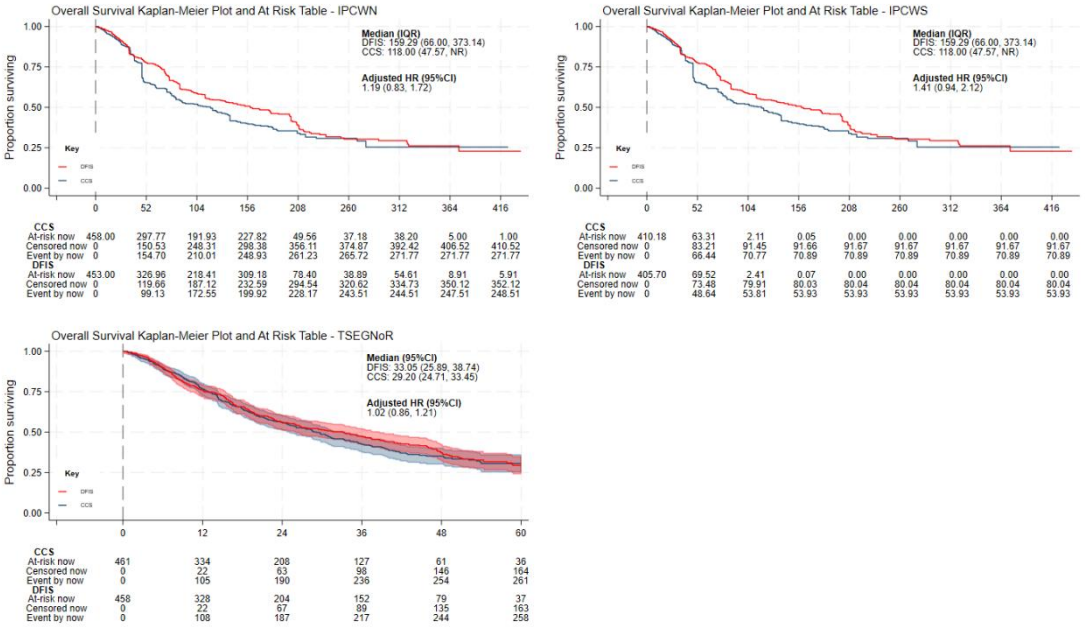
The Kaplan-Meier survival curves for the simple and advanced methods are shown in Figure 4.4 and Figure 4.5 respectively.

Figure 4.4: Kaplan-Meier Survival Curves for the Simple Methods



ITT: Intention-to-Treat; PPEx: Per-protocol Exclude; PPCens: Per-protocol Censor

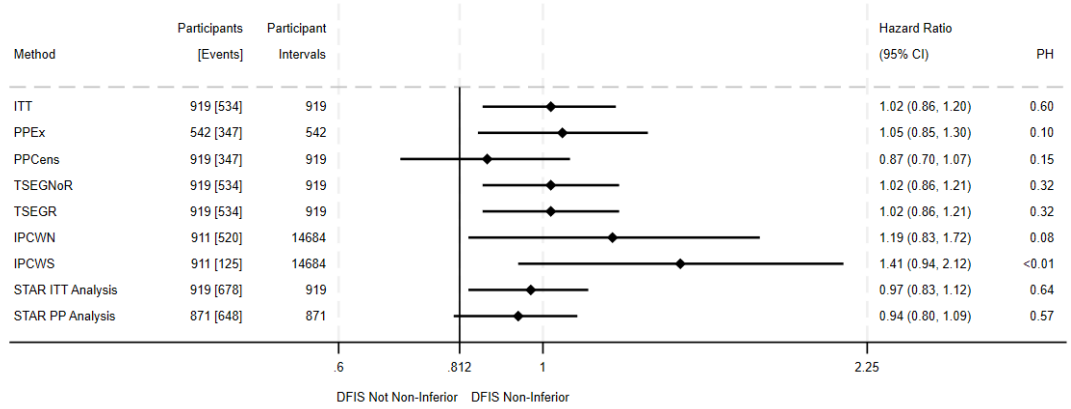
Figure 4.5: Kaplan-Meier Survival Curves for the Advanced Methods



TSEGNoR: Two-stage estimation generalised gamma without recensoring; IPCWN: inverse probability of censoring weights with unstabilised weights; IPCWs: inverse probability of censoring weights with stabilised weights.

Figure 4.6 displays the corresponding summary information including the Cox proportional hazard model results. The median (IQR) of the unstabilised and stabilised weights for the IPCW methods were 1.00 (1.00,1.00) and 0.125 (0.00142, 0.498) respectively. The generalised gamma AFT model was selected for the TSE method, with an AIC of 1022.7 compared to 1043.9 for the Weibull model. The time ratio was 1.13 indicating that second-line immunotherapy increases the time to death compared to no second-line immunotherapy. Note the results were not different when re-censoring was incorporated into the TSE method as only one censored event was recensored.

Figure 4.6: STAR trial: Summary of Model Results



ITT: Intention-to-Treat; PPEx: Per-protocol Exclude; PPCens: Per-protocol Censor; TSEGNoR: Two-stage estimation generalised gamma without re-censoring; TSEGR: Two-stage estimation generalised gamma with recensoring; IPCWN: inverse probability of censoring weights with unstabilised weights; IPCWs: inverse probability of censoring weights with stabilised weights; PH: p-value for the proportional hazards test. Non-inferiority is concluded if the lower-bound of the confidence interval does not cross the solid line. Note that eight individuals were excluded from IPCW due to switching to second line in the first 6-week interval. The hazard ratio presented is CCS vs DFIS from the Cox Proportional Hazards model adjusted for the stratification factors of the trial. Std error and 95% CI are the associated standard error and 95% confidence interval. Prop H is the p-value for the supremum test for proportional hazards.

Only PPCens, had a lower bound of the 95% two-sided CI below 0.812 and would have therefore reached a different conclusion to the original STAR ITT analysis which concluded that DFIS was non-inferior to CCS in terms of OS (Figure 4.6). However, PPEx unnecessarily removed participants from the analysis, ITT did not make any adjustments for not-receiving immunotherapy in second line, and IPCWs did not perform well in terms of the proportional hazard assumption. In the case study the results from the TSE methods are very similar to the ITT method. This is like the simulation study. Therefore, like the simulation study, the TSE method is considered to be the most appropriate method of analysis.

4.7 Discussion

The overarching aim of this work was to determine whether an alternative hypothetical estimand for OS could be estimated both theoretically through a simulation study and practically using a case study. The estimand assumed a scenario where all participants, where appropriate, receive the same one of two subsequent anti-cancer therapies. For the case study this was translated into all participants who commenced second-line therapy being assumed to have received immunotherapy.

The work was implemented in stages. The first aim of the simulation study was "*To explore whether the alternative estimand can be estimated with minimal bias, adequate precision and suitable coverage in a base scenario*." The base scenario represented a plausible clinical trial dataset and considered four types of methods (ITT, per-protocol, TSE and IPCW). On the review of the base scenario results it was concluded that whilst some methods performed worse than others in specific performance measures, a fuller picture of their performance would be needed across different scenarios before making a conclusion. Therefore, all methods were taken forward to the full simulation study in which key parameters were varied across eleven different scenarios to consider alternative plausible and extreme scenarios, with and without a constant treatment effect assumption. This equated to 315 combinations.

There were concerns around the bias of the ITT method in the base scenario as an aim of the ITT method is to protect against post-randomisation events.[27] Investigations showed that the ITT method underestimated the survival compared to the DGM (Appendix Figure D.12). This was in part by design of the DGM, where the experimental arm was less likely to discontinue, and the control arm were less likely to have the uplift of subsequent anti-cancer therapy which may imply a bigger difference between the two arms than exists. This difference was exaggerated further by the constant trial treatment effect assumption. This was investigated by relaxing the constant trial treatment effect assumption and repeating the simulation study. Overall, there was little difference between the performance of the methods with and without the assumption. The ITT method remained biased without the assumption. However, a new issue related to proportional hazards was introduced. Given the motivation behind relaxing this assumption was to investigate ITT bias, which was still present, no further investigations were conducted. Future work could investigate alternative analysis models which allow for non-proportional hazards.

The second aim of the simulation study was "*To identify the most stable method to estimate the alternative estimand i.e. the method which provided the most consistent results across a range of plausible and extreme scenarios.*" This simulation study

provides evidence why the more naïve methods (ITT and per-protocol) should not be used to address this intercurrent event. In addition, the more advanced method of IPCW also fails to perform consistently well across the different scenarios. A limitation of the IPCW method is the need to estimate weights and the estimation and stabilisation of weights is a somewhat subjective process as a decision must be made on covariates to use in the censoring model. The same could be said for the TSE methods, where a decision must be made on the functional form and covariates to include in the AFT model. However, the same covariates were considered in the censoring model and the AFT model for second baseline, and the TSE models performed the most consistently across all the scenarios. Therefore, from the simulation study it is TSE methodology that we would recommend should be implemented in the situation where trial participants discontinue their trial intervention and begin a subsequent anti-cancer therapy.

The recommendation to use the TSE method in this context was supported by the case study. However, it is important to recognise the decisions and assumptions which must be put in place to apply TSE in practice. The first decision is what subsequent therapy to consider in the estimand. In STAR there was a natural choice due to the introduction of immunotherapy during the course of the trial. However, for other trials there may not be. We would recommend considering the treatment which is closest to current practice. The second decision was which variables to include in the AFT model to make the assumption of no unmeasured confounding plausible. For STAR these decisions were made post-hoc in collaboration with a clinical expert and consideration of the literature and available data.[28] However, going forward it is recommended that researchers collaborate with other stakeholders when making decisions and where possible consider them at trial outset, and ongoing in line with any changes to the treatment landscape.

There are some limitations to STAR being used as the case study. Whilst STAR did collect subsequent anti-cancer therapy, it did not collect prognostic factors at the point of initiation of a new therapy which may have been more appropriate for inclusion in the AFT model. In addition, STAR followed a risk proportionate approach to data management which prioritised cleaning the data required for the primary analysis which was defined in the statistical analysis plan. Going forward we would recommend that trial teams not only collect this data but also perform basic checks e.g. sequential dates on the information during the course of the trial. This is in line with recent recommendations for trial teams to collect subsequent therapies[29] which our previous work found was not consistently collected.[5] We also recognise that STAR provides one

example. Therefore, future work will consider additional examples including the impact of sequential trial participation.

From the implemented methods, there is potential for the TSE method to be extended by incorporating external information around the effect of subsequent anti-cancer therapy A compared to subsequent anti-cancer therapy B into the switching model. This could either take a frequentist approach using network meta-analysis techniques or a Bayesian approach considering informative and uninformative priors. These approaches will be discussed with the various stakeholder groups as to whether this is an appropriate direction for future research.

Another direction for future research could be extending the DGM further to consider a third line of therapy. This limitation in the case study resulted in 203 participants being censored at the end of their second line anti-cancer therapy and the loss of 114 OS events. Once multiple individual discontinuation events are being considered there is the potential to introduce recurrent event methodology.[30] This again will be explored as an option for future research with different stakeholder groups in the next stage of this work.

There are additional limitations to, and assumptions made, in this work which could be further investigated. In the simulation study, unlike the case study, the SEs of the TSE and IPCW methods were not estimated using robust SEs. This was a practical choice due to computational resource required but means that the SEs presented are underestimates which could impact method performance in terms of coverage and power. In addition, the effect of baseline covariates, including treatment, on the value of follow-up covariates was simulated to have a diminishing effect over time. This was considered to be plausible but could be investigated in future work. Furthermore, whilst the impact of no unmeasured confounding was explored, the inputted effect of the unmeasured confounders was small and this may have biased the results. This aligns with findings from research into the specific scenario of treatment switching where methods reliant on the no unmeasured confounding assumption were suggested to perform adequately, only if no important independent predictors were excluded.[6] Further scenarios could be implemented in the future to investigate the breaking point for this assumption. Other potential extensions include considering more than two subsequent anti-cancer therapies; changing the effect of the follow-up variables over time; considering other types of censoring other than administrative censoring; changing the magnitude of the other covariate effects and adding a time-lag between participants discontinuing and commencing their subsequent therapy.

Since the design of this simulation study, an updated technical support document on treatment switching has been published with additional methods and recommendations for considering treatment switching in trials.[31] The work presented here complements the update as treatment switching is no longer defined as only participants in the control arm switching to the experimental arm, and our recommendations for trial teams considering the impact of subsequent treatments upfront, so that appropriate data is collected for sensitivity analysis do align. This work also provides simulation study evidence for this new extended definition of treatment switching which previous simulation studies did not. Furthermore it provides simulation evidence for the work conducted by Ouwens et al[32] who considered participants receiving subsequent immunotherapy in a real data case study. Future work could extend the simulation study considered here to compare the TSE method performance to the recent methodological advances. For example, TSE with g-estimation.

This simulation study and real data case study provides a potential answer to the previously unanswered question; how should OS be analysed when trial participants receive subsequent therapy post-trial? The hypothetical question of interest: "*What is the experimental trial treatment effect on OS compared to the control treatment, if all participants who discontinued prior to death went on to receive the same treatment?"* is a starting point which has the potential to be extended now that we have evidence that TSE is an appropriate method of estimation. As the simulation study was designed following a stakeholder consultation, the next phase of this research will aim to close the feedback loop, agree collaboratively what future research in this area should investigate, and provide guidance around data collection and make recommendation to researchers to apply the TSE method as a sensitivity analysis of OS in cancer clinical trials where participants receive subsequent therapy post-trial discontinuation.

4.8 Declarations

4.8.1 Ethics approval and consent to participate

Ethical approval has been obtained for this project from the Yorkshire & The Humber – Bradford Leeds Research Ethics Committee (22/YH/0155). Ethical approval for the STAR trial was obtained from Liverpool Central Research Ethics Committee (11/NW/0246).

4.8.2 Consent for publication

Not applicable.

4.8.3 Availability of data and material

All code for the simulation study is provided in Github (https://github.com/KaraLouiseRoyle/SOLVE_Simulation). Data from the case study supporting this work are available on reasonable request. All requests will be reviewed by relevant stakeholders, based on the principles of a controlled access approach. Requests to access data should be made to CTRU-DataAccess@leeds.ac.uk in the first instance.

4.8.4 Competing Interests

The authors declare that they have no competing interests.

4.8.5 Funding

This project is funded by the NIHR Doctoral Fellowship Programme [NIHR301510]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also supported by Core Clinical Trials Unit Infrastructure from Cancer Research UK [C7852/A25447]. IRW was supported by the Medical Research Council [Programme MC_UU_00004/07].

4.8.6 Author contributions

Under supervision from DM, JKV-R, DAC and IRW, K-LR devised and conducted the quantitative analysis, and drafted the manuscript. All authors read and approved the final manuscript.

4.8.7 Acknowledgements

We acknowledge the contributions of the co-design and stakeholder advisory group and participants in earlier work packages which enabled this work to take place.

4.9 Chapter 4 References

1. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (2017, accessed 1st March 2023).
2. FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry, https://www.fda.gov/media/71195/download (2018, accessed 6th September 2022).
3. Fontrier AM, Visintin E and Kanavos P. Similarities and Differences in Health Technology Assessment Systems and Implications for Coverage Decisions: Evidence from 32 Countries. *Pharmacoecon Open* 2022; 6: 315-328. 20211129. DOI: 10.1007/s41669-021-00311-5.
4. Royle K-L, Meads D, Visser-Rogers JK, et al. How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review. *Trials* 2023; 24: 708.

5. Royle KL, Wheatstone P, Meads D, et al. How should overall survival be analysed in randomised clinical trials in cancer if participants receive subsequent treatment lines? A stakeholder consultation. *Trials* 2025; 26: 434. DOI: https://dx.doi.org/10.1186/s13063-025-09148-3.
6. Latimer NR and Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
7. Morris TP, White IR and Crowther MJ. Using simulation studies to evaluate statistical methods. 2019; 38: 2074-2102. DOI: 10.1002/sim.8086.
8. Carril A. RANDTREAT: Stata module to randomly assign treatments uneven treatments and deal with misfits. 2017.
9. Carril A. Dealing with misfits in random treatment assignment. *The Stata Journal* 2017; 17: 652-667.
10. Crowther MJ and Lambert PC. Simulating biologically plausible complex survival data. *Statistics in Medicine* 2013; 32: 4118-4134. DOI: https://doi.org/10.1002/sim.5823.
11. Williamson EJ, Forbes A and White IR. Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Stat Med* 2014; 33: 721-737. 20130930. DOI: 10.1002/sim.5991.
12. Hollis S and White I. Intention-To-Treat (ITT). In: Everitt B and Palmer CR (eds) *Encyclopaedic companion to medical statistics*. Second edition ed. Oxford: Wiley, 2011.
13. Hollis S and White I. Per-Protocol. In: Everitt B and Palmer CR (eds) *Encyclopaedic companion to medical statistics*. Second edition ed. Oxford: Wiley, 2011, pp.329.
14. Robins JM and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; 56: 779-788. DOI: 10.1111/j.0006-341x.2000.00779.x.
15. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting for treatment switching in randomised controlled trials - A simulation study and a simplified two-stage method. *Stat Methods Med Res* 2017; 26: 724-751. 20141121. DOI: 10.1177/0962280214557578.
16. Collett D. *Modelling survival data in medical research*. Chapman and Hall/CRC, 2023.
17. White IR. simsum: Analyses of simulation studies including Monte Carlo error. *Stata Journal* 2010; 10: 369-385.
18. E Marley, I R White and T Morris. Suite of Stata programs for the analysis of simulation studies, https://github.com/UCL/siman, Downloaded 11 September 2025].
19. White IR, Pham TM, Quartagno M and Morris TP. How to check a simulation study. *International Journal of Epidemiology* 2023; 53. DOI: 10.1093/ije/dyad134.
20. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228-247. DOI: 10.1016/j.ejca.2008.10.026.
21. Motzer RJ, Mazumdar M, Bacik J, et al. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol* 1999; 17: 2530-2530.
22. Collinson FJ, Gregory WM, McCabe C, et al. The STAR trial protocol: a randomised multi-stage phase II/III study of Sunitinib comparing temporary cessation with allowing continuation, at the time of maximal radiological response, in the first-line treatment of locally advanced/metastatic Renal Cancer. *BMC Cancer* 2012; 12: 598. DOI: 10.1186/1471-2407-12-598.
23. Brown JE, Royle K-L, Gregory W, et al. Temporary treatment cessation versus continuation of first-line tyrosine kinase inhibitor in patients with advanced clear cell renal cell carcinoma (STAR): an open-label, non-inferiority, randomised, controlled, phase 2/3 trial. *The Lancet Oncology* 2023; 24: 213-227. DOI: 10.1016/S1470-2045(22)00793-8.

24. NICE. Nivolumab for previously treated advanced renal cell carcinoma, https://www.nice.org.uk/guidance/ta417 (2016, accessed 8th Jan 2025).
25. Scottish Medicines Consortium. nivolumab (Opdivo), https://scottishmedicines.org.uk/medicines-advice/nivolumab-opdivo-resubmission-118816/ (2017, accessed 28th January 2025).
26. Collinson F, Royle K-L, Swain J, et al. Temporary treatment cessation compared with continuation of tyrosine kinase inhibitors for adults with renal cancer: the STAR non-inferiority RCT. 2024; 28: 45. DOI: 10.3310/JWTR4127.
27. Gupta SK. Intention-to-treat concept: A review. *Perspect Clin Res* 2011; 2: 109-112. DOI: 10.4103/2229-3485.83221.
28. Tannir NM, Pal SK and Atkins MB. Second-Line Treatment Landscape for Renal Cell Carcinoma: A Comprehensive Review. *The Oncologist* 2018; 23: 540-555. DOI: 10.1634/theoncologist.2017-0534.
29. Lee D and Torres GM. Randomized controlled trial reporting guidelines should be updated to include information on subsequent treatments. *Cancer* 2025; 131: e35922. DOI: 10.1002/cncr.35922.
30. Charles-Nelson A, Katsahian S and Schramm C. How to analyze and interpret recurrent events data in the presence of a terminal event: An application on readmission after colorectal cancer surgery. *Statistics in Medicine* 2019; 38: 3476-3502.
31. Bell Gorrod H, Latimer NR and Abrams KR. NICE DSU Technical Support Document 24: Adjusting survival time estimates in the presence of treatment switching: An update to TSD 16, Available from https://www.sheffield.ac.uk/nice-dsu (2024).
32. Ouwens M, Darilay A, Zhang Y, et al. Assessing the Influence of Subsequent Immunotherapy on Overall Survival in Patients with Unresectable Stage III Non–Small Cell Lung Cancer from the PACIFIC Study. *Current Therapeutic Research* 2021; 95: 100640. DOI: https://doi.org/10.1016/j.curtheres.2021.100640.

Chapter 5 Discussion

5.1 Thesis Overview

The overall aim of this thesis was to "*develop new, or extend existing, methodology which assesses overall survival (OS) whilst accounting for subsequent lines of anti-cancer treatment in randomised clinical trials.*" Chapter 2 identified current research practice in terms of what is published in the literature through a systematic literature review. It found that subsequent lines of anti-cancer treatment are generally discussed but often only descriptively analysed, if analysed at all. It was hypothesized following the literature review that this was either a reporting bias or educational gap. Chapter 3 investigated this further within a stakeholder consultation, by asking statisticians and other data analysts their current practices and knowledge around the existing methods for treatment switching. The majority of those identifying as statisticians and other data analysts had not implemented or heard of the methods, identifying an educational gap. The stakeholder consultation was used to inform the development of the estimand investigated within the simulation study in Chapter 4. It also garnered the opinions of all stakeholders (patients and the public, statisticians and other data analysts, healthcare professionals, payers and industry partners) on topics which are relevant to the development of statistical methodology such as how results are presented, the use of statistical assumptions and data collection in clinical trials when participants have stopped their trial intervention. This use of participation and the associated PPI is rare in a PhD focused on statistical methodology and this work provides an example from other researchers to learn from. The simulation study investigated the hypothetical estimand which estimates: the experimental trial intervention effect compared to the control, if all participants received the same subsequent anti-cancer treatment following trial intervention discontinuation, where appropriate. The simulation study then determined whether existing statistical methods could be used to estimate the estimand. It found that the two-stage estimation (TSE) method performed adequately across the different scenarios and was the most appropriate when applied in a real data case study. The simulation study also highlighted how the naïve methods (ITT and PP) do not perform well in this scenario

In summary, the preceding chapters have met the overall aim of the thesis by considering the existing literature and stakeholder opinions to identify an appropriate estimand in the context of treatment discontinuation and the uptake of subsequent anti-cancer therapy, which can be estimated using the TSE method. The remaining sections of this chapter consider:

- The contribution of this thesis to the literature, including the implications of recent developments, in more detail.
- Recommendations for future cancer clinical trials including extensions to other endpoints and clinical areas outside of OS and cancer.
- Reflections on the limitations of the methods and scope of the thesis and how they could be addressed through future research.

5.2 Developments in and Contributions to the Existing Literature and Infrastructure

5.2.1 Treatment Switching and Uptake of New Statistical Methodology

At PhD outset the accepted definition of treatment switching reported in technical support documentation (TSD)[1] was:

"Treatment switching can occur when patients in the control group of a clinical trial are allowed to switch on to the experimental treatment at some point during follow-up."

This limited definition had been the sole focus of methodological developments in the area to date. The work presented in this thesis aimed to investigate developing this definition to consider participants switching on to any subsequent anti-cancer treatment following trial treatment discontinuation after progression, intolerability or choice.

Initial investigations through the literature review presented in Chapter 2 and stakeholder consultation in Chapter 3 concluded that an expanded definition of treatment switching was of interest to stakeholders and that researchers were not consistently investigating it or attempting to model its impact. The work presented in Chapter 4 aimed to begin to address this by applying existing methods for treatment switching to the problem using an alternative hypothetical estimand for OS. The systematic and robust methodology implemented in the thesis complement recent advances in the literature in this area.

Recently, in April 2024, the accepted definition of treatment switching reported in updated TSD[2] has been extended to:

"Treatment switching describes a situation where participants in a randomised controlled trial switch on to a treatment that they were not randomised to receive."

As well as extending the definition, the updated TSD summarised recent methodological developments in the area through a literature review.[2] The rank preserving structural failure time model (RPSFTM) was not included in the simulation

study due to concerns with the ability to estimate two parameters. This decision was based on previous work[3], over whether the method would converge and have sufficient power. Bell Gorrod et al discuss that it is unclear whether the recent attempt by Xu et al[4] provides a feasible solution to the issue. They also highlighted work by Li et al[5] who proposed an enhanced RPSFTM. This method essentially has two stages in order to successfully estimate two parameters. This, whilst promising, was explicitly designed to consider switching from the control arm alone, so is not relevant to the estimand considered in this thesis. As a result of these findings, and the fact the review did not highlight any proposed updates to IPCW methodology, it seems reasonable that the decisions and conclusions made around the RPSFTM and IPCW methods in this thesis remain valid despite updates in the literature.

However, Bell Gorrod et al's discussion of the work by Ouwens et al[6] and methodological extensions to the TSE method recommended in Chapter 4, require further consideration. Ouwens et al investigated a similar estimand using the point of treatment switch as the secondary baseline and applied both TSE and RPSFTM to remove the effect of subsequent immunotherapy. They concluded, similar to the real data example presented in Chapter 4, that as their results were similar to the ITT analysis, treatment with the immunotherapy durvalumab was beneficial for patients with non-small cell lung cancer irrespective of subsequent immunotherapy. However, they only applied their methodology to a real data case study. They did not support the application of the methods in a new scenario with simulation evidence. The simulation study conducted in Chapter 4 therefore provides independent evidence for the use of TSE in a comparable context. However, it may be beneficial to assess the performance of TSE with g-estimation, to alleviate some concerns highlighted by Bell Gorrod around the potential for time-dependent confounding.[7] The move to the extended TSE may also mean alternative secondary baselines, such as the point of treatment discontinuation, could be considered in this context. However, the potential extension does come with disadvantages including data required and computational complexity that should be considered. It is important to note that the Ouwens et al work was not identified during the literature review described in Chapter 2. This limitation is considered in Section 5.5.

The alternative secondary baseline was discussed more recently in a proposal for new methods for TSE in treatment switching.[8] The authors explored alternative hypothetical estimands considering participants in the control arm receiving immunotherapy as their first subsequent treatment, where immunotherapy was the experimental trial intervention. The different approaches and considerations to applying TSE in practice included, as indicated above, the definition of the alternative secondary baseline,

incorporating propensity scoring in the AFT model rather than covariate adjustment and removing the AFT model entirely to use a quantile matching approach. They concluded that their approach to considering different estimands helps to quantify the effect of treatment switching on the trial's overall conclusion from the ITT analysis. They also discussed the benefits of considering the methods in further work as they had not exhausted all possible avenues. The work presented in Chapter 4 provides further evidence that the effect of subsequent treatment can be quantified. It also provides more robust evidence for the use TSE in this setting through the inclusion of the simulation study for at least the methods which use AFT models. However, the authors focus on statistical significance alone to determine covariate adjustment in their models, and I recommend that future researchers also incorporate stakeholder engagement (e.g. clinical and patient) rather than just statistical significance to determine in the model covariates.

The updated TSD also identified a number of alternative methods proposed to adjust for treatment switching which do not fall into extensions of RPSFTM, IPCW or TSE. These included: random forest based prediction[4], regression imputation[9], semi-competing risks[10], semi-parametric copula-based approach[11] and a decision analytic modelling approach[12]. The latter of the two are not relevant extensions to this thesis; the semi-parametric copula-based approach does not derive counterfactuals and the decision analytic modelling approach does not include individual patient data. However, the remaining three methods provide interesting potential alternatives which require investigation before they can be recommended or discounted. This is considered as part of the future work in Section 5.6.

Outside of methodology developments, the updated TSD also reviewed the uptake and reporting of treatment switching methods within NICE technology appraisals published between 1st January 2020 and 20th April 2022. Only a quarter of the reviewed cancer appraisals mentioned switching or crossover (10/40, 25%) and how the methods were implemented (7/10 (70%) implemented adjustment methods) was not transparent, with the majority of information around modelling choices e.g. variable selection and consideration of assumptions not included in publicly published material. These findings resulted in the development of reporting guidelines for treatment switching adjustment methods, which whilst developed for technology appraisals, could be used to guide the reporting of similar analyses in clinical trial publications which are recommended as a result of this thesis (See Section 5.3). It is important to note these guidelines update the previously published recommendations from Sullivan et al[13] to include TSE which are particularly relevant to the findings in this thesis. Comparing the recommendations to the reporting of the adjustment methods in Chapter 4, in addition

to the considerations around the secondary baseline discussed above, D6 "*Describe which switches do not represent standard treatment pathways in the NHS in England and Wales*" would also need amending for the hypothetical estimand considered within the thesis. This is because the recommendations in the TSD have been written from the perspective of HTA decision making, which generally does not recommend adjusting for treatment switching when the switch represents standard treatment pathways. However, this thesis considers the perspective of the research team, investigating the impact of subsequent treatments as a whole. Interestingly, the real data case study presented in Chapter 4 does provide an example of when adjusting for a standard treatment pathway may be beneficial. This consideration of different estimands for different perspectives reiterates the second recommendation from the updated TSD:

"*We recommend that when planning trials and pre-specifying analysis, trial sponsors should consider questions of interest not only for regulatory agencies, but also for HTA agencies, and especially the implications that this can have for data collection. Recommendations made in the ICH E9 addendum around trial planning, pre-specification of analyses, and reporting of results, should be adopted*"

However, the uptake and understanding of the ICH E9 addendum has been questioned by a recent review on the topic. In total 81 out of a possible 122 (66.4%) protocols submitted to the United Kingdom's Health Research Authority between January 2011 to December 2020 attempted to define their primary estimand.[14] The review authors concluded that, whilst estimands are being more widely used, more training and resources are required to support researchers in correctly implementing the guidance. Interestingly, the authors did review how the intercurrent event of treatment switching was being addressed and whilst three protocols mentioned treatment switching, the two which said it was going to be dealt with hypothetically did not specify their hypothetical scenario. Whilst it is not clear which definition of treatment switching was being used by the protocols in question, I believe the estimand considered in the real data case study in Chapter 4 meets the review requirements and can be used as an example for future researchers when implementing the recommendations in this thesis (Section 5.3). More recently, a survey has aimed to understand the uptake of estimands across the UKCRC CTU network[15] Whilst this did not consider the quality of the estimand, of the 46 CTUs who completed the survey, 91% had used them in the last three years and 59% included them in standard operating procedures and guidelines. The preliminary results of the survey suggest, like the review, that the uptake of estimands is improving. However, there is more work to be done to make them part of day-to-day practice. Further work in this area is ongoing and the cross-

reporting of recommendations can only increase researchers awareness and have a positive impact on both research and reporting practices.

Following the recent developments in the literature, a recommendation paper on treatment switching was recently published.[16] The authors considered in detail the methods available for adjusting for treatment switching with the aim to provide a comprehensive summary for researchers. I welcome the publication of this overview to highlight the issue which this thesis has been investigating in a streamlined way for researchers to consider in their trials. Considering the authors recommendations listed in Table 2 and Table 3 of the article, I agree that, when implemented, researchers need to clearly report their decisions and assumptions. However, I question whether this article is one step ahead of the current landscape. From the literature review in Chapter 2, a separate independent review[17], a recent commentary[18] and recommendation 2 from TSD 24 quoted above, the barrier at present is the actual collection of subsequent treatment data in the first place. Section 5.2.2 discusses this area in more detail. That being said, it will be interesting to see if the publication of this guidance and TSD 24 supports the implementation and reporting of the methods in years to come. It is also important to note the work presented in Chapter 4 adds reproducible worked examples of the different methods which the authors implied was lacking.

Given the recent changes in the technology assessment and pharmaceutical industry guidance, the literature review searches detailed in Chapter 2 were rerun on the 11th November 2025. In total 159 papers were identified, after screening 37 (23%) papers remained and were read in full (Table 5.1, Appendix Table E.1).

Table 5.1: Screening Results from Literature Search Update

Reason for Exclusion	N	Percent
Review / Meta-Analysis / Pooled Analysis	47	39%
Protocol Paper	27	22%
Not Randomised / Real-World Data / Pilot Study	20	16%
Not Investigating Treatment (e.g. Screening or Surgical Practices)	9	7%
Not Reporting OS	8	7%
Cost-Effectiveness	7	6%
Surrogate Endpoints / Outcomes	2	2%
Full Text Not Available	1	1%
Not Cancer	1	1%
Total	122	100%

Twenty-one of the 37 papers (57%) mentioned subsequent treatment lines in some form. Note this included crossover by design and sequential trials. This is a similar percentage overall to what was reported in Chapter 2 (60%). Six of the 21 papers (29%), performed some analysis to consider subsequent treatment lines. Adam et al[19],

who were investigating a combination therapy of liver transplantation and chemotherapy compared to chemotherapy alone for patients with liver cancer, utilised the old definition of treatment switching to assess the impact of participants allocated to chemotherapy alone undertaking a subsequent liver transplantation. In their analysis they considered the naïve approach of excluding the participants who underwent subsequent transplantation in a per-protocol analysis. Parker et al implemented a different approach by analysing the time to subsequent treatment to investigate the issue in two different prostate cancer trials.[20, 21]They concluded in the first that as the time was similar in both arms there was no concern. In the second they concluded that, whilst time to salvage androgen deprivation therapy was increased, as there was no improvement in metastases-free survival for their short-course intervention, it was unlikely that OS was improved. Interestingly, the more recent trial RADICALS-HD[21] considered the old definition of treatment switching to “non-protocol androgen deprivation therapy” and the older trial STAMPEDE[20] considered the new definition of any second-line treatments. In my opinion it would have been more appropriate if the authors analysed OS using RPSFTM, IPCW or TSE rather than infer the impact from other endpoints. Nonetheless, three papers identified by the literature review did implement either RPSFTM or IPCW. [22-24] Two papers considered crossover by design in their trial using RPSFTM [22, 24]. Rao et al also applied IPCW in this scenario.[22] However, Yong et al [23] considered RPSFTM and IPCW to investigate participants going on to receive subsequent anti-CD38 antibodies in a multiple myeloma trial. They conclude that the impact of subsequent anti-CD38 antibody is the reason why there is a difference in progression-free survival observed in their trial but not OS. Of all the papers included in the review, in my opinion their summary of subsequent treatment lines (Figure 5.1) is an example of best practice as it clearly shows: the proportion of participants receiving subsequent treatments; the number of subsequent treatments and the types of subsequent treatments. Note that none of the papers in this review implemented TSE.

Figure 5.1: Example of Reporting Subsequent Treatment Lines from IKEMA[1]

Supplemental table 2: Summary of subsequent antimyeloma treatments

	Isa-Kd (n=179)	**Kd (n=123)**
Any subsequent antimyeloma treatment, n (%)	89 (49·7)	81 (65·9)
Number of subsequent regimens, median (range)	2·0 (1–8)	2·0 (1–13)
Number of subsequent regimens, n (%)		
1	44 (49·4)	25 (30·9)
2	19 (21·3)	30 (37·0)
≥3	26 (29·2)	26 (32·1)
Subsequent antimyeloma treatments, n (%)		
Corticosteroids	80 (89·9)	72 (88·9)
Immunomodulators	76 (85·4)	66 (81·5)
Alkylating agents	44 (49·4)	45 (55·6)
Proteasome inhibitors	42 (47·2)	33 (40·7)
Anti-CD38 agents	26 (29·2)	51 (63·0)
Daratumumab	26 (29·2)	44 (54·3)
Isatuximab	1 (1·1)	11 (13·6)
Other anti-CD38 agents	0	2 (2·5)
Anti-BCMA	18 (20·2)	5 (6·2)
Belantamab mafodotin	14 (15·7)	2 (2·5)
Elranatamab	3 (3·4)	1 (1·2)
Teclistamab	3 (3·4)	2 (2·5)
Further transplant	12 (13·5)	14 (17·3)
Anthracyclins	9 (10·1)	14 (17·3)
Topoisomerase inhibitors	7 (7·9)	6 (7·4)
Platinum compound	7 (7·9)	4 (4·9)
CAR T cell	2 (2·2)	4 (4·9)
Other	12 (13·5)	4 (4·9)
Selinexor	6 (6·7)	3 (3·7)

Applying the results of this thesis to the papers, 13 (62%) of the papers that mentioned subsequent treatments had potential applications for TSE considering either the old definition of treatment switching or as considered in this thesis. Of these, six have been discussed above as they performed some analysis. Considering the remaining seven in turn.

Izumi et al[25] report on the impact of enzalutamide versus abiraterone plus prednisolone before chemotherapy in prostate cancer. They provide clear descriptive summaries of the subsequent treatments participants went on to receive and focus on a specific type of treatment called androgen receptor signalling–targeted agents (ARSTs). They consider that in the future the traditional sequence of patients receiving multiple ARSTs may not be required due to more options available to patients. They report that "*only 46 of the 184 patients in our real-world cohort received sequential treatment with these ARSTs*". Applying the results of this thesis, a possible sensitivity analysis could be analysing OS as if no one went on to receive sequential ARSTs but instead received one of the more current sequential treatments.

[1] Reprinted from The Lancet Haematology, 11. 10, Yong, Kwee et al, Isatuximab plus carfilzomib–dexamethasone versus carfilzomib–dexamethasone in patients with relapsed multiple myeloma (IKEMA): overall survival analysis of a phase 3, randomised, controlled trial, e741-e750., Copyright (2024), with permission from Elsevier

Kadono et al[26] presented the results of a sequential trial where the second-line treatment was defined as part of the trial. However, in their paper they present information on participants not receiving the protocol defined second-line and summarise the other second-line treatments received. Considering the results of this thesis it could be of interest for the authors to analyse their data as if everyone went on to receive their defined second-line treatment.

Chu et al[27] considered the addition of the immunotherapy pembrolizumab to chemotherapy for advanced plural mesothelioma. They summarised the number of participants in both arms who subsequently went on to receive immunotherapy post-trial treatment. Whilst they discussed the potential delayed effects of immunotherapy, they did not fully consider the impact of subsequent immunotherapy. The authors could apply the same hypothetical estimand considered in this thesis; what would have happened if everyone went on to receive subsequent immunotherapy? The same hypothetical scenario could have been applied by Mita et al[28] who discussed the potential impact of immunotherapies being added as a second-line option during the study in their results. The use of the hypothetical estimand and two-stage estimation would enable the effect, if any, to be quantified.

Khalil et al[29] and Liu et al[30] only provided descriptive summaries of subsequent treatment lines. The impact of these were not commented on in the discussion. If one of the types of subsequent treatments reflect a more realistic patient pathway, the methods applied in this thesis could be used to analyse OS in this hypothetical scenario.

Finally, Morris et al[31] summarised their subsequent treatment lines and acknowledged that allowing crossover as part of the trial design may have impacted their OS analysis. However, they did not perform any sensitivity analysis to consider this. They did make a commitment to investigate it at the final analysis of OS. It may also be of interest to them to investigate the impact of other influential subsequent treatment lines as considered in this thesis.

From this updated review it is clear that, whilst subsequent treatment lines are being described, their consideration and reporting is inconsistent; and whilst the RPSFTM and IPCW are being considered more than reported in Chapter 2, TSE is not. This thesis provides evidence for when to consider the TSE in plausible clinical trial scenarios.

In summary, the preceding chapters of this thesis complement the recent movement away from a single definition of treatment switching by providing robust independent

evidence for existing methods considering the extended definition. However, the additional methodological developments highlight potential gaps in the findings. The impact of these potential gaps are discussed in Sections 5.5 and 5.6.

5.2.2 Healthcare Systems Data Access and Usage

A known and consistent barrier to the implementation of the methods considered in this thesis and discussed in Section 5.2.1 is the collection of suitable data to fit the models. The stakeholder consultation described in Chapter 3 asked stakeholders their opinions on the collection and recording of data once participants have stopped taking their trial intervention. The focus group, which formed part of the consultation, asked participant's for their preferences in the current research world, as well as the ideal research world. In the current research world, the group reached consensus on recording this information on a database specifically made for the trial. However, a database completed normally as part of standard practice was preferred in an ideal research world.

The ideal research world requires the use of data from electronic health records, often referred to as healthcare systems data, this all-encompassing term covers data collected throughout the healthcare ecosystem from primary care delivered by general practitioners to highly specialised treatment given in tertiary care.[32] At PhD outset, as discussed in Chapter 3, there had been an intention to take this information and use it to be able to signpost individuals on how implement the methods in practice using routine data sources, moving towards the ideal research world. However, during the course of the PhD the infrastructure around routine data in England has been in flux.

At PhD outset, July 2021, the data controller of the National Disease Registration Service (NDRS), which includes the collection of systemic anti-cancer therapy (SACT) in the NHS[33], was Public Health England (PHE). On 1st October 2021 this changed to NHS Digital.[34] On 1st February 2023 this changed again to NHS England through the merger of NHS Digital and NHS England.[35] At present, requests to access NDRS data are conducted through the Data Access Request Service (DARS).[36] However, this only includes data on patients in England. The Researcher Access Service from Research Data Scotland controls access to data from patients in Scotland[37], SAIL (Secure Anonymised Information Linkage) Wales[38] and the Northern Ireland Cancer Registry, Northern Ireland.[39]

As a result of the changing infrastructure, I was not confident that any investigations conducted and recommendations made during the time of this PhD would be relevant in the next 5 years, or fit for purpose given the different databases for each of the devolved nations given the likelihood that trialists recruit from throughout the United

Kingdom. A recent Delphi survey identified seven priority methodological questions on the use of routine data in clinical trials.[40] This in turn led to further research including recommendations for trialists to consider at the design stage of the trial.[41] The recommendations included a consideration of data quality assurance. One way to assess data quality is through data utility comparison studies. Data utility comparison studies (DUCkS) have been conducted in solid tumours such as prostate cancer. [42] Researchers found that whilst the routine data collected could be used it required substantial preparation and shifted the burden from sites dual reporting the information, to the research team requesting and cleaning the data. It also only considered data from England therefore the patients recruited outside of England could not have their data compared. Whilst there were a reduced number of recruiting sites outside of England this is still a limitation of the analysis. [43]

Therefore at this time I recommend that as per the recent commentary,[18] reporting guidelines should be updated to ensure that trial teams who intend to analyse OS in their statistical analysis plans collect data on subsequent anti-cancer treatment. However, the work presented in this thesis around the concerns of site and patient burden, and calls to action from the recent reviews, cannot be ignored. Therefore, future work discussed in Section 5.6 should consider more examples of how healthcare systems data can be accessed efficiently across the devolved nations and incorporated into trials.

In summary this thesis provides further evidence as to why good quality and accessible healthcare systems data is important, and that whilst it is not available trialists should be collecting the data themselves.

5.2.3 Patient and Public Involvement in Statistical Methodology Research

At PhD outset there were limited examples in the literature of using patient and public involvement (PPI) within the development of statistical methodology. However, funders of health and social care research have a shared commitment for public involvement in research they fund.[44] This is also the case for a PhD focussed on statistical methodology. Therefore, the PPI activities were designed in collaboration with experienced patient partners and discussions with researchers who had also worked in similar areas including Flight[45] and Goulau[46]. As described previously in the thesis, the PPI activities included: the co-design of the questionnaire for the stakeholder consultation; agreement on the contents and facilitation of the focus group meeting.

During the course of the PhD this resource gap has been explored by the PPI-SMART (Patient and Public Involvement for Statistical Methods Research Techniques) group

led by Professor Laura Gray.[47] A survey by the group identified that the barriers to PPI for statistical methodology include a lack of understanding of the benefits and training for both statisticians and public partners. The team then began to address this through; the development of resources, bringing researchers interested in this space together to share their work in a workshop[48], and growing a PPI for Statisticians Community of Practice (CoP).[49]

The work conducted within Chapter 3 of this thesis has been at the forefront of developments in this area. As a result of the implemented PPI activities I was invited to speak at the aforementioned workshop which led to an invited session at the Royal Statistical Society Conference in 2024[50] and becoming a founding member of the CoP.

The workshop and seminar enabled me to reflect on the co-design aspects of the questionnaire development and the format of the supporting materials. If I was designing the project again, I would use the UK standards for Public Involvement[51], the NIHR learning for involvement resources[52] and the resources developed by the PPI-SMART group[47] to design the PPI activities in conjunction with working with patient partners. I would also work with a company which specialises in science communication to develop the summary videos rather than editing them myself.[53-55] As even though they received good feedback from the members of the co-design group, they could have been more professional and benefited from further testing. Furthermore, through the networking opportunities, I learned about the work conducted by Suzie Cro on incorporating patient and public involvement in the development of estimands.[56] This resource would have been useful to consider when aiming to co-design the questionnaire with patient partners in order to identify the estimand of interest. However, despite the resource being published following the design of the questionnaire discussed in Chapter 3. The development of the questionnaire does align with the public partners four key topics presented in the resource:

1. *Topic 1: Trials should address questions that are relevant to patients and public:* In this PhD PPI activities were conducted during grant application to ensure the topic was of interest to patients and the public. In addition, whilst participatory rather than involvement, the stakeholder consultation was conducted to ensure that any proposed methodology addresses questions which are relevant to patients and public. From the stakeholder consultation I have ensured that the hypothetical scenario does not assume that no one stopped their trial treatment prior to death, only consider those who did not stop their trial treatment prior to death or shorten OS to be the time on treatment as these were of concern to the patient participants.

2. *Topic 2: Involve public partners early on in trial planning:* Public partners were involved in the early stages of both the development of the PhD project itself and the co-design of the questionnaire to shape the direction of the research.
3. *Topic 3: Education and communication is necessary:* Supporting materials on the topic of the PhD were developed and discussed with the public partners to address any questions.
4. *Topic 4: A change of culture is required of public partners and researchers working together:* By sharing my experiences, I have been helping to work towards changing the culture, where researchers work in collaboration with public partners.

In summary, this thesis provides a successful example for future researchers to learn from to develop their PPI activities in their statistical methodology research.

5.2.4 Research Participation in Statistical Methodology Research

It is important, in this thesis, to understand the difference between patient and public involvement and participation in research. Participation in research is defined as:

"The act of taking part in a research study, for example people being recruited to take part in a clinical trial or another kind of research study, joining in a focus group or completing a questionnaire." [57]

Therefore, whilst the opinions of patients and the public were used to guide the statistical methodology research conducted here, only the co-design of the questionnaire, facilitation of the focus group and inclusion on the stakeholder advisory committee (SAG) were true involvement. The remaining patient and public opinions were gathered through participation in the form of a stakeholder consultation.

As previously discussed in Chapter 3, a modified version of the RAND/UCLA appropriateness method was implemented for the stakeholder consultation. Alternative qualitative methods to obtain consensus include the Delphi and nominal group methods.

The Delphi method is usually conducted through multiple rounds of the same survey to stakeholder groups until consensus is reached.[58] There is no discussion between stakeholders and therefore the reasons behind answers are harder to determine and disagreement harder to resolve. One example of how the Delphi method has been used is to determine how patients and the public should be included in the development of statistical analysis plans.[59] As the stakeholder consultation aimed to reach consensus across patients and the public, health care professionals, statisticians and other data analysts, and payers and industry partners on a technical topic, it was

considered important for them to be able to discuss any potential disagreements. Therefore, it was deemed inappropriate to only implement a Delphi survey.

Alternatively, the nominal group method is a structured meeting where all participants provide their own thoughts on a topic, any required clarification is obtained, and then all suggestions are then ranked or rated as required. The nominal group method has been implemented in the past to obtain stakeholder perspectives on how to improve efficiency in the care of cancer patients.[60] Whilst the nominal group method has the added advantage of discussion over the Delphi method, it requires participants to have prior knowledge of the topic being discussed and ability to summarise these thoughts as topics for ranking. Therefore, it was also considered to not be an appropriate method for the stakeholder consultation, as non-statisticians / data analysts would have been at a disadvantage and consensus may not have been reached on all the topics the consultation aimed to cover.

The RAND/UCLA appropriateness method can be seen as a combination of the Delphi and nominal group methods.[61] Participants first complete a questionnaire and then meet to discuss any aspects which did not meet consensus, where rounds can be repeated as required. It has been used previously to determine quality measures in the care of older adults.[62] This was considered to be the most appropriate methodology for the stakeholder consultation, as it allowed the topics to be pre-specified through the use of the questionnaire and facilitated discussion. However, as previously noted the implemented RAND/UCLA was modified to allow a wider range of opinions to be considered through the questionnaire.

It is, in my understanding, rare to include participation within statistical methodological research, especially from non-statisticians and data analysts. However, it has been an invaluable experience in order to understand the current perspectives of all stakeholder groups and be confident that the research is of interest and relevant to all it impacts. Whilst patient and public involvement and stakeholder advisory groups are important, the added layer of participation widens the perspectives considered within the research and strengthens the conclusions.

In summary, this thesis provides a successful example for future researchers to learn how to utilise research participation from all stakeholders in statistical methodology research.

5.3 Recommendations for Cancer Clinical Trials

Following the completion of all elements of this thesis, I would recommend that cancer clinical trials incorporate TSE as a sensitivity analysis for the analysis of OS in their statistical analysis plans. This analysis has been written into the estimand framework for adoption into statistical analysis plans in Table 5.2.

Table 5.2: Template Estimand Framework for Hypothetical Estimation

Treatment Conditions	[*Summarise the trial interventions including an overview of the dose schedule and include the subsequent treatment of interest here*]
Population	All individuals with [*insert disease of interest*] provided they met the trial eligibility criteria*
Endpoint	Overall Survival: time from randomisation to death from any cause
Population Level Summaries	Hazard ratio of the experimental vs control
Intercurrent Events	Treatment discontinuation and uptake of subsequent treatment • Trial intervention discontinuation and started second-line with [*insert subsequent treatment of interest here*]: Treatment Policy approach • Trial intervention discontinuation and started second-line but did not have [*insert subsequent treatment of interest here*]: Hypothetical approach: what would have happened if everyone starting second-line treatment received [*insert subsequent treatment of interest here*]. • Trial intervention discontinuation without starting a second-line prior to death: Treatment Policy approach • Trial intervention discontinuation and palliative care: Treatment Policy approach • Any Third Line or Later: Hypothetical approach: what would have happened had no one received a third-line of treatment.¥ [*insert consideration of other intercurrent events here*]

** As per guidance the population in an estimand is the patients targeted by the research question not the analysis population.*

¥ This limitation is discussed in Chapter 4

In order to perform this robustly trial teams should:

1. Collect subsequent anti-cancer treatment information and verify through data management that the participant's subsequent treatment pathway is plausible.
2. Consult clinical colleagues on the participant and disease characteristics which influence subsequent anti-cancer treatment and participant outcome in order to collect this information at the point of subsequent treatment initiation to define a second baseline.

3. Review the clinical landscape on an ongoing basis and collaborate with clinical colleagues and patient partners to determine the most appropriate hypothetical scenario.
4. Report their findings and rationale for choices in their journal articles presenting OS results.

These recommendations align with the recent updates in the literature discussed above (Section 5.2) and the findings that stakeholders are comfortable with data being collected on participants following trial treatment discontinuation, provided there is a clear rationale. They also aim to overcome the challenges identified in the real data case study where assumptions had to be made on the timing of some subsequent therapies due to inconsistencies in the data, and enable a clear definition of the second baseline required for TSE.

5.4 Extensions to Other Areas

The focus of this thesis has been the analysis of OS in cancer clinical trials. However, there are potential parallels which the findings could be applied to.

The literature review identified that surrogate endpoints such as progression-free, disease-free and relapse-free survival are being used as primary endpoints in clinical trials. For the majority, in line with clinical guidance, participants will not begin a new systemic anti-cancer treatment prior to progression, relapse or disease re-occurrence. However, there may be a clinical need to start a new systemic anti-cancer treatment following participant or clinician withdrawal due to intolerance or choice. Recent reflections from the EORTC (European Organisation for Research and Treatment of Cancer) have highlighted poor practice in this area where researchers have censored or excluded participants who received new systemic treatment prior to progression (manuscript in draft). This poor practice reflects the naïve methods discussed in this thesis, and researchers may benefit from considering the use of a similar hypothetical estimand to that proposed here.

The literature review excluded any studies which presented meta-analyses as outside the scope of this thesis. However, the recommendations for cancer clinical trials to consider a hypothetical estimand and clearly report their findings, would align with and extend a recent investigation into the topic of treatment switching in meta-analysis.[63] A simulation study found that when the results of different estimands are pooled the resulting meta-analysis does not target any estimand of interest and provides misleading results. Therefore, if the individual trials included in a meta-analysis have a consistent approach to the uptake of subsequent treatments, as proposed here, this may improve the performance of the meta-analysis.

The work presented in this thesis may also be relevant for the development of hypothetical estimands for time-to-event endpoints which are impacted by similar intercurrent events in other relapsing conditions areas such as alcoholism or mental health. For example, researchers may investigate how an intervention impacts on the time in which participants are employed, in the event that interventions fail, other interventions may be implemented at disease-relapse.

5.5 Limitations

The limitations of the individual components of this work were previously discussed in the respective chapters. Here I consider the limitations to the work as a whole and any points which would benefit from further deliberation.

A key strength of the thesis is the inclusion of all stakeholder perspectives in the development of statistical methodology through the use of a stakeholder advisory group, co-design group and participation. However, the research supervisory team did not include an individual who specialised in qualitative methods. This was an oversight during the development of the project, and whilst there were expertise within my institution where informal advice could be sought, including an experienced qualitative supervisor could have helped to refine the questionnaire further and support the co-design process and analysis. On reflection, the questionnaire contained too many avenues and some of the results may be subject to reporting bias. The additional questions for statisticians and data analysts would have been more appropriate if answered in a separate questionnaire on an institute level to understand clinical trials units and similar institutions current practice as a whole rather than the potentially narrow and repeated view of what one or more individuals at an institution knew. This alternative approach has been utilised recently to consider the uptake of the estimand framework across the UKCRC CTU network.[15]

The literature review in Chapter 2 considered that the use of two search mechanisms should have provided a full picture of the current practice around subsequent treatment lines. However, as indicated above, the work by Ouwens et al was excluded. On review of the literature search, Current Therapeutic Research, where the article was published, is not indexed in the MEDLINE database which was used for the literature search.[64] As MEDLINE includes over 5000 journals it was considered to be a thorough search of the literature. However, with this finding it may have been useful to include a second database such as EMBASE which on review does include Current Therapeutic Research.[65] It may have also been prudent to include a citation search for TSD 16 to fully appreciate the ongoing work in the area. However, with the update to TSD 24 discussed above, it is unlikely further developments have been missed.

The update to TSD 24 during the course of the thesis has resulted in some differences with the implemented methods and the recommendations in the guidance. As indicated above, TSE with g-estimation could be implemented as part of future work. However, TSD 24 also suggests that separate AFT models should be implemented for each randomisation arm. This was not implemented in Chapter 4. Instead randomisation allocation was used as a covariate in the model. As the same subsequent treatment effect was used in the DGM; this decision is unlikely to have affected the results of the thesis. However, a separate scenario investigating different subsequent treatment effects depending on randomisation could be considered in future work.

5.6 Future work

Considering the current methodological landscape, there seems to be a disjoint between the pharmaceutical, health technology and academic sectors in the consideration given to subsequent treatment lines with the latter lagging in the collection and implementation of existing and new methodology. Therefore, as an academic statistician, my primary focus for future work has to be on increasing the awareness of the impact subsequent treatment lines can have on the analysis of OS, and encouraging academic trialists to conduct the necessary data collection to investigate the impact by disseminating my findings through journal articles, conference abstracts and professional networks. This in turn would support the impact of any future methodological work discussed in Chapter 4, which as previously stated, will be decided upon following further stakeholder engagement and closure of the feedback loop.

It is likely that, in addition to lack of awareness, resource is a barrier to the current collection of information on subsequent treatment. Therefore alternative routes to obtain this information should be explored. As indicated in Section 5.2.2 this became outside the scope of this thesis due to changes in national infrastructure. However, future work will build on the work conducted by other researchers to investigate accessing healthcare systems data, including subsequent treatment information, for participants enrolled in multiple myeloma trials conducted through Leeds CTRU (Integrated Research Application System Number 343775). The feasibility and practicability of trial teams accessing this data and using it to quantify the effect of subsequent treatment lines in trials will be assessed as a secondary aim of this project. The lessons learned in myeloma will then be applied to other blood cancers and supplement the existing evidence from solid tumours.

5.7 Conclusion

In conclusion, this thesis is an example of how qualitative and quantitative research can be combined, with PPI intertwined, to answer a question on statistical methodological. As a result it provides robust evidence for the use of TSE in order to quantify the effect of subsequent treatments on the analysis of OS in cancer clinical trials. It complements the current perspective of the research community and highlights the potential inconsistencies between the different disciplines. There is a role for the findings to be used to increase awareness and improve research practices around the collection of data and consideration of the issue. Indeed the findings should support the 473 cancer clinical trials which are currently listed as "ongoing" on the ISCRTN database[66] or any trial impacted by one of the 110 technology appraisal guidelines which have been updated during this PhD.[67] However, as always, unknowns remain and these will be investigated through future work, ensuring that any future direction is co-designed to consider the perspective of all stakeholders interested in improving and understanding the effect clinical trial interventions have on OS.

5.8 Chapter 5 References

1. Latimer NR and Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
2. Bell Gorrod H, Latimer NR and Abrams KR. NICE DSU Technical Support Document 24: Adjusting survival time estimates in the presence of treatment switching: An update to TSD 16, Available from https://www.sheffield.ac.uk/nice-dsu (2024).
3. White IR, Babiker AG, Walker S and Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Stat Med* 1999; 18: 2617-2634. DOI: 10.1002/(sici)1097-0258(19991015)18:19<2617::aid-sim187>3.0.co;2-e.
4. Xu Y, Wu M, He W, et al. Teasing Out the Overall Survival Benefit With Adjustment for Treatment Switching to Multiple Treatments. *Statistics in Biopharmaceutical Research* 2022; 14: 592-601. DOI: 10.1080/19466315.2021.1914716.
5. Li L, Tang S and Jiang L. On an enhanced rank-preserving structural failure time model to handle treatment switch, crossover, and dropout. *Stat Med* 2017; 36: 1532-1547. 20170122. DOI: 10.1002/sim.7224.
6. Ouwens M, Darilay A, Zhang Y, et al. Assessing the Influence of Subsequent Immunotherapy on Overall Survival in Patients with Unresectable Stage III Non–Small Cell Lung Cancer from the PACIFIC Study. *Current Therapeutic Research* 2021; 95: 100640. DOI: https://doi.org/10.1016/j.curtheres.2021.100640.
7. Latimer NR, White IR, Tilling K and Siebert U. Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding. *Stat Methods Med Res* 2020; 29: 2900-2918. 20200330. DOI: 10.1177/0962280220912524.
8. Jackson D, Ran D, Zhang F, et al. New Methods for Two-Stage Treatment Switching Estimation. *Pharm Stat* 2025; 24: e2462. DOI: https://doi.org/10.1002/pst.2462.
9. Luo X, Li M, Wu C, et al. A proposed approach for analyzing post-study therapy effect in survival analysis. *J Biopharm Stat* 2016; 26: 790-800. 20150806. DOI: 10.1080/10543406.2015.1074916.

10. Chen Q, Zhang F, Chen MH and Cong XJ. Estimation of treatment effects and model diagnostics with two-way time-varying treatment switching: an application to a head and neck study. *Lifetime Data Anal* 2020; 26: 685-707. 20200303. DOI: 10.1007/s10985-020-09495-0.
11. Huang CH, Chen YH, Wang JL and Wang M. Semiparametric copula-based analysis for treatment effects in the presence of treatment switching. *Stat Med* 2020; 39: 2936-2948. 20200624. DOI: 10.1002/sim.8585.
12. Kuehne F, Rochau U, Paracha N, et al. Estimating Treatment-Switching Bias in a Randomized Clinical Trial of Ovarian Cancer Treatment: Combining Causal Inference with Decision-Analytic Modeling. *Med Decis Making* 2022; 42: 194-207. 20211020. DOI: 10.1177/0272989x211026288.
13. Sullivan TR, Latimer NR, Gray J, et al. Adjusting for Treatment Switching in Oncology Trials: A Systematic Review and Recommendations for Reporting. *Value Health* 2020; 23: 388-396. DOI: 10.1016/j.jval.2019.10.015.
14. Clark TP, Wicentowski RH, Cro S, et al. Are estimands being correctly used? A review of UK research protocols. *Trials* 2025; 26: 310. DOI: 10.1186/s13063-025-08991-8.
15. Stiles M, Cafferty F, Goulao B, et al. CP8-3 Evaluating Estimand Implementation in Clinical Trials in the UK in Abstracts from the Society for Clinical Trials 46th Annual Meeting (2025). *Clinical Trials* 2025; 22. DOI: https://doi.org/10.1177/17407745251393006.
16. Watkins C, Kleine E, Miranda M, et al. Further Practical Guidance on Adjusting Time-To-Event Outcomes for Treatment Switching. *Pharm Stat* 2025; 24: e70019. DOI: https://doi.org/10.1002/pst.70019.
17. Olivier T, Haslam A and Prasad V. Post-progression treatment in cancer randomized trials: a cross-sectional study of trials leading to FDA approval and published trials between 2018 and 2020. *BMC Cancer* 2023; 23: 448. 20230517. DOI: 10.1186/s12885-023-10917-z.
18. Lee D and Torres GM. Randomized controlled trial reporting guidelines should be updated to include information on subsequent treatments. *Cancer* 2025; 131: e35922. DOI: 10.1002/cncr.35922.
19. Adam R, Piedvache C, Chiche L, et al. Liver transplantation plus chemotherapy versus chemotherapy alone in patients with permanently unresectable colorectal liver metastases (TransMet): results from a multicentre, open-label, prospective, randomised controlled trial. *Lancet* 2024; 404: 1107-1118. Multicenter Study Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S0140-6736(24)01595-2.
20. Parker CC, James ND, Brawley CD, et al. Radiotherapy to the prostate for men with metastatic prostate cancer in the UK and Switzerland: Long-term results from the STAMPEDE randomised controlled trial. *PLoS Med* 2022; 19: e1003998. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pmed.1003998.
21. Parker CC, Clarke NW, Cook AD, et al. Adding 6 months of androgen deprivation therapy to postoperative radiotherapy for prostate cancer: a comparison of short-course versus no androgen deprivation therapy in the RADICALS-HD randomised controlled trial. *Lancet* 2024; 403: 2405-2415. Randomized Controlled Trial Multicenter Study Comparative Study Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(24)00548-8.
22. Rao S, Samalin-Scalzi E, Evesque L, et al. Retifanlimab with carboplatin and paclitaxel for locally recurrent or metastatic squamous cell carcinoma of the anal canal (POD1UM-303/InterAACT-2): a global, phase 3 randomised controlled trial. *Lancet* 2025; 405: 2144-2152. Randomized Controlled Trial Clinical Trial, Phase III Multicenter Study. DOI: https://dx.doi.org/10.1016/S0140-6736(25)00631-2.
23. Yong K, Martin T, Dimopoulos MA, et al. Isatuximab plus carfilzomib-dexamethasone versus carfilzomib-dexamethasone in patients with relapsed multiple myeloma (IKEMA): overall survival analysis of a phase 3, randomised, controlled trial. *The Lancet Haematology* 2024; 11: e741-e750. Randomized Controlled Trial Clinical

Trial, Phase III Multicenter Study. DOI: https://dx.doi.org/10.1016/S2352-3026(24)00148-0.
24. Qi C, Liu C, Peng Z, et al. Claudin-18 isoform 2-specific CAR T-cell therapy (satri-cel) versus treatment of physician's choice for previously treated advanced gastric or gastro-oesophageal junction cancer (CT041-ST-01): a randomised, open-label, phase 2 trial. *Lancet* 2025; 405: 2049-2060. Clinical Trial, Phase II Randomized Controlled Trial Multicenter Study. DOI: https://dx.doi.org/10.1016/S0140-6736(25)00860-8.
25. Izumi K, Shima T, Mita K, et al. Enzalutamide Versus Abiraterone plus Prednisolone Before Chemotherapy for Castration-resistant Prostate Cancer: A Multicenter Randomized Controlled Trial. *Eur Urol Open Sci* 2022; 41: 16-23. DOI: https://dx.doi.org/10.1016/j.euros.2022.04.016.
26. Kadono Y, Konaka H, Nohara T, et al. Efficacy and Safety of First-Line Cytokines versus Sunitinib and Second-Line Axitinib for Patients with Metastatic Renal Cell Carcinoma (ESCAPE Study): A Phase III, Randomized, Sequential Open-Label Study. *Cancers (Basel)* 2023; 15: 13. DOI: https://dx.doi.org/10.3390/cancers15102745.
27. Chu Q, Perrone F, Greillier L, et al. Pembrolizumab plus chemotherapy versus chemotherapy in untreated advanced pleural mesothelioma in Canada, Italy, and France: a phase 3, open-label, randomised controlled trial. *Lancet* 2023; 402: 2295-2306. Randomized Controlled Trial Clinical Trial, Phase III Research Support, Non-U.S. Gov't DOI: https://dx.doi.org/10.1016/S0140-6736(23)01613-6.
28. Mita K, Izumi K, Goriki A, et al. Enzalutamide versus Abiraterone Plus Prednisolone for Nonmetastatic Castration-Resistant Prostate Cancer: A Sub-Analysis from the ENABLE Study for PCa. *Cancers (Basel)* 2024; 16: 24. DOI: https://dx.doi.org/10.3390/cancers16030508.
29. Khalil KA, Musallam HS, Hassan MA and Mahmoud IA. Triplet (FOLFOXIRI) Versus Doublet (FOLFOX or FOLFIRI) Regimen as First Line Treatment in Metastatic Colorectal Carcinoma, a Prospective Phase II, Randomized Controlled Trial. *Asian Pac J Cancer Prev* 2022; 23: 3421-3429. Randomized Controlled Trial Clinical Trial, Phase II. DOI: https://dx.doi.org/10.31557/APJCP.2022.23.10.3421.
30. Liu X, Zhang X, Jiang S, et al. Site-specific therapy guided by a 90-gene expression assay versus empirical chemotherapy in patients with cancer of unknown primary (Fudan CUP-001): a randomised controlled trial. *Lancet Oncol* 2024; 25: 1092-1102. Randomized Controlled Trial Comparative Study DOI: https://dx.doi.org/10.1016/S1470-2045(24)00313-9.
31. Morris MJ, Castellano D, Herrmann K, et al. ¹⁷⁷Lu-PSMA-617 versus a change of androgen receptor pathway inhibitor therapy for taxane-naive patients with progressive metastatic castration-resistant prostate cancer (PSMAfore): a phase 3, randomised, controlled trial. *Lancet* 2024; 404: 1227-1239. Randomized Controlled Trial Clinical Trial, Phase III Multicenter Study Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(24)01653-2.
32. NHS England. The healthcare ecosystem, https://digital.nhs.uk/developer/guides-and-documentation/introduction-to-healthcare-technology/the-healthcare-ecosystem (accessed 29th October 2025).
33. NHS England. National Disease Registration Service, https://digital.nhs.uk/ndrs (accessed 28th October 2025).
34. Department of Health and Social Care. Technical notice: transfer of National Disease Registration Service from Public Health England to NHS Digital, https://www.gov.uk/government/publications/location-of-public-health-england-phe-functions-from-1-october-2021/technical-notice-transfer-of-national-disease-registration-service-from-public-health-england-to-nhs-digital (accessed 28th October 2025).
35. NHS England. Organisation information for NHS Digital, https://digital.nhs.uk/about-nhs-digital/organisation-information-for-nhs-digital (accessed 28th October 2025).

36. NHS England. Data Access Request Service (DARS), https://digital.nhs.uk/services/data-access-request-service-dars (accessed 28th October 2025).
37. Research Data Scotland. Researcher Access Service, https://www.researchdata.scot/work-and-impact/current-projects/researcher-access-service/ (accessed 28th October 2025).
38. SAIL. SAIL Databank, https://saildatabank.com/ (accessed 28th October 2025).
39. Queen's University Belfast. Northern Ireland Cancer Registry, https://www.qub.ac.uk/research-centres/nicr/ (accessed 28th October 2025).
40. Williams ADN, Davies G, Farrin AJ, et al. A DELPHI study priority setting the remaining challenges for the use of routinely collected data in trials: COMORANT-UK. *Trials* 2023; 24: 243. DOI: 10.1186/s13063-023-07251-x.
41. Toader A-M, Campbell MK, Quint JK, et al. Using healthcare systems data for outcomes in clinical trials: issues to consider at the design stage. *Trials* 2024; 25: 94. DOI: 10.1186/s13063-024-07926-z.
42. Murray ML, Love SB, Yorke-Edwards V, et al. Data utility comparison studies (DUCkS) of adjuvant chemotherapy & radiotherapy treatments in the STAMPEDE trial: Can we use NHS England data?, https://www.mrcctu.ucl.ac.uk/posters/data-utility-comparison-studies-ducks-of-adjuvant-chemotherapy-radiotherapy-treatments-in-the-stampede-trial/ (accessed 5th November 2025).
43. STAMPEDE Trial. STAMPEDE Centres, https://www.google.com/maps/d/viewer?mid=1Z-dSzfcFXd2GgIhIRTbUxddx1tk&ll=53.846788181012165%2C-2.6012179844423895&z=6 (accessed 5th November 2025).
44. NHS Health Research Authority. The Shared Commitment to Public Involvement in health and social care research, https://www.hra.nhs.uk/planning-and-improving-research/best-practice/public-involvement/shared-commitment-public-involvement-health-and-social-care-research/ (accessed 15th October 2025).
45. Flight L, Julious S, Brennan A, et al. How can health economics be used in the design and analysis of adaptive clinical trials? A qualitative analysis. *Trials* 2020; 21: 252. DOI: 10.1186/s13063-020-4137-2.
46. Beatriz Goulao, Marion Campbell, Katie Gillies and Ramsay C. P-96 What methods are used to involve patients and the public in

numerical aspects of research? A scoping review in Meeting abstracts from the 5th International Clinical Trials Methodology Conference (ICTMC 2019). *Trials* 2019; 20: 579. DOI: 10.1186/s13063-019-3688-6.
47. NIHR Leicester Biomedical Research Centre. PPI for Statistical Methodology and Research Techniques, https://leicesterbrc.nihr.ac.uk/ppismart/ (accessed 15th October 2025).
48. Smith A, Worboys H, Begum S, et al. Incorporation of Patient and Public Involvement in Statistical Methodology Research: Summary of Workshop Proceedings. *Stat Med* 2025; 44: e70159. DOI: 10.1002/sim.70159.
49. NIHR Leicester Biomedical Research Centre. PPI for Statisticians Community of Practice, https://leicesterbrc.nihr.ac.uk/ppismart/ppi-stats-community-of-practice/ (accessed 15th October 2025).
50. SOLVE Study. SOLVE PPI RSS Presentation, https://youtu.be/C2oZxLz97KU (2024, accessed 15th October 2025).
51. UK Public Involvement Standards Development Partnership group. UK Standards for Public Involvement, https://sites.google.com/nihr.ac.uk/pi-standards/home (accessed 15th October 2025).
52. NIHR. Resources and training for public involvement in research, https://www.learningforinvolvement.org.uk/ (2024, accessed 9th April 2026).
53. SOLVE Study. What is the SOLVE Study?, https://youtu.be/dZsle9e-Npl (2022, accessed 8th February 2023).
54. SOLVE Study. How will the SOLVE study answer it's question?, https://youtu.be/AdfS1FeBhBs (2022, accessed 8th February 2023).

55. SOLVE Study. SOLVE Summary Video, https://youtu.be/HKYbl8vcCaw (2022, accessed 8th February 2023).
56. Cro S, Kahan BC, Patel A, et al. Starting a conversation about estimands with public partners involved in clinical trials: a co-developed tool. *Trials* 2023; 24: 443. DOI: 10.1186/s13063-023-07469-9.
57. NIHR. Glossary https://www.nihr.ac.uk/glossary (accessed 26th October 2025).
58. Keeney S, Hasson F and McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *International Journal of Nursing Studies* 2001; 38: 195-200. DOI: https://doi.org/10.1016/S0020-7489(00)00044-4.
59. Goulão B, Morris TP, Blazeby J, et al. Involving patients and the public In sTatistIcal Analysis pLans (INITIAL): A delphi survey. *PLoS ONE* 2023; 18: e0292257. DOI: 10.1371/journal.pone.0292257.
60. Ferrara L, Otto M, Aapro M, et al. How to improve efficiency in cancer care: Dimensions, methods, and areas of evaluation. *Journal of Cancer Policy* 2022; 34: 100355. DOI: https://doi.org/10.1016/j.jcpo.2022.100355.
61. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA appropriateness method user's manual*. 2001. Rand Corp Santa Monica CA.
62. Correia RH, Dash D, Jones A, et al. Primary care quality for older adults: Practice-based quality measures derived from a RAND/UCLA appropriateness method study. *PLoS ONE* 2024; 19: e0297505. DOI: 10.1371/journal.pone.0297505.
63. Metcalfe RK, Remiro-Azócar A, Vuong Q, et al. Estimands and Their Implications for Evidence Synthesis for Oncology: A Simulation Study of Treatment Switching in Meta-Analysis V2. arXiv, 2025.
64. NIH. NLM Catalog, https://www.ncbi.nlm.nih.gov/nlmcatalog/?term=%22Current+therapeutic+research%22 (accessed 28th October 2025).
65. ELSEVIER. 2025-07_Embase-journals, https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fsupportcontent.elsevier.com%2FRightNow%2520Next%2520Gen%2FEmbase%2F2025-07_Embase-journals.xlsx&wdOrigin=BROWSELINK (accessed 28th October 2025).
66. ISCRTN. ISCRTN Registry, https://www.isrctn.com/search?q=&filters=trialStatus:ongoing,conditionCategory:Cancer&page=1&sort=&pageSize=10 (accessed 14th April 2026).
67. NICE. NICE Guidance and Advice List, https://www.nice.org.uk/guidance/published?q=cancer&from=2021-06-01&to=2026-04-14&ngt=Technology+appraisal+guidance (accessed 14th April 2026).

Appendix A Ethical Approval

In line with University of Leeds guidance the confirmation letters of ethical approval for the original application and substantial amendment are included here.

NHS
Health Research Authority
Yorkshire & The Humber - Bradford Leeds Research Ethics Committee
NHSBT Newcastle Blood Donor Centre
Holland Drive
Newcastle upon Tyne
NE2 4NQ

Telephone: 0207 104 8210

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

01 August 2022

Miss Kara-Louise Royle
Senior Medical Statistician
University of Leeds
Leeds Institute of Clinical Trials Research
University of Leeds
Leeds
LS2 9NL

Dear Miss Royle

Study title:	**The analysis of overall survival in a changing clinical world; the methodologicalimplications of relapsing cancers in randomised controlled trials.**
REC reference:	**22/YH/0155**
Protocol number:	**N/A**
IRAS project ID:	**299766**

Thank you for your letter of 27 July 2022, responding to the Research Ethics Committee's (REC) request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Good practice principles and responsibilities

The UK Policy Framework for Health and Social Care Research sets out principles of good practice in the management and conduct of health and social care research. It also outlines the responsibilities of individuals and organisations, including those related to the four elements of research transparency:

1. registering research studies
2. reporting results
3. informing participants
4. sharing study data and tissue

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

All research should be registered in a publicly accessible database and we expect all researchers, research sponsors and others to meet this fundamental best practice standard.

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a publicly accessible database within six weeks of recruiting the first research participant. For this purpose, 'clinical trials' are defined as:

- clinical trial of an investigational medicinal product
- clinical investigation or other study of a medical device

- combined trial of an investigational medicinal product and an investigational medical device
- other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice.

Failure to register a clinical trial is a breach of these approval conditions, unless a deferral has been agreed by the HRA (for more information on registration and requesting a deferral see: Research registration and research project identifiers).

If you have not already included registration details in your IRAS application form you should notify the REC of the registration details as soon as possible.

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter.

Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit:
https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/

N.B. If your study is related to COVID-19 we will aim to publish your research summary within 3 days rather than three months.

During this public health emergency, it is vital that everyone can promptly identify all relevant research related to COVID-19 that is taking place globally. If you haven't already done so, please register your study on a public registry as soon as possible and provide the REC with the registration detail, which will be posted alongside other information relating to your project. We are also asking sponsors not to request deferral of publication of research summary for any projects relating to COVID-19. In addition, to facilitate finding and extracting studies related to COVID-19 from public databases, please enter the WHO official acronym for the coronavirus disease (COVID-19) in the full title of your study. Approved COVID-19 studies can be found at: https://www.hra.nhs.uk/covid-19-research/approved-covid-19-research/

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol

- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report
- Reporting results

The latest guidance on these topics can be found at https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

NHS/HSC sites

The favourable opinion applies to all NHS/HSC sites taking part in the study, subject to confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or management permission (in Scotland) being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS/HSC sites

I am pleased to confirm that the favourable opinion applies to any non-NHS/HSC sites listed in the application, subject to site management permission being obtained prior to the start of the study at the site.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Covering letter on headed paper [Cover Letter]	1	14 June 2022
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Proof of Professional Indemnity Cover]	1	01 October 2021
IRAS Checklist XML [Checklist_17062022]		17 June 2022
IRAS Checklist XML [Checklist_27072022]		27 July 2022
Letter from funder [Award Acceptance Letter]	1	18 December 2020
Letters of invitation to participant [SOLVE Advertising Materal]	5	16 June 2022
Non-validated questionnaire [Study Information Sheet, Consent and Questionnaire]	2	27 July 2022
Other [UoL Public Indemnity Proof]	1	29 September 2021
Other [Response to comments from Ethics]	1	27 July 2022
REC Application Form [REC_Form_17062022]		17 June 2022
Referee's report or other scientific critique report [NIHR Application Reviewer Comments]	1	04 September 2020
Research protocol or project proposal [NIHR Doctoral Fellowship Application Form]	1	29 October 2020
Summary CV for Chief Investigator (CI) [Kara-Louise Royle CV]	1	11 May 2022
Summary CV for supervisor (student research) [CVs of all academic supervisors]	1	01 November 2021

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at:
https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS project ID: 299766 Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely

Pp.

Dr Janet Holt
Chair

Email:bradfordleeds.rec@hra.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

After ethical review guidance for sponsors and investigators – **Non CTIMP Standard Conditions of Approval**]

Copy to: Ms Jean Uniacke

NHS
Health Research Authority

Yorkshire & The Humber - Bradford Leeds Research Ethics Committee
NHSBT Newcastle Blood Donor Centre
Holland Drive
Newcastle upon Tyne
NE2 4NQ

Tel: 0207 104 8388

02 February 2023

Miss Kara-Louise Royle
Leeds Institute of Clinical Trials Research
University of Leeds
Leeds
LS2 9NL

Dear Miss Royle

Study title:	**The analysis of overall survival in a changing clinical world; the methodologicalimplications of relapsing cancers in randomised controlled trials.**
REC reference:	**22/YH/0155**
Protocol number:	**N/A**
Amendment number:	**Amendment One**
Amendment date:	**17 January 2023**
IRAS project ID:	**299766**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation. However, suggested that the researchers' changes "As a reminder the meeting will be recorded to aid note taking" in Discussion Group documentation to that both audio and video will be recorded.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Completed Amendment Tool [22-PR-0858-amendment one]	1	17 January 2023

Other [DiscussionGroupMeetingSlides]	1	16 January 2023
Other [SOLVEDiscussionGroup]	1	16 January 2023
Other [SOLVE Post-Discussion Groups V1 20230116]	1	16 January 2023

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

Working with NHS Care Organisations

Sponsors should ensure that they notify the R&D office for the relevant NHS care organisation of this amendment in line with the terms detailed in the categorisation email issued by the lead nation for the study.

Amendments related to COVID-19

We will update your research summary for the above study on the research summaries section of our website. During this public health emergency, it is vital that everyone can promptly identify all relevant research related to COVID-19 that is taking place globally. If you have not already done so, please register your study on a public registry as soon as possible and provide the HRA with the registration detail, which will be posted alongside other information relating to your project.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS Project ID - 299766:	**Please quote this number on all correspondence**

Yours sincerely



On Behalf Of
Dr Janet Holt
Chair

Yorkshire & The Humber - Bradford Leeds Research Ethics Committee

Attendance at Sub-Committee of the REC meeting on 02 February 2023

Committee Members:

Name	*Profession*	*Present*	*Notes*
Mr Farhan Joy Allahyar	Clinical Research Fellow for Workforce Race Equality Standard	Yes	
Dr Janet Holt	Associate Professor in Healthcare Ethics	Yes	Chair

Also in attendance:

Name	*Position (or reason for attending)*
Mrs Stacey Bamford	Approvals Administrator

Appendix B Chapter 2 Supplementary Material

How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review – Supplementary Material

Appendix Table B.1: PRISMA Abstract Checklist

Section and Topic	Item #	Checklist item	Reported (Yes/No)
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	N/A
Synthesis of results	6	Specify the methods used to present and synthesise results.	N/A
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	Yes

Section and Topic	Item #	Checklist item	Reported (Yes/No)
Registration	12	Provide the register name and registration number.	N/A

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71 For more information, visit: http://www.prisma-statement.org/

Appendix Table B.2: PRISMA Reporting Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 22
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Supplementary Table B.1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 23-25
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 25
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Figure 2.1
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 25
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Table 2.1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 28
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 28
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Table 2.2

	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Table 2.2
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Not a meta-analysis
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 28
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Not a meta-analysis
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Not a meta-analysis
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Not a meta-analysis
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Not a meta-analysis
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not a meta-analysis
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not a meta-analysis
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not a meta-analysis
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not a meta-analysis
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 2.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Figure 2.1
Study characteristics	17	Cite each included study and present its characteristics.	Figure 2.1
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Not a meta-analysis

Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Not a meta-analysis
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Not a meta-analysis
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not a meta-analysis
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not a meta-analysis
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not a meta-analysis
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not a meta-analysis
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not a meta-analysis
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 36-40
	23b	Discuss any limitations of the evidence included in the review.	Page 39-40
	23c	Discuss any limitations of the review processes used.	Page 39-40
	23d	Discuss implications of the results for practice, policy, and future research.	Page 40
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Not a meta-analysis
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not prepared
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 41
Competing interests	26	Declare any competing interests of review authors.	Page 41
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 41

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Appendix Table B.3: Summarised Clinical Effectiveness Papers

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	Hepatectomy Followed by mFOLFOX6 Versus Hepatectomy Alone for Liver-Only Metastatic Colorectal Cancer (JCOG0603): A Phase II or III Randomized Controlled Trial. [1]	2021	Bowel / Lower GI	Secondary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for the discrepancy between OS and surrogate endpoints.	No
2	Late Toxicity After Adjuvant Conventional Radiation Versus Image-Guided Intensity-Modulated Radiotherapy for Cervical Cancer (PARCER): A Randomized Controlled Trial. [2]	2021	Gynaecological	Secondary	Hazard ratio and P-value from a logrank test	No	N/A	N/A	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
5	Five-Year Outcomes With Pembrolizumab Versus Chemotherapy for Metastatic Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score >= 50. [3]	2021	Lung	Secondary	Hazard ratio and 95% CI from an adjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
7	Modified FOLFIRINOX versus S-1 as second-line chemotherapy in gemcitabine-failed metastatic pancreatic cancer patients: A randomised controlled trial (MPACA-3). [4]	2021	Upper GI/HPB	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for non-stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
13	Bevacizumab Plus mFOLFOX6 Versus mFOLFOX6 Alone as First-Line Treatment for RAS Mutant Unresectable Colorectal Liver-Limited Metastases: The BECOME Randomized Controlled Trial. [5]	2020	Bowel / Lower GI	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	Yes	No	No	No	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
23	Bevacizumab biosimilar LY01008 compared with bevacizumab (Avastin) as first-line treatment for Chinese patients with unresectable, metastatic, or recurrent non-squamous non-small-cell lung cancer: A multicenter, randomized, double-blinded, phase III trial. [6]	2021	Lung	Secondary	Other (Median OS and logrank p-value.)	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
24	Prostate-Only Versus Whole-Pelvic Radiation Therapy in High-Risk and Very High-Risk Prostate Cancer (POP-RT): Outcomes From Phase III Randomized Controlled Trial. [7]	2021	Genito-urinary	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	No	N/A	N/A	N/A	N/A	N/A	N/A
37	Second-line FOLFOX chemotherapy versus active symptom control for advanced biliary tract cancer (ABC-06): a phase 3, open-label, randomised, controlled trial. [8]	2021	Upper GI/HPB	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	Yes	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	Yes	Yes	Yes	Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	No
45	Intracranial Efficacy and Survival With Tucatinib Plus Trastuzumab and Capecitabine for Previously Treated HER2-Positive Breast Cancer With Brain Metastases in the HER2CLIMB Trial. [9]	2020	Breast	Exploratory	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A
52	Sunitinib Versus Sorafenib as Initial Targeted Therapy for mCC-RCC With Favorable/Intermediate Risk: Multicenter Randomized Trial CROSS-J-RCC. [10]	2020	Genito-urinary	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	Yes	Yes	Yes	None or limited options of subsequent treatment lines for patients are given as a reason as to why subsequent treatment lines will not have affected the OS results.	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
54	Cemiplimab monotherapy for first-line treatment of advanced non-small-cell lung cancer with PD-L1 of at least 50%: a multicentre, open-label, global, phase 3, randomised, controlled trial. [11]	2021	Lung	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
67	Metronomic capecitabine as maintenance treatment after first line induction with XELOX for metastatic colorectal cancer patients. [12]	2020	Bowel / Lower GI	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	No	N/A	Yes	Not Specified	Unsure - Not Specified	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
69	Carfilzomib or bortezomib in combination with lenalidomide and dexamethasone for patients with newly diagnosed multiple myeloma without intention for immediate autologous stem-cell transplantation (ENDURANCE): a multicentre, open-label, phase 3, randomised, controlled trial. [13]	2020	Haematological	Secondary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
71	Sequencing of Androgen-Deprivation Therapy With External-Beam Radiotherapy in Localized Prostate Cancer: A Phase III Randomized Controlled Trial. [14]	2020	Genito-urinary	Secondary	Hazard ratio from an unadjusted cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A
78	Liver transplantation in hepatocellular carcinoma after tumour downstaging (XXL): a randomised, controlled, phase 2b/3 trial. [15]	2020	Upper GI/HPB	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A
82	Weekly dose-dense chemotherapy in first-line epithelial ovarian, fallopian tube, or primary peritoneal carcinoma treatment (ICON8): primary progression free survival analysis results from a GCIG phase 3 randomised controlled trial. [16]	2019	Gynaecological	Primary or Co-Primary	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
91	Oxaliplatin-Based Adjuvant Chemotherapy for Rectal Cancer After Preoperative Chemoradiotherapy (ADORE): Long-Term Results of a Randomized Controlled Trial. [17]	2019	Bowel / Lower GI	Secondary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
96	The effect of modulated electro-hyperthermia on local disease control in HIV-positive and -negative cervical cancer women in South Africa: Early results from a phase III randomised controlled trial. [18]	2019	Gynaecological	Secondary	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
97	Combination of icotinib and chemotherapy as first-line treatment for advanced lung adenocarcinoma in patients with sensitive EGFR mutations: A randomized controlled study. [19]	2019	Lung	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
98	Erlotinib Versus Gemcitabine Plus Cisplatin as Neoadjuvant Treatment of Stage IIIA-N2 EGFR-Mutant Non-Small-Cell Lung Cancer (EMERGING-CTONG 1103): A Randomized Phase II Study. [20]	2019	Lung	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
103	Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. [21]	2019	Genito-urinary	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No
105	Atezolizumab with or without cobimetinib versus regorafenib in previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase 3, randomised, controlled trial. [22]	2019	Bowel / Lower GI	Primary or Co-Primary	Hazard ratio and P-value from a stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Yes	No	Yes	No	N/A	No
126	Randomized Trial of Systemic Therapy After Involved-Field Radiotherapy in Patients With Early-Stage Follicular Lymphoma: TROG 99.03. [23]	2018	Haematological	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	Yes	No	No	No	Yes	The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
135	A randomized, controlled trial evaluating the efficacy and safety of BTH1677 in combination with bevacizumab, carboplatin, and paclitaxel in first-line treatment of advanced non-small cell lung cancer. [24]	2018	Lung	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Unsure	N/A	N/A	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
138	Neoadjuvant Chemotherapy Followed by Radical Surgery Versus Concomitant Chemotherapy and Radiotherapy in Patients With Stage IB2, IIA, or IIB Squamous Cervical Cancer: A Randomized Controlled Trial. [25]	2018	Gynaecological	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for both strat and non-strat factors	Yes	Yes	No	Yes	No	N/A	No
142	Bevacizumab Maintenance Versus No Maintenance During Chemotherapy-Free Intervals in Metastatic Colorectal Cancer: A Randomized Phase III Trial (PRODIGE 9). [26]	2018	Bowel / Lower GI	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
144	Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. [27]	2018	Genito-urinary	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No
152	Adding Celecoxib With or Without Zoledronic Acid for Hormone-Naive Prostate Cancer: Long-Term Survival Results From an Adaptive, Multiarm, Multistage, Platform, Randomized Controlled Trial. [28]	2017	Genito-urinary	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	Yes	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No
153	Randomized controlled trial of S-1 versus docetaxel in patients with non-small-cell lung cancer previously treated with platinum-based chemotherapy (East Asia S-1 Trial in Lung Cancer). [29]	2017	Lung	Primary or Co-Primary	Hazard ratio and p-value from an unadjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Both Adjusted and Unadjusted Analysis	Yes - Unclear on what adjusted for	Yes	Yes	No	Yes	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
155	Peri-operative chemotherapy with or without bevacizumab in operable oesophagogastric adenocarcinoma (UK Medical Research Council ST03): primary analysis results of a multicentre, open-label, randomised phase 2-3 trial. [30]	2017	Upper GI/HPB	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	Yes	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
160	Results of a phase II randomized controlled clinical trial comparing efficacy of Cabazitaxel versus Docetaxel as second line or above therapy in recurrent head and neck cancer. [31]	2017	Other	Secondary	Hazard ratio and p-value from an unadjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	No	N/A	N/A	N/A	N/A	N/A	N/A
161	Effect of docetaxel duration on clinical outcomes: exploratory analysis of CLEOPATRA, a phase III randomized controlled trial. [32]	2017	Breast	Exploratory	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	No	No	No	No	N/A	Yes[1]

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
166	Carfilzomib or bortezomib in relapsed or refractory multiple myeloma (ENDEAVOR): an interim overall survival analysis of an open-label, randomised, phase 3 trial. [33]	2017	Haematological	Secondary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	Yes[2]
167	Randomized Controlled Trial Testing the Efficacy of Platinum-Free Interval Prolongation in Advanced Ovarian Cancer: The MITO-8, MaNGO, BGOG-Ov1, AGO-Ovar2.16, ENGOT-Ov1, GCIG Study. [34]	2017	Gynaecological	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	No	N/A	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for both strat and non-strat factors	No	N/A	N/A	N/A	N/A	N/A	N/A
177	Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. [35]	2017	Lung	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
179	A randomized trial of TLR-2 agonist CADI-05 targeting desmocollin-3 for advanced non-small-cell lung cancer. [36]	2017	Lung	Primary or Co-Primary	Hazard ratio and P-value from a stratified logrank test.	Yes	No	N/A	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	N/A	N/A	No	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
183	Randomized controlled trial of lobaplatin plus etoposide vs. cisplatin plus etoposide as first-line therapy in patients with extensive-stage small cell lung cancer. [37]	2019	Lung	Secondary	Hazard ratio from an unadjusted cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No
187	A Phase II Randomized Controlled Trial: Definitive Concurrent Chemoradiotherapy with Docetaxel Plus Cisplatin versus 5-Fluorouracil plus Cisplatin in Patients with Oesophageal Squamous Cell Carcinoma. [38]	2017	Upper GI/HPB	Primary or Co-Primary	Other (p-value from a logrank test with estimates of OS at 1 and 2 years)	Yes	No	N/A	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
192	Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. [39]	2016	Genito-urinary	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	Yes	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
199	Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. [40]	2016	Other	Exploratory	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	Yes	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	Yes[3]
201	Adjuvant Oral Uracil-Tegafur with Leucovorin for Colorectal Cancer Liver Metastases: A Randomized Controlled Trial. [41]	2016	Bowel / Lower GI	Secondary	Hazard ratio from an adjusted Cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Unsure	N/A	N/A	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
202	TRAPEZE: a randomised controlled trial of the clinical effectiveness and cost-effectiveness of chemotherapy with zoledronic acid, strontium-89, or both, in men with bony metastatic castration-refractory prostate cancer. [42]	2016	Genito-urinary	Secondary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for non-stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
208	Salvage radiotherapy with or without short-term hormone therapy for rising prostate-specific antigen concentration after radical prostatectomy (GETUG-AFU 16): a randomised, multicentre, open-label phase 3 trial. [43]	2016	Genito-urinary	Secondary	Hazard ratio and P-value from a stratified logrank test.	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
213	Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. [44]	2016	Lung	Primary or Co-Primary	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	No
215	Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. [45]	2016	Lung	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
224	Gastrectomy plus chemotherapy versus chemotherapy alone for advanced gastric cancer with a single non-curable factor (REGATTA): a phase 3, randomised controlled trial. [46]	2016	Upper GI/HPB	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for non-stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for non-stratification factors	Yes	Yes	No	Yes	Yes	Other (Cross-over trial design)	No
226	Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. [47]	2016	Lung	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	No
234	Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. [48]	2015	Gynaecological	Secondary	Other (P-value from a logrank test difference in restricted mean survival time at 5 years.)	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Not Specified	Unsure - Not Specified	Yes	No	No	No	Yes	Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
235	Methotrexate, Doxorubicin, and Cisplatin (MAP) Plus Maintenance Pegylated Interferon Alfa-2b Versus MAP Alone in Patients With Resectable High-Grade Osteosarcoma and Good Histologic Response to Preoperative MAP: First Results of the EURAMOS-1 Good Response Randomized Controlled Trial. [49]	2015	Other	Secondary	Hazard ratio from an adjusted Cox model and p-value from a logrank test.	Yes	Yes	Yes	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A
236	Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. [50]	2015	Gynaecological	Primary or Co-Primary	Hazard ratio and 95% CI from an adjusted Cox model	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A
237	Final Report of the Intergroup Randomized Study of Combined Androgen-Deprivation Therapy Plus Radiotherapy Versus Androgen-Deprivation Therapy Alone in Locally Advanced Prostate Cancer. [51]	2015	Genito-urinary	Primary or Co-Primary	Hazard ratio and p-value from an unadjusted Cox model	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for both strat and non-strat factors	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
245	Locoregional treatment versus no treatment of the primary tumour in metastatic breast cancer: an open-label randomised controlled trial. [52]	2015	Breast	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	No	Yes	No	Yes	Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	No
247	Neoadjuvant plus adjuvant bevacizumab in early breast cancer (NSABP B-40 [NRG Oncology]): secondary outcomes of a phase 3, randomised controlled trial. [53]	2015	Breast	Exploratory	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A
248	Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. [54]	2015	Upper GI/HPB	Primary or Co-Primary	Hazard ratio from an unadjusted cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for both strat and non-strat factors	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
250	Adjuvant lymph-node field radiotherapy versus observation only in patients with melanoma at high risk of further lymph-node field relapse after lymphadenectomy (ANZMTG 01.02/TROG 02.01): 6-year follow-up of a phase 3, randomised controlled trial. [55]	2015	Other	Secondary	Hazard ratio and P-value from a stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No
254	Afatinib versus erlotinib as second-line treatment of patients with advanced squamous cell carcinoma of the lung (LUX-Lung 8): an open-label randomised controlled phase 3 trial. [56]	2015	Lung	Secondary	Hazard ratio from an unadjusted cox model and p-value from an stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	No	Yes	No	N/A	No
263	High-dose radiotherapy with short-term or long-term androgen deprivation in localised prostate cancer (DART01/05 GICOR): a randomised, controlled, phase 3 trial. [57]	2015	Genito-urinary	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
267	Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. [58]	2015	Breast	Secondary	Other (5 year OS rate and I think a p-value from a logrank test.)	No	N/A	N/A	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
269	Whole-breast irradiation with or without a boost for patients treated with breast-conserving surgery for early breast cancer: 20-year follow-up of a randomised phase 3 trial. [59]	2015	Breast	Primary or Co-Primary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Not Specified	Unsure - Not Specified	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
275	Use of thoracic radiotherapy for extensive stage small-cell lung cancer: a phase 3 randomised controlled trial. [60]	2015	Lung	Primary or Co-Primary	Other (Overall survival at 1 year but their HR is clearly for the entire follow-up period.)	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
283	Gemcitabine and capecitabine with or without telomerase peptide vaccine GV1001 in patients with locally advanced or metastatic pancreatic cancer (TeloVac): an open-label, randomised, phase 3 trial. [61]	2014	Upper GI/HPB	Primary or Co-Primary	Hazard ratio from an adjusted Cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	Yes[4]
285	Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. [62]	2014	Haematological	Primary or Co-Primary	Other (Odds Ratio with 95% CI and p-value calculated at 5 years.)	No	N/A	N/A	No	N/A	N/A	No	N/A	N/A	Yes	Yes	No	Yes	No	N/A	No
286	Adjuvant bevacizumab in patients with melanoma at high risk of recurrence (AVAST-M): preplanned interim results from a multicentre, open-label, randomised controlled phase 3 study. [63]	2014	Other	Primary or Co-Primary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for both strat and non-strat factors	Yes	Yes	No	Yes	Yes	Other (Limited Follow-up)	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
290	Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: long-term results from the MRC RT01 randomised controlled trial. [64]	2014	Genito-urinary	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	No	Yes	The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	No
299	Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. [65]	2014	Other	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	No	N/A	N/A	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
300	Concomitant cisplatin plus radiotherapy and high-dose-rate brachytherapy versus radiotherapy alone for stage IIIB epidermoid cervical cancer: a randomized controlled trial. [66]	2014	Gynaecological	Secondary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
301	Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial. [67]	2014	Lung	Secondary	Hazard ratio from an unadjusted cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	No
302	Docetaxel versus active symptom control for refractory oesophagogastric adenocarcinoma (COUGAR-02): an open-label, phase 3 randomised controlled trial. [68]	2014	Upper GI/HPB	Primary or Co-Primary	Hazard Ratio and 95% CI from an unadjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
306	Randomized, multicenter, phase II study of CO-101 versus gemcitabine in patients with metastatic pancreatic ductal adenocarcinoma: including a prospective evaluation of the role of hENT1 in gemcitabine or CO-101 sensitivity. [69]	2013	Upper GI/HPB	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	Yes	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	Yes	Yes	No	No	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
312	2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. [70]	2013	Breast	Secondary	Hazard ratio from an unadjusted cox model and p-value from a logrank test.	No	N/A	N/A	Yes	Not Specified	Unsure - Not Specified	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	No	Yes	Yes	Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	Yes[5]
318	Randomized controlled trial of toremifene 120 mg compared with exemestane 25 mg after prior treatment with a non-steroidal aromatase inhibitor in postmenopausal women with hormone receptor-positive metastatic breast cancer. [71]	2013	Breast	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
319	Chemoradiotherapy with or without cetuximab in patients with oesophageal cancer (SCOPE1): a multicentre, phase 2/3 randomised trial. [72]	2013	Upper GI/HPB	Secondary	Hazard ratio from an adjusted Cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	No	Yes	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
326	Intraoperative radiotherapy versus external radiotherapy for early breast cancer (ELIOT): a randomised controlled equivalence trial. [73]	2013	Breast	Secondary	Hazard ratio from an unadjusted cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	No	Yes	No	N/A	No
328	Long-term results of dose-dense paclitaxel and carboplatin versus conventional paclitaxel and carboplatin for treatment of advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer (JGOG 3016): a randomised, controlled, open-label trial. [74]	2013	Gynaecological	Secondary	Hazard ratio from an adjusted Cox model and p-value from a logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	No	No	No	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
329	Erlotinib versus docetaxel as second-line treatment of patients with advanced non-small-cell lung cancer and wild-type EGFR tumours (TAILOR): a randomised controlled trial. [75]	2013	Lung	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for both strat and non-strat factors	Yes	Yes	No	Yes	No	N/A	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
333	Randomized controlled trial of cetuximab plus chemotherapy for patients with KRAS wild-type unresectable colorectal liver-limited metastases. [76]	2013	Bowel / Lower GI	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
335	Taking into account successive treatment lines in the analysis of a colorectal cancer randomised trial. [77]	2013	Bowel / Lower GI	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	No	N/A	N/A	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Yes	Yes	Yes	Yes	Other (Sequence)	Yes[6]
337	Radiofrequency ablation with or without transcatheter arterial chemoembolization in the treatment of hepatocellular carcinoma: a prospective randomized trial. [78]	2013	Upper GI/HPB	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for both strat and non-strat factors	Yes	Yes	No	Yes	No	N/A	No
344	Vinorelbine and gemcitabine vs vinorelbine and carboplatin as first-line treatment of advanced NSCLC. A phase III randomised controlled trial by the Norwegian Lung Cancer Study Group. [79]	2012	Lung	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Not Specified	Unsure - Not Specified	Yes	Yes	No	Yes	Yes	The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	No

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
345	Celecoxib plus hormone therapy versus hormone therapy alone for hormone-sensitive prostate cancer: first results from the STAMPEDE multiarm, multistage, randomised controlled trial. [80]	2012	Genito-urinary	Primary or Co-Primary	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A	Yes	The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	N/A
349	Bevacizumab plus oxaliplatin-based chemotherapy as adjuvant treatment for colon cancer (AVANT): a phase 3 randomised controlled trial. [81]	2012	Bowel / Lower GI	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Unsure	N/A	N/A	Unsure	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
351	Adjuvant chemotherapy with doxorubicin, ifosfamide, and lenograstim for resected soft-tissue sarcoma (EORTC 62931): a multicentre randomised controlled trial. [82]	2012	Other	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
354	Adjuvant radiotherapy versus observation alone for patients at risk of lymph-node field relapse after therapeutic lymphadenectomy for melanoma: a randomised trial. [83]	2012	Other	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Both Adjusted and	Yes - Unclear on what adjusted for	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
358	Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. [84]	2012	Upper GI/HPB	Secondary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	N/A	N/A	N/A	N/A
366	Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. [85]	2011	Bowel / Lower GI	Primary or Co-Primary	Other (Upper bound of the confidence interval from an 80% logrank test.)	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	Yes	Yes	No	No	Yes	The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	No
367	Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. [86]	2011	Bowel / Lower GI	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	Yes	Yes	No	Yes	Yes	The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	No
373	First-line treatment with zoledronic acid as compared with clodronic acid in multiple myeloma (MRC Myeloma IX): a randomised controlled trial. [87]	2010	Haematological	Primary or Co-Primary	Hazard ratio and p-value from an adjusted Cox model	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for non-stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for both strat and non-strat factors	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
374	Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. [88]	2010	Breast	Secondary	Hazard ratio and p-value from an unadjusted Cox model	No	N/A	N/A	No	N/A	N/A	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	No	N/A	N/A	N/A	N/A	N/A	N/A
376	Early versus delayed treatment of relapsed ovarian cancer (MRC OV05/EORTC 55955): a randomised trial. [89]	2010	Gynaecological	Primary or Co-Primary	Hazard ratio and P-value from a stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for non-stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for both strat and non-strat factors	Yes	Yes	No	No	Yes	Other (Early Access)	No
379	Neo-adjuvant chemotherapy alone or with regional hyperthermia for localised high-risk soft-tissue sarcoma: a randomised phase 3 multicentre study. [90]	2010	Other	Secondary	Hazard ratio and P-value from a stratified logrank test.	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Unclear on what adjusted for	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A
381	Randomized trial comparing conventional-dose with high-dose conformal radiation therapy in early-stage adenocarcinoma of the prostate: long-term results from proton radiation oncology group/american college of radiology 95-09. [91]	2010	Genito-urinary	Secondary	Other (Proportions and a logrank p-value)	No	N/A	N/A	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
382	Overall survival analysis of a phase II randomized controlled trial of a Poxviral-based PSA-targeted immunotherapy in metastatic castration-resistant prostate cancer. [92]	2010	Genito-urinary	Secondary	Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	Yes	No	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	No	No	No	Yes	None or limited options of subsequent treatment lines for patients are given as a reason as to why subsequent treatment lines will not have affected the OS results.	No
384	Lenalidomide plus high-dose dexamethasone versus lenalidomide plus low-dose dexamethasone as initial therapy for newly diagnosed multiple myeloma: an open-label randomised controlled trial. [93]	2010	Haematological	Exploratory	Other (OS at one year and p-value from logrank test.)	Yes	Yes	No	No	N/A	N/A	Yes	Both Adjusted and Unadjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	No	No	No	Yes	The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	No
385	Autologous stem cell transplantation as a first-line treatment strategy for chronic lymphocytic leukemia: a multicenter, randomized, controlled trial from the SFGM-TC and GFLLC. [94]	2011	Haematological	Secondary	Other (% of OS at 3 and 5 years)	Yes	No	N/A	No	N/A	N/A	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	No	N/A	N/A	N/A	N/A	N/A	N/A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
390	Single-injection depot progesterone before surgery and survival in women with operable breast cancer: a randomized controlled trial. [95]	2011	Breast	Secondary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Not Specified	Unsure - Not Specified	Yes	Adjusted Analysis	Yes - Adjusted for both strat and non-strat factors	No	N/A	N/A	N/A	N/A	N/A	N/A
395	Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. [96]	2011	Breast	Secondary	Hazard ratio and p-value from an unadjusted Cox model	Yes	Yes	No	No	N/A	N/A	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for the discrepancy between OS and surrogate endpoints.	Yes[7]
404	Best supportive care compared with chemotherapy for unresectable gall bladder cancer: a randomized controlled study. [97]	2010	Upper GI/HPB	Primary or Co-Primary	Hazard ratio from a Cox model and p-value from a logrank test	Yes	No	N/A	Yes	Not Specified	Unsure - Not Specified	Yes	Not Specified	Unsure - Not Specified	No	N/A	N/A	N/A	N/A	N/A	N/A
405	Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. [98]	2010	Upper GI/HPB	Primary or Co-Primary	Hazard ratio and P-value from a logrank test	Yes	Yes	No	Yes	Adjusted Analysis	Yes - Adjusted for some or all stratification factors	Yes	Unadjusted Analysis	No - Only Unadjusted analysis	Yes	Yes	No	Yes	Yes	The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	No

Column names: 1 = ORN , 2 = Title , 3 = Year , 4 = Cancer , 5 = OS Endpoint , 6 = Primary analysis , 7 = Was a Kaplan-Meier curve for OS presented in the main text? , 8 = Were the numbers at risk included in the Kaplan-Meier curve? , 9 = Were the number of patients censored or number of events included in the at risk table , 10 = Was a logrank test for OS presented in the main text? , 11 = If included, was an logrank test for OS presented using unadjusted or adjusted or both techniques in the main text (separate from subgroup analysis)? , 12 = If included, what level of adjustment was considered for an logrank test for OS in the main text (separate from subgroup analysis)? , 13 = Was a Cox-PH Model for OS presented in the main text (separate from subgroup analysis)? , 14 = If included, was a Cox-PH Model for OS presented using unadjusted or adjusted or both techniques in the main text (separate from subgroup analysis)? , 15 = If included, what level of adjustment was considered for a Cox-PH Model for OS in the main text (separate from subgroup analysis)? , 16 = Were subsequent treatments mentioned in the paper? , 17 = Was the percentage or number of participants who received later lines included in the main text or appendix? , 18 = Was a breakdown or summary of the number of subsequent treatments (Two, Three, Four) included in the main text or appendix?, 19 = Was a breakdown or summary of the type of subsequent treatments (Treatment 1, Treatment 2, Treatment 3) included in the main text or appendix?, 20 = Were subsequent treatment lines mentioned in the discussion? , 21 = How were subsequent treatments mentioned in the discussion? , 22 = Was additional analysis conducted to account for subsequent treatment lines?

Additional Analysis 1=Time Dependent Covariate, 2=Landmark analysis to calculate post-progression survival, 3=Censor point of subsequent therapies, 4=Post-Hoc separating out the groups in an OS analysis, 5 = Cut-off of follow-up time to see how the effect changed., 6 = Shared Frailty Models and Time Dependent Covariates. 7 Censored at point of cross-over.

Appendix Table B.4: Summary of the Clinical Effectiveness Papers by Year of Publication

	2010 (n=9)	2011 (n=5)	2012 (n=6)	2013 (n=10)	2014 (n=8)	2015 (n=13)	2016 (n=9)	2017 (n=10)	2018 (n=5)	2019 (n=8)	2020 (n=7)	2021 (n=8)	Total (n=98)
OS Endpoint													
Primary or Co-Primary	4 (44.4%)	2 (40.0%)	3 (50.0%)	4 (40.0%)	6 (75.0%)	6 (46.2%)	5 (55.6%)	7 (70.0%)	1 (20.0%)	3 (37.5%)	1 (14.3%)	3 (37.5%)	45 (45.9%)
Secondary	4 (44.4%)	3 (60.0%)	3 (50.0%)	6 (60.0%)	2 (25.0%)	6 (46.2%)	3 (33.3%)	2 (20.0%)	4 (80.0%)	5 (62.5%)	5 (71.4%)	5 (62.5%)	48 (49.0%)
Exploratory	1 (11.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.7%)	1 (11.1%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	5 (5.1%)
Was a Kaplan-Meier curve for OS presented in the main text?													
Yes	7 (77.8%)	5 (100.0%)	5 (83.3%)	8 (80.0%)	7 (87.5%)	12 (92.3%)	7 (77.8%)	10 (100.0%)	5 (100.0%)	6 (75.0%)	7 (100.0%)	7 (87.5%)	86 (87.8%)
No	2 (22.2%)	0 (0.0%)	0 (0.0%)	2 (20.0%)	1 (12.5%)	1 (7.7%)	1 (11.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	8 (8.2%)
N/A[1]	0 (0.0%)	0 (0.0%)	1 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (11.1%)	0 (0.0%)	0 (0.0%)	2 (25.0%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Were the numbers at risk included in the Kaplan-Meier curve?													
Yes	5 (55.6%)	4 (80.0%)	5 (83.3%)	8 (80.0%)	7 (87.5%)	12 (92.3%)	7 (77.8%)	7 (70.0%)	5 (100.0%)	6 (75.0%)	6 (85.7%)	7 (87.5%)	79 (80.6%)
No	2 (22.2%)	1 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (30.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	7 (7.1%)
N/A	2 (22.2%)	0 (0.0%)	1 (16.7%)	2 (20.0%)	1 (12.5%)	1 (7.7%)	2 (22.2%)	0 (0.0%)	0 (0.0%)	2 (25.0%)	0 (0.0%)	1 (12.5%)	12 (12.2%)
Were the number of patients censored or number of events included in the at risk table													
Yes	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	1 (7.7%)	2 (22.2%)	3 (30.0%)	0 (0.0%)	2 (25.0%)	2 (28.6%)	2 (25.0%)	13 (13.3%)
No	5 (55.6%)	4 (80.0%)	5 (83.3%)	7 (70.0%)	7 (87.5%)	11 (84.6%)	5 (55.6%)	4 (40.0%)	5 (100.0%)	4 (50.0%)	4 (57.1%)	5 (62.5%)	66 (67.3%)
N/A	4 (44.4%)	1 (20.0%)	1 (16.7%)	2 (20.0%)	1 (12.5%)	1 (7.7%)	2 (22.2%)	3 (30.0%)	0 (0.0%)	2 (25.0%)	1 (14.3%)	1 (12.5%)	19 (19.4%)
Were confidence intervals included in the Kaplan-Meier curve?													
No	7 (77.8%)	5 (100.0%)	5 (83.3%)	8 (80.0%)	7 (87.5%)	12 (92.3%)	7 (77.8%)	10 (100.0%)	5 (100.0%)	6 (75.0%)	7 (100.0%)	7 (87.5%)	86 (87.8%)
N/A	2 (22.2%)	0 (0.0%)	1 (16.7%)	2 (20.0%)	1 (12.5%)	1 (7.7%)	2 (22.2%)	0 (0.0%)	0 (0.0%)	2 (25.0%)	0 (0.0%)	1 (12.5%)	12 (12.2%)

	2010 (n=9)	2011 (n=5)	2012 (n=6)	2013 (n=10)	2014 (n=8)	2015 (n=13)	2016 (n=9)	2017 (n=10)	2018 (n=5)	2019 (n=8)	2020 (n=7)	2021 (n=8)	Total (n=98)
Primary analysis													
Hazard Ratio and 95% CI from an unadjusted Cox model	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.0%)
Hazard ratio and 95% CI from an adjusted Cox model	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	2 (2.0%)
Hazard ratio and P-value from a logrank test	1 (11.1%)	2 (40.0%)	3 (50.0%)	3 (30.0%)	1 (12.5%)	1 (7.7%)	0 (0.0%)	1 (10.0%)	4 (80.0%)	1 (12.5%)	1 (14.3%)	1 (12.5%)	19 (19.4%)
Hazard ratio and P-value from a stratified logrank test.	2 (22.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.7%)	1 (11.1%)	1 (10.0%)	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	6 (6.1%)
Hazard ratio and p-value from an adjusted Cox model	1 (11.1%)	0 (0.0%)	0 (0.0%)	3 (30.0%)	2 (25.0%)	0 (0.0%)	2 (22.2%)	3 (30.0%)	1 (20.0%)	1 (12.5%)	0 (0.0%)	2 (25.0%)	15 (15.3%)
Hazard ratio and p-value from an unadjusted Cox model	1 (11.1%)	1 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.7%)	0 (0.0%)	2 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (5.1%)
Hazard ratio from a Cox model and p-value from a logrank test	1 (11.1%)	0 (0.0%)	2 (33.3%)	0 (0.0%)	1 (12.5%)	2 (15.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	2 (28.6%)	1 (12.5%)	10 (10.2%)
Hazard ratio from an adjusted Cox model and p-value from a logrank test.	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (20.0%)	1 (12.5%)	1 (7.7%)	1 (11.1%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (6.1%)
Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	1 (11.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.7%)	4 (44.4%)	1 (10.0%)	0 (0.0%)	1 (12.5%)	3 (42.9%)	2 (25.0%)	13 (13.3%)
Hazard ratio from an unadjusted cox model and p-value from a logrank test.	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (20.0%)	1 (12.5%)	1 (7.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	5 (5.1%)
Hazard ratio from an unadjusted cox model and p-value from an stratified logrank test.	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	2 (2.0%)
N/A[1]	0 (0.0%)	0 (0.0%)	1 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (11.1%)	0 (0.0%)	0 (0.0%)	2 (25.0%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Other	2 (22.2%)	2 (40.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	3 (23.1%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	10 (10.2%)
Were subsequent treatments mentioned in the paper?													

	2010 (n=9)	2011 (n=5)	2012 (n=6)	2013 (n=10)	2014 (n=8)	2015 (n=13)	2016 (n=9)	2017 (n=10)	2018 (n=5)	2019 (n=8)	2020 (n=7)	2021 (n=8)	Total (n=98)
Yes	4 (44.4%)	3 (60.0%)	2 (33.3%)	8 (80.0%)	7 (87.5%)	5 (38.5%)	7 (77.8%)	6 (60.0%)	4 (80.0%)	5 (62.5%)	3 (42.9%)	5 (62.5%)	59 (60.2%)
No	5 (55.6%)	2 (40.0%)	4 (66.7%)	2 (20.0%)	1 (12.5%)	8 (61.5%)	2 (22.2%)	4 (40.0%)	1 (20.0%)	3 (37.5%)	4 (57.1%)	3 (37.5%)	39 (39.8%)
Was the percentage or number of participants who received later lines included in the main text or appendix?													
Yes	2 (22.2%)	3 (60.0%)	1 (16.7%)	7 (70.0%)	7 (87.5%)	3 (23.1%)	7 (77.8%)	5 (50.0%)	3 (60.0%)	5 (62.5%)	2 (28.6%)	5 (62.5%)	50 (51.0%)
No	2 (22.2%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	2 (15.4%)	0 (0.0%)	1 (10.0%)	1 (20.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	8 (8.2%)
N/A	5 (55.6%)	2 (40.0%)	5 (83.3%)	2 (20.0%)	1 (12.5%)	8 (61.5%)	2 (22.2%)	4 (40.0%)	1 (20.0%)	3 (37.5%)	4 (57.1%)	3 (37.5%)	40 (40.8%)
Was a breakdown or summary of the number of subsequent treatments (Two, Three, Four) included in the main text or appendix?													
Yes	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	1 (7.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	1 (12.5%)	4 (4.1%)
No	4 (44.4%)	3 (60.0%)	1 (16.7%)	7 (70.0%)	7 (87.5%)	4 (30.8%)	7 (77.8%)	5 (50.0%)	4 (80.0%)	5 (62.5%)	2 (28.6%)	4 (50.0%)	53 (54.1%)
N/A[2]	5 (55.6%)	2 (40.0%)	5 (83.3%)	2 (20.0%)	1 (12.5%)	8 (61.5%)	2 (22.2%)	5 (50.0%)	1 (20.0%)	3 (37.5%)	4 (57.1%)	3 (37.5%)	41 (41.8%)
Was a breakdown or summary of the type of subsequent treatments (Treatment 1, Treatment 2, Treatment 3) included in the main text or appendix?													
Yes	1 (11.1%)	2 (40.0%)	1 (16.7%)	6 (60.0%)	6 (75.0%)	3 (23.1%)	7 (77.8%)	4 (40.0%)	3 (60.0%)	5 (62.5%)	2 (28.6%)	5 (62.5%)	45 (45.9%)
No	3 (33.3%)	1 (20.0%)	0 (0.0%)	2 (20.0%)	1 (12.5%)	2 (15.4%)	0 (0.0%)	1 (10.0%)	1 (20.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	12 (12.2%)
N/A[2]	5 (55.6%)	2 (40.0%)	5 (83.3%)	2 (20.0%)	1 (12.5%)	8 (61.5%)	2 (22.2%)	5 (50.0%)	1 (20.0%)	3 (37.5%)	4 (57.1%)	3 (37.5%)	41 (41.8%)
Was additional analysis conducted to account for subsequent treatment lines?													
Yes	0 (0.0%)	1 (20.0%)	0 (0.0%)	2 (20.0%)	1 (12.5%)	0 (0.0%)	1 (11.1%)	2 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	7 (7.1%)
No	4 (44.4%)	2 (40.0%)	1 (16.7%)	6 (60.0%)	6 (75.0%)	5 (38.5%)	6 (66.7%)	3 (30.0%)	4 (80.0%)	5 (62.5%)	3 (42.9%)	5 (62.5%)	50 (51.0%)
N/A[2]	5 (55.6%)	2 (40.0%)	5 (83.3%)	2 (20.0%)	1 (12.5%)	8 (61.5%)	2 (22.2%)	5 (50.0%)	1 (20.0%)	3 (37.5%)	4 (57.1%)	3 (37.5%)	41 (41.8%)
Were subsequent treatment lines mentioned in the discussion?													
Yes	4 (44.4%)	3 (60.0%)	2 (33.3%)	3 (30.0%)	6 (75.0%)	3 (23.1%)	6 (66.7%)	2 (20.0%)	1 (20.0%)	1 (12.5%)	3 (42.9%)	5 (62.5%)	39 (39.8%)

	2010 (n=9)	2011 (n=5)	2012 (n=6)	2013 (n=10)	2014 (n=8)	2015 (n=13)	2016 (n=9)	2017 (n=10)	2018 (n=5)	2019 (n=8)	2020 (n=7)	2021 (n=8)	Total (n=98)
No	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (50.0%)	1 (12.5%)	2 (15.4%)	1 (11.1%)	4 (40.0%)	3 (60.0%)	4 (50.0%)	0 (0.0%)	0 (0.0%)	20 (20.4%)
N/A	5 (55.6%)	2 (40.0%)	4 (66.7%)	2 (20.0%)	1 (12.5%)	8 (61.5%)	2 (22.2%)	4 (40.0%)	1 (20.0%)	3 (37.5%)	4 (57.1%)	3 (37.5%)	39 (39.8%)
How were subsequent treatments mentioned in the discussion?													
The uptake of subsequent treatments is given as a reason for the discrepancy between OS and surrogate endpoints.	0 (0.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	2 (2.0%)
The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	1 (11.1%)	0 (0.0%)	1 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (33.3%)	0 (0.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (6.1%)
The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	1 (11.1%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	2 (25.0%)	1 (7.7%)	1 (11.1%)	2 (20.0%)	0 (0.0%)	1 (12.5%)	2 (28.6%)	3 (37.5%)	14 (14.3%)
The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	0 (0.0%)	2 (40.0%)	1 (16.7%)	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	1 (12.5%)	0 (0.0%)	1 (11.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (3.1%)

	2010 (n=9)	2011 (n=5)	2012 (n=6)	2013 (n=10)	2014 (n=8)	2015 (n=13)	2016 (n=9)	2017 (n=10)	2018 (n=5)	2019 (n=8)	2020 (n=7)	2021 (n=8)	Total (n=98)
Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	2 (15.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	4 (4.1%)
None or limited options of subsequent treatment lines for patients are given as a reason as to why subsequent treatment lines will not have affected the OS results.	1 (11.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	2 (2.0%)
Other	1 (11.1%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	1 (12.5%)	0 (0.0%)	1 (11.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
N/A	5 (55.6%)	2 (40.0%)	4 (66.7%)	7 (70.0%)	2 (25.0%)	10 (76.9%)	3 (33.3%)	8 (80.0%)	4 (80.0%)	7 (87.5%)	4 (57.1%)	3 (37.5%)	59 (60.2%)

[1] Not-Applicable here refers to papers where OS is a pre-specified endpoint, but the data is not mature at the time of the publication.
[2] Note one paper stated no patients received subsequent treatments so is classed as N/A for these summaries.
N/A: Not Applicable; OS: Overall survival.

Appendix Table B.5: Summary of the Effectiveness Papers by Cancer

	Bowel Lower GI (n=12)	Breast (n=12)	Genito-urinary (n=15)	Gynaecological (n=10)	Haematological (n=7)	Lung (n=18)	Upper GI HPB (n=15)	Other (n=9)	Total (n=98)
OS Endpoint									
Primary or Co-Primary	4 (33.3%)	2 (16.7%)	7 (46.7%)	4 (40.0%)	2 (28.6%)	10 (55.6%)	13 (86.7%)	3 (33.3%)	45 (45.9%)
Secondary	8 (66.7%)	7 (58.3%)	8 (53.3%)	6 (60.0%)	4 (57.1%)	8 (44.4%)	2 (13.3%)	5 (55.6%)	48 (49.0%)
Exploratory	0 (0.0%)	3 (25.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	0 (0.0%)	1 (11.1%)	5 (5.1%)
Was a Kaplan-Meier curve for OS presented in the main text?									
Yes	11 (91.7%)	9 (75.0%)	12 (80.0%)	7 (70.0%)	6 (85.7%)	17 (94.4%)	15 (100.0%)	9 (100.0%)	86 (87.8%)
No	1 (8.3%)	3 (25.0%)	2 (13.3%)	1 (10.0%)	1 (14.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (8.2%)
N/A[1]	0 (0.0%)	0 (0.0%)	1 (6.7%)	2 (20.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Were the numbers at risk included in the Kaplan-Meier curve?									
Yes	10 (83.3%)	9 (75.0%)	11 (73.3%)	6 (60.0%)	5 (71.4%)	16 (88.9%)	13 (86.7%)	9 (100.0%)	79 (80.6%)
No	1 (8.3%)	0 (0.0%)	1 (6.7%)	1 (10.0%)	1 (14.3%)	1 (5.6%)	2 (13.3%)	0 (0.0%)	7 (7.1%)
N/A	1 (8.3%)	3 (25.0%)	3 (20.0%)	3 (30.0%)	1 (14.3%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	12 (12.2%)
Were the number of patients censored or number of events included in the at risk table									
Yes	1 (8.3%)	0 (0.0%)	3 (20.0%)	0 (0.0%)	2 (28.6%)	1 (5.6%)	4 (26.7%)	2 (22.2%)	13 (13.3%)
No	9 (75.0%)	9 (75.0%)	8 (53.3%)	6 (60.0%)	3 (42.9%)	15 (83.3%)	9 (60.0%)	7 (77.8%)	66 (67.3%)
N/A	2 (16.7%)	3 (25.0%)	4 (26.7%)	4 (40.0%)	2 (28.6%)	2 (11.1%)	2 (13.3%)	0 (0.0%)	19 (19.4%)
Were confidence intervals included in the Kaplan-Meier curve?									
No	11 (91.7%)	9 (75.0%)	12 (80.0%)	7 (70.0%)	6 (85.7%)	17 (94.4%)	15 (100.0%)	9 (100.0%)	86 (87.8%)
N/A	1 (8.3%)	3 (25.0%)	3 (20.0%)	3 (30.0%)	1 (14.3%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	12 (12.2%)
Primary analysis									
Hazard Ratio and 95% CI from an unadjusted Cox model	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	1 (1.0%)
Hazard ratio and 95% CI from an adjusted Cox model	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	2 (2.0%)
Hazard ratio and P-value from a logrank test	4 (33.3%)	3 (25.0%)	0 (0.0%)	2 (20.0%)	1 (14.3%)	3 (16.7%)	3 (20.0%)	3 (33.3%)	19 (19.4%)

	Bowel Lower GI (n=12)	Breast (n=12)	Genito-urinary (n=15)	Gynaecological (n=10)	Haematological (n=7)	Lung (n=18)	Upper GI HPB (n=15)	Other (n=9)	Total (n=98)
Hazard ratio and P-value from a stratified logrank test.	1 (8.3%)	0 (0.0%)	1 (6.7%)	1 (10.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	2 (22.2%)	6 (6.1%)
Hazard ratio and p-value from an adjusted Cox model	2 (16.7%)	1 (8.3%)	4 (26.7%)	2 (20.0%)	1 (14.3%)	2 (11.1%)	3 (20.0%)	0 (0.0%)	15 (15.3%)
Hazard ratio and p-value from an unadjusted Cox model	0 (0.0%)	2 (16.7%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	1 (11.1%)	5 (5.1%)
Hazard ratio from a Cox model and p-value from a logrank test	2 (16.7%)	1 (8.3%)	3 (20.0%)	0 (0.0%)	0 (0.0%)	1 (5.6%)	2 (13.3%)	1 (11.1%)	10 (10.2%)
Hazard ratio from an adjusted Cox model and p-value from a logrank test.	1 (8.3%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	1 (5.6%)	2 (13.3%)	1 (11.1%)	6 (6.1%)
Hazard ratio from an adjusted Cox model and p-value from an stratified logrank test.	1 (8.3%)	2 (16.7%)	3 (20.0%)	0 (0.0%)	2 (28.6%)	2 (11.1%)	2 (13.3%)	1 (11.1%)	13 (13.3%)
Hazard ratio from an unadjusted cox model and p-value from a logrank test.	0 (0.0%)	2 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (11.1%)	1 (6.7%)	0 (0.0%)	5 (5.1%)
Hazard ratio from an unadjusted cox model and p-value from an stratified logrank test.	0 (0.0%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	2 (2.0%)
N/A[1]	0 (0.0%)	0 (0.0%)	1 (6.7%)	2 (20.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Other	1 (8.3%)	1 (8.3%)	1 (6.7%)	1 (10.0%)	3 (42.9%)	2 (11.1%)	1 (6.7%)	0 (0.0%)	10 (10.2%)
Were subsequent treatments mentioned in the paper?									
Yes	10 (83.3%)	6 (50.0%)	8 (53.3%)	4 (40.0%)	4 (57.1%)	14 (77.8%)	9 (60.0%)	4 (44.4%)	59 (60.2%)
No	2 (16.7%)	6 (50.0%)	7 (46.7%)	6 (60.0%)	3 (42.9%)	4 (22.2%)	6 (40.0%)	5 (55.6%)	39 (39.8%)
Was the percentage or number of participants who received later lines included in the main text or appendix?									
Yes	9 (75.0%)	4 (33.3%)	6 (40.0%)	2 (20.0%)	2 (28.6%)	14 (77.8%)	9 (60.0%)	4 (44.4%)	50 (51.0%)
No	1 (8.3%)	2 (16.7%)	1 (6.7%)	2 (20.0%)	2 (28.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (8.2%)
N/A	2 (16.7%)	6 (50.0%)	8 (53.3%)	6 (60.0%)	3 (42.9%)	4 (22.2%)	6 (40.0%)	5 (55.6%)	40 (40.8%)

	Bowel Lower GI (n=12)	Breast (n=12)	Genito-urinary (n=15)	Gynaecological (n=10)	Haematological (n=7)	Lung (n=18)	Upper GI HPB (n=15)	Other (n=9)	Total (n=98)
Was a breakdown or summary of the number of subsequent treatments (Two, Three, Four) included in the main text or appendix?									
Yes	1 (8.3%)	1 (8.3%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	4 (4.1%)
No	9 (75.0%)	5 (41.7%)	6 (40.0%)	4 (40.0%)	4 (57.1%)	13 (72.2%)	8 (53.3%)	4 (44.4%)	53 (54.1%)
N/A[2]	2 (16.7%)	6 (50.0%)	8 (53.3%)	6 (60.0%)	3 (42.9%)	5 (27.8%)	6 (40.0%)	5 (55.6%)	41 (41.8%)
Was a breakdown or summary of the type of subsequent treatments (Treatment 1, Treatment 2, Treatment 3) included in the main text or appendix?									
Yes	8 (66.7%)	4 (33.3%)	5 (33.3%)	1 (10.0%)	2 (28.6%)	13 (72.2%)	8 (53.3%)	4 (44.4%)	45 (45.9%)
No	2 (16.7%)	2 (16.7%)	2 (13.3%)	3 (30.0%)	2 (28.6%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	12 (12.2%)
N/A[2]	2 (16.7%)	6 (50.0%)	8 (53.3%)	6 (60.0%)	3 (42.9%)	5 (27.8%)	6 (40.0%)	5 (55.6%)	41 (41.8%)
Was additional analysis conducted to account for subsequent treatment lines?									
Yes	1 (8.3%)	3 (25.0%)	0 (0.0%)	0 (0.0%)	1 (14.3%)	0 (0.0%)	1 (6.7%)	1 (11.1%)	7 (7.1%)
No	9 (75.0%)	3 (25.0%)	7 (46.7%)	4 (40.0%)	3 (42.9%)	13 (72.2%)	8 (53.3%)	3 (33.3%)	50 (51.0%)
N/A[2]	2 (16.7%)	6 (50.0%)	8 (53.3%)	6 (60.0%)	3 (42.9%)	5 (27.8%)	6 (40.0%)	5 (55.6%)	41 (41.8%)
Were subsequent treatment lines mentioned in the discussion?									
Yes	7 (58.3%)	4 (33.3%)	5 (33.3%)	3 (30.0%)	3 (42.9%)	8 (44.4%)	6 (40.0%)	3 (33.3%)	39 (39.8%)
No	3 (25.0%)	2 (16.7%)	3 (20.0%)	1 (10.0%)	1 (14.3%)	6 (33.3%)	3 (20.0%)	1 (11.1%)	20 (20.4%)
N/A	2 (16.7%)	6 (50.0%)	7 (46.7%)	6 (60.0%)	3 (42.9%)	4 (22.2%)	6 (40.0%)	5 (55.6%)	39 (39.8%)
How were subsequent treatments mentioned in the discussion?									
The uptake of subsequent treatments is given as a reason for the discrepancy between OS and surrogate endpoints.	1 (8.3%)	1 (8.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (2.0%)
The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	0 (0.0%)	0 (0.0%)	2 (13.3%)	0 (0.0%)	1 (14.3%)	1 (5.6%)	1 (6.7%)	1 (11.1%)	6 (6.1%)
The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.	3 (25.0%)	1 (8.3%)	0 (0.0%)	1 (10.0%)	2 (28.6%)	4 (22.2%)	2 (13.3%)	1 (11.1%)	14 (14.3%)

	Bowel Lower GI (n=12)	Breast (n=12)	Genito-urinary (n=15)	Gynaecological (n=10)	Haematological (n=7)	Lung (n=18)	Upper GI HPB (n=15)	Other (n=9)	Total (n=98)
The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.	2 (16.7%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	0 (0.0%)	1 (8.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (5.6%)	1 (6.7%)	0 (0.0%)	3 (3.1%)
Randomisation / balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	0 (0.0%)	1 (8.3%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	1 (5.6%)	1 (6.7%)	0 (0.0%)	4 (4.1%)
None or limited options of subsequent treatment lines for patients are given as a reason as to why subsequent treatment lines will not have affected the OS results.	0 (0.0%)	0 (0.0%)	2 (13.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (2.0%)
Other	1 (8.3%)	0 (0.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	1 (6.7%)	1 (11.1%)	4 (4.1%)
N/A	5 (41.7%)	8 (66.7%)	10 (66.7%)	7 (70.0%)	4 (57.1%)	10 (55.6%)	9 (60.0%)	6 (66.7%)	59 (60.2%)

[1] Not-Applicable here refers to papers where OS is a pre-specified endpoint, but the data is not mature at the time of the publication.
[2] Note one paper stated no patients received subsequent treatments so is classed as N/A for these summaries.
N/A: Not Applicable; OS: Overall survival; GI: Gastrointestinal; HPB: Hepato-biliary

Appendix Table B.6: Summarised Cost-effectiveness Papers

ORN	Title	What analysis method was used for the Cost-Effectiveness analysis?	What OS estimate is used for the Cost-Effectiveness analysis?	Are the costs of subsequent treatment taken into account (in any form)	Was any additional analysis (e.g. RSPFTM, Two-stage etc.) presented to account for subsequent treatment lines or cross-over?
36	The cost-effectiveness of atezolizumab in first-line for metastatic triple negative breast cancer is heavily linked to PD-L1 level. [99]	Unclear	HR	No Adjustment for subsequent costs	No
49	Cost-effectiveness analysis of pembrolizumab monotherapy versus chemotherapy for previously untreated advanced non-small cell lung cancer. [100]	Partitioned Survival Model	KM	Adjustment for subsequent costs	No
53	Cost-effectiveness of Pembrolizumab as Second-line Therapy for the Treatment of Locally Advanced or Metastatic Urothelial Carcinoma in Sweden. [101]	Partitioned Survival Model	KM	Adjustment for subsequent costs	Yes
74	Cost-effectiveness of pembrolizumab for advanced non-small cell lung cancer patients with varying comorbidity burden. [102]	Microsimulation	HR	Unclear	No
84	Cost effectiveness of pembrolizumab vs chemotherapy as first-line treatment for metastatic NSCLC that expresses high levels of PD-L1 in Switzerland. [103]	Partitioned Survival Model	KM	Adjustment for subsequent costs	Yes
147	Economic evaluation of rituximab in addition to standard of care chemotherapy for adult patients with acute lymphoblastic leukemia. [104]	Partitioned Survival Model	KM	Adjustment for subsequent costs	No
173	Cost-effectiveness of Pazopanib Versus Sunitinib as First-line Treatment for Locally Advanced or Metastatic Renal Cell Carcinoma from an Italian National Health Service Perspective. [105]	Partitioned Survival Model	KM	Adjustment for subsequent costs	No
211	Exploratory Cost-Effectiveness Analysis of Response-Guided Neoadjuvant Chemotherapy for Hormone Positive Breast Cancer Patients. [106]	Markov Model	HR	Unclear	Yes
334	Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for KRAS wild-type patients with metastatic colorectal cancer. [107]	Partitioned Survival Model	Mean Survival	No Adjustment for subsequent costs	Yes

Appendix Table B.7: Summarised Methodology / Review papers

ORN	Title	Year of Publication	Methodology / Review Category	If Mentioned - Summary
8	Two decades of research toward the treatment of locally advanced and metastatic pancreatic cancer: Remarkable effort and limited gain. [108]	2021	Treatment overview / Analysis - mention	First and second treatment lines and the evidence used to make them available are discussed. However, the impact of subsequent treatment lines on that evidence is not mentioned.
14	Sunitinib malate for gastrointestinal stromal tumour in imatinib mesylate-resistant patients: recommendations and evidence. [109]	2010	Treatment overview / Analysis - no mention	N/A
32	A comparison of partitioned survival analysis and state transition multi-state modelling approaches using a case study in oncology. [110]	2020	Methodology paper - mention	Only considered in the sense that subsequent costs have been considered in various models.
35	The Landmark Series: Locally Advanced Pancreatic Cancer and Ablative Therapy Options. [Review] [111]	2021	Treatment overview / Analysis - no mention	N/A
42	Hazard ratio of progression-free survival is an excellent predictor of overall survival in phase III randomized controlled trials evaluating the first-line chemotherapy for extensive-disease small-cell lung cancer. [112]	2020	Methodology paper - mention	It is mentioned in passing that sensitivity to first-line chemotherapy predict response to later-line chemotherapy and post-progression survival. They don't mention that the use of surrogate endpoints is motivated by the effect of subsequent lines.
85	Propensity Score Weighting Using Overlap Weights: A New Method Applied to Regorafenib Clinical Data and a Cost-Effectiveness Analysis. [113]	2019	Methodology paper - no mention	N/A
115	Pembrolizumab monotherapy versus chemotherapy for treatment of advanced urothelial carcinoma with disease progression during or following platinum-containing chemotherapy. A Cochrane Rapid Review. [Review] [114]	2018	Treatment overview / Analysis - no mention	N/A
146	Sample size determination for the current strategy in oncology Phase 3 trials that tests progression-free survival and overall survival in a two-stage design framework. [115]	2018	Methodology paper - mention	Subsequent lines are one of the motivations for the publication and said to be one of the reasons PFS is becoming commonly the primary endpoint and OS the secondary endpoint, but appropriate sample size calculations are required when they are Co-primary
174	Checkpoint inhibition: new treatment options in urologic cancer. [Review] [116]	2017	Treatment overview / Analysis - mention	Discussing the patient treatment pathway as a whole and where different types of treatment should fall.

ORN	Title	Year of Publication	Methodology / Review Category	If Mentioned - Summary
175	First-line non-cytotoxic therapy in chemotherapy-naive patients with metastatic castration-resistant prostate cancer: a systematic review of 10 randomised clinical trials. [Review] [117]	2017	Treatment overview / Analysis - no mention	N/A
185	Cabozantinib in the treatment of advanced renal cell carcinoma in adults following prior vascular endothelial growth factor targeted therapy: clinical trial evidence and experience. [Review] [118]	2018	Treatment overview / Analysis - mention	Discussing the patient treatment pathway and where a certain drug should fall. However, the do mention that the availability of subsequent treatments may have affected the OS and PFS results they are considering.
209	Review of the Reporting of Survival Analyses within Randomised Controlled Trials and the Implications for Meta-Analysis. [Review] [119]	2016	Methodology paper - no mention	N/A
217	Evolution of Randomized Trials in Advanced/Metastatic Soft Tissue Sarcoma: End Point Selection, Surrogacy, and Quality of Reporting. [Review] [120]	2016	Methodology paper - mention	They use the issue of subsequent treatment lines as motivation for why they want to investigate surrogate endpoints.
219	First-Line Treatments for Poor-Prognosis Metastatic Renal Cell Carcinoma: Experts' Prescribing Practices and Systematic Literature Review. [Review] [121]	2016	Treatment overview / Analysis - no mention	N/A
228	Systemic therapy beyond first-line in advanced gastric cancer: An overview of the main randomized clinical trials. [Review] [122]	2016	Treatment overview / Analysis - no mention	N/A
229	Evaluating Intermittent Androgen-Deprivation Therapy Phase III Clinical Trials: The Devil Is in the Details. [123]	2016	Methodology paper - mention	Subsequent lines are mentioned in passing as affecting OS.
232	Why post-progression survival and post-relapse survival are not appropriate measures of efficacy in cancer randomized clinical trials. [124]	2015	Methodology paper - mention	The objective of the paper is to show that post-progression survival is not a way to account for subsequent treatments. Instead, causal methods should be used. The author suggests using a weighted cox-model to find the "per-protocol" effect and estimate what would have happened had everyone followed the protocol.
239	Systemic therapies for hepatocellular carcinoma. [Review] [125]	2015	Treatment overview / Analysis - no mention	N/A

ORN	Title	Year of Publication	Methodology / Review Category	If Mentioned - Summary
244	Adaptive truncated weighting for improving marginal structural model estimation of treatment effects informally censored by subsequent therapy. [126]	2015	Methodology paper - mention	The objective of the paper is to compare different ways to account for subsequent lines and suggest a new way to calculate the weights in a marginal structural model.
258	MET inhibitors for treatment of advanced hepatocellular carcinoma: A review. [Review] [127]	2015	Treatment overview / Analysis - no mention	N/A
274	Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. [Review] [128]	2015	Methodology paper - mention	The objective of the paper is to discuss the issues around subsequent treatment lines and whether PFS can be trusted. PFS2 and time to next therapy are suggested to be used alongside OS to provide a more comprehensive interpretation of the overall clinical benefit of an experimental treatment in a trial.
276	Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma. Correlation with overall survival, benchmarking and power analysis. [Review] [129]	2015	Methodology paper - mention	The objective of the paper is to consider whether PFS can be classed as a surrogate endpoint for OS because of the potential confounding of subsequent lines of therapy.
313	Systematic review of surgical resection vs radiofrequency ablation for hepatocellular carcinoma. [Review] [130]	2013	Treatment overview / Analysis - no mention	N/A
336	Systematic chemotherapy for inoperable, locally advanced, recurrent, or metastatic uterine leiomyosarcoma: a systematic review. [Review] [131]	2013	Treatment overview / Analysis - no mention	N/A
362	Potential pitfalls of crossover and thoughts on iniparib in triple-negative breast cancer. [Review] [132]	2011	Methodology paper - mention	The paper discusses the impact of patients crossing over from the experimental to the control and control to the experimental and how this can impact the overall survival results. It's a good argument for why we need treatment switching methods.
369	Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. [Review] [133]	2011	Methodology paper - mention	This talks about what the clinical benefit is in a number of scenarios. If the subsequent treatments are standard practice then the OS results even if non-sig show what would happen in the real world.
397	Current and future directions of clinical trials for ovarian cancer. [Review] [134]	2011	Treatment overview / Analysis - no mention	N/A

B.1 Appendix B References

1. Kanemitsu Y, Shimizu Y, Mizusawa J, et al. Hepatectomy Followed by mFOLFOX6 Versus Hepatectomy Alone for Liver-Only Metastatic Colorectal Cancer (JCOG0603): A Phase II or III Randomized Controlled Trial. *J Clin Oncol* 2021; 39: 3789-3799. DOI: https://dx.doi.org/10.1200/JCO.21.01032.
2. Chopra S, Gupta S, Kannan S, et al. Late Toxicity After Adjuvant Conventional Radiation Versus Image-Guided Intensity-Modulated Radiotherapy for Cervical Cancer (PARCER): A Randomized Controlled Trial. *J Clin Oncol* 2021; 39: 3682-3692. DOI: https://dx.doi.org/10.1200/JCO.20.02530.
3. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Five-Year Outcomes With Pembrolizumab Versus Chemotherapy for Metastatic Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score >= 50. *J Clin Oncol* 2021; 39: 2339-2349. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.21.00174.
4. Go SI, Lee SC, Bae WK, et al. Modified FOLFIRINOX versus S-1 as second-line chemotherapy in gemcitabine-failed metastatic pancreatic cancer patients: A randomised controlled trial (MPACA-3). *Eur J Cancer* 2021; 157: 21-30. Clinical Trial, Phase III, Comparative Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.ejca.2021.08.002.
5. Tang W, Ren L, Liu T, et al. Bevacizumab Plus mFOLFOX6 Versus mFOLFOX6 Alone as First-Line Treatment for RAS Mutant Unresectable Colorectal Liver-Limited Metastases: The BECOME Randomized Controlled Trial. *J Clin Oncol* 2020; 38: 3175-3184. Comparative Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.20.00174.
6. Shi Y, Lei K, Jia Y, et al. Bevacizumab biosimilar LY01008 compared with bevacizumab (Avastin) as first-line treatment for Chinese patients with unresectable, metastatic, or recurrent non-squamous non-small-cell lung cancer: A multicenter, randomized, double-blinded, phase III trial. *Cancer Commun (Lond)* 2021; 41: 889-903. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1002/cac2.12179.
7. Murthy V, Maitre P, Kannan S, et al. Prostate-Only Versus Whole-Pelvic Radiation Therapy in High-Risk and Very High-Risk Prostate Cancer (POP-RT): Outcomes From Phase III Randomized Controlled Trial. *J Clin Oncol* 2021; 39: 1234-1242. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.20.03282.
8. Lamarca A, Palmer DH, Wasan HS, et al. Second-line FOLFOX chemotherapy versus active symptom control for advanced biliary tract cancer (ABC-06): a phase 3, open-label, randomised, controlled trial. *Lancet Oncol* 2021; 22: 690-701. Clinical Trial, Phase III Comparative Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(21)00027-9.
9. Lin NU, Borges V, Anders C, et al. Intracranial Efficacy and Survival With Tucatinib Plus Trastuzumab and Capecitabine for Previously Treated HER2-Positive Breast Cancer With Brain Metastases in the HER2CLIMB Trial. *J Clin Oncol* 2020; 38: 2610-2619. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.20.00775.
10. Tomita Y, Naito S, Sassa N, et al. Sunitinib Versus Sorafenib as Initial Targeted Therapy for mCC-RCC With Favorable/Intermediate Risk: Multicenter Randomized Trial CROSS-J-RCC. *Clin Genitourin Cancer* 2020; 18: e374-e385. Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.clgc.2020.01.001.
11. Sezer A, Kilickap S, Gumus M, et al. Cemiplimab monotherapy for first-line treatment of advanced non-small-cell lung cancer with PD-L1 of at least 50%: a multicentre, open-label, global, phase 3, randomised, controlled trial. *Lancet* 2021; 397: 592-604. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial,

Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(21)00228-2.
12. Geng R, Wang G, Qiu L, et al. Metronomic capecitabine as maintenance treatment after first line induction with XELOX for metastatic colorectal cancer patients. *Medicine (Baltimore)* 2020; 99: e23719. Randomized Controlled Trial. DOI: https://dx.doi.org/10.1097/MD.0000000000023719.
13. Kumar SK, Jacobus SJ, Cohen AD, et al. Carfilzomib or bortezomib in combination with lenalidomide and dexamethasone for patients with newly diagnosed multiple myeloma without intention for immediate autologous stem-cell transplantation (ENDURANCE): a multicentre, open-label, phase 3, randomised, controlled trial. *The Lancet Oncology* 2020; 21: 1317-1330. DOI: 10.1016/s1470-2045(20)30452-6.
14. Malone S, Roy S, Eapen L, et al. Sequencing of Androgen-Deprivation Therapy With External-Beam Radiotherapy in Localized Prostate Cancer: A Phase III Randomized Controlled Trial. *J Clin Oncol* 2020; 38: 593-601. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.19.01904.
15. Mazzaferro V, Citterio D, Bhoori S, et al. Liver transplantation in hepatocellular carcinoma after tumour downstaging (XXL): a randomised, controlled, phase 2b/3 trial. *Lancet Oncol* 2020; 21: 947-956. Clinical Trial, Phase II, Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(20)30224-2.
16. Clamp AR, James EC, McNeish IA, et al. Weekly dose-dense chemotherapy in first-line epithelial ovarian, fallopian tube, or primary peritoneal carcinoma treatment (ICON8): primary progression free survival analysis results from a GCIG phase 3 randomised controlled trial. *Lancet* 2019; 394: 2084-2095. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(19)32259-7.
17. Hong YS, Kim SY, Lee JS, et al. Oxaliplatin-Based Adjuvant Chemotherapy for Rectal Cancer After Preoperative Chemoradiotherapy (ADORE): Long-Term Results of a Randomized Controlled Trial. *J Clin Oncol* 2019; 37: 3111-3123. Clinical Trial, Phase II, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.19.00016.
18. Minnaar CA, Kotzen JA, Ayeni OA, et al. The effect of modulated electro-hyperthermia on local disease control in HIV-positive and -negative cervical cancer women in South Africa: Early results from a phase III randomised controlled trial. *PLoS ONE* 2019; 14: e0217894. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0217894.
19. Xu L, Qi Q, Zhang Y, et al. Combination of icotinib and chemotherapy as first-line treatment for advanced lung adenocarcinoma in patients with sensitive EGFR mutations: A randomized controlled study. *Lung Cancer* 2019; 133: 23-31. Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/j.lungcan.2019.05.008.
20. Zhong WZ, Chen KN, Chen C, et al. Erlotinib Versus Gemcitabine Plus Cisplatin as Neoadjuvant Treatment of Stage IIIA-N2 EGFR-Mutant Non-Small-Cell Lung Cancer (EMERGING-CTONG 1103): A Randomized Phase II Study. *J Clin Oncol* 2019; 37: 2235-2245. Clinical Trial, Phase II, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.19.00075.
21. Rini BI, Powles T, Atkins MB, et al. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. *Lancet* 2019; 393: 2404-2415. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(19)30723-8.

22. Eng C, Kim TW, Bendell J, et al. Atezolizumab with or without cobimetinib versus regorafenib in previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase 3, randomised, controlled trial. *Lancet Oncol* 2019; 20: 849-861. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(19)30027-0.
23. MacManus M, Fisher R, Roos D, et al. Randomized Trial of Systemic Therapy After Involved-Field Radiotherapy in Patients With Early-Stage Follicular Lymphoma: TROG 99.03. *J Clin Oncol* 2018; 36: 2918-2925. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2018.77.9892.
24. Engel-Riedel W, Lowe J, Mattson P, et al. A randomized, controlled trial evaluating the efficacy and safety of BTH1677 in combination with bevacizumab, carboplatin, and paclitaxel in first-line treatment of advanced non-small cell lung cancer. *J Immunother Cancer* 2018; 6: 16. Clinical Trial, Phase II, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1186/s40425-018-0324-z.
25. Gupta S, Maheshwari A, Parab P, et al. Neoadjuvant Chemotherapy Followed by Radical Surgery Versus Concomitant Chemotherapy and Radiotherapy in Patients With Stage IB2, IIA, or IIB Squamous Cervical Cancer: A Randomized Controlled Trial. *J Clin Oncol* 2018; 36: 1548-1555. Clinical Trial, Phase III, Comparative Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2017.75.9985.
26. Aparicio T, Ghiringhelli F, Boige V, et al. Bevacizumab Maintenance Versus No Maintenance During Chemotherapy-Free Intervals in Metastatic Colorectal Cancer: A Randomized Phase III Trial (PRODIGE 9). *J Clin Oncol* 2018; 36: 674-681. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2017.75.2931.
27. Powles T, Duran I, van der Heijden MS, et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* 2018; 391: 748-757. Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(17)33297-X.
28. Mason MD, Clarke NW, James ND, et al. Adding Celecoxib With or Without Zoledronic Acid for Hormone-Naive Prostate Cancer: Long-Term Survival Results From an Adaptive, Multiarm, Multistage, Platform, Randomized Controlled Trial. *J Clin Oncol* 2017; 35: 1530-1541. Randomized Controlled Trial. DOI: https://dx.doi.org/10.1200/JCO.2016.69.0677.
29. Nokihara H, Lu S, Mok TSK, et al. Randomized controlled trial of S-1 versus docetaxel in patients with non-small-cell lung cancer previously treated with platinum-based chemotherapy (East Asia S-1 Trial in Lung Cancer). *Ann Oncol* 2017; 28: 2698-2706. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1093/annonc/mdx419.
30. Cunningham D, Stenning SP, Smyth EC, et al. Peri-operative chemotherapy with or without bevacizumab in operable oesophagogastric adenocarcinoma (UK Medical Research Council ST03): primary analysis results of a multicentre, open-label, randomised phase 2-3 trial. *Lancet Oncol* 2017; 18: 357-370. Clinical Trial, Phase II, Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(17)30043-8.
31. Joshi A, Patil V, Noronha V, et al. Results of a phase II randomized controlled clinical trial comparing efficacy of Cabazitaxel versus Docetaxel as second line or above therapy in recurrent head and neck cancer. *Oral Oncol* 2017; 75: 54-60. Clinical Trial, Phase II, Comparative Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/j.oraloncology.2017.10.018.

32. Miles D, Im YH, Fung A, et al. Effect of docetaxel duration on clinical outcomes: exploratory analysis of CLEOPATRA, a phase III randomized controlled trial. *Ann Oncol* 2017; 28: 2761-2767. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1093/annonc/mdx406.
33. Dimopoulos MA, Goldschmidt H, Niesvizky R, et al. Carfilzomib or bortezomib in relapsed or refractory multiple myeloma (ENDEAVOR): an interim overall survival analysis of an open-label, randomised, phase 3 trial. *Lancet Oncol* 2017; 18: 1327-1337. Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(17)30578-8.
34. Pignata S, Scambia G, Bologna A, et al. Randomized Controlled Trial Testing the Efficacy of Platinum-Free Interval Prolongation in Advanced Ovarian Cancer: The MITO-8, MaNGO, BGOG-Ov1, AGO-Ovar2.16, ENGOT-Ov1, GCIG Study. *J Clin Oncol* 2017; 35: 3347-3353. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1200/JCO.2017.73.4293.
35. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet* 2017; 389: 255-265. DOI: 10.1016/s0140-6736(16)32517-x.
36. Belani CP, Chakraborty BC, Modi RI and Khamar BM. A randomized trial of TLR-2 agonist CADI-05 targeting desmocollin-3 for advanced non-small-cell lung cancer. *Ann Oncol* 2017; 28: 298-304. Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1093/annonc/mdw608.
37. Cheng Y, Fan Y, Liu X, et al. Randomized controlled trial of lobaplatin plus etoposide vs. cisplatin plus etoposide as first-line therapy in patients with extensive-stage small cell lung cancer. *Oncol* 2019; 17: 4701-4709. DOI: https://dx.doi.org/10.3892/ol.2019.10125.
38. Zhu Y, Zhang W, Li Q, et al. A Phase II Randomized Controlled Trial: Definitive Concurrent Chemoradiotherapy with Docetaxel Plus Cisplatin versus 5-Fluorouracil plus Cisplatin in Patients with Oesophageal Squamous Cell Carcinoma. *J* 2017; 8: 3657-3666. DOI: https://dx.doi.org/10.7150/jca.20053.
39. James ND, Sydes MR, Clarke NW, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet* 2016; 387: 1163-1177. Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(15)01037-5.
40. Hodi FS, Chesney J, Pavlick AC, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol* 2016; 17: 1558-1568. Clinical Trial, Phase II, Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S1470-2045(16)30366-7.
41. Hasegawa K, Saiura A, Takayama T, et al. Adjuvant Oral Uracil-Tegafur with Leucovorin for Colorectal Cancer Liver Metastases: A Randomized Controlled Trial. *PLoS ONE* 2016; 11: e0162400. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1371/journal.pone.0162400.
42. James N, Pirrie S, Pope A, et al. TRAPEZE: a randomised controlled trial of the clinical effectiveness and cost-effectiveness of chemotherapy with zoledronic acid, strontium-89, or both, in men with bony metastatic castration-refractory prostate cancer. *Health Technol Assess* 2016; 20: 1-288. Clinical Trial, Phase II, Clinical Trial, Phase III, Randomized Controlled Trial. DOI: https://dx.doi.org/10.3310/hta20530.
43. Carrie C, Hasbini A, de Laroche G, et al. Salvage radiotherapy with or without short-term hormone therapy for rising prostate-specific antigen concentration after radical prostatectomy (GETUG-AFU 16): a randomised, multicentre, open-label phase 3 trial. *Lancet Oncol* 2016; 17: 747-756. Clinical Trial, Phase III, Comparative Study,

Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S1470-2045(16)00111-X.
44. Park K, Tan EH, O'Byrne K, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. *Lancet Oncol* 2016; 17: 577-589. Clinical Trial, Phase II, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S1470-2045(16)30033-X.
45. Fehrenbacher L, Spira A, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* 2016; 387: 1837-1846. Clinical Trial, Phase II Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(16)00587-0.
46. Fujitani K, Yang HK, Mizusawa J, et al. Gastrectomy plus chemotherapy versus chemotherapy alone for advanced gastric cancer with a single non-curable factor (REGATTA): a phase 3, randomised controlled trial. *Lancet Oncol* 2016; 17: 309-318. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00553-7.
47. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2016; 387: 1540-1550. Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(15)01281-7.
48. Oza AM, Cook AD, Pfisterer J, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. *Lancet Oncol* 2015; 16: 928-936. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00086-8.
49. Bielack SS, Smeland S, Whelan JS, et al. Methotrexate, Doxorubicin, and Cisplatin (MAP) Plus Maintenance Pegylated Interferon Alfa-2b Versus MAP Alone in Patients With Resectable High-Grade Osteosarcoma and Good Histologic Response to Preoperative MAP: First Results of the EURAMOS-1 Good Response Randomized Controlled Trial. *J Clin Oncol* 2015; 33: 2279-2287. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't Webcast. DOI: https://dx.doi.org/10.1200/JCO.2014.60.0734.
50. Kehoe S, Hook J, Nankivell M, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet* 2015; 386: 249-257. Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(14)62223-6.
51. Mason MD, Parulekar WR, Sydes MR, et al. Final Report of the Intergroup Randomized Study of Combined Androgen-Deprivation Therapy Plus Radiotherapy Versus Androgen-Deprivation Therapy Alone in Locally Advanced Prostate Cancer. *J Clin Oncol* 2015; 33: 2143-2150. Randomized Controlled Trial, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2014.57.7510.
52. Badwe R, Hawaldar R, Nair N, et al. Locoregional treatment versus no treatment of the primary tumour in metastatic breast cancer: an open-label randomised controlled trial. *Lancet Oncol* 2015; 16: 1380-1388. Comparative Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00135-7.
53. Bear HD, Tang G, Rastogi P, et al. Neoadjuvant plus adjuvant bevacizumab in early breast cancer (NSABP B-40 [NRG Oncology]): secondary outcomes of a phase 3, randomised controlled trial. *Lancet Oncol* 2015; 16: 1037-1048. Clinical Trial, Phase III,

Randomized Controlled Trial, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00041-8.
54. Shapiro J, van Lanschot JJB, Hulshof M, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol* 2015; 16: 1090-1098. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00040-6.
55. Henderson MA, Burmeister BH, Ainslie J, et al. Adjuvant lymph-node field radiotherapy versus observation only in patients with melanoma at high risk of further lymph-node field relapse after lymphadenectomy (ANZMTG 01.02/TROG 02.01): 6-year follow-up of a phase 3, randomised controlled trial. *Lancet Oncol* 2015; 16: 1049-1060. Clinical Trial, Phase III, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00187-4.
56. Soria JC, Felip E, Cobo M, et al. Afatinib versus erlotinib as second-line treatment of patients with advanced squamous cell carcinoma of the lung (LUX-Lung 8): an open-label randomised controlled phase 3 trial. *Lancet Oncol* 2015; 16: 897-907. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)00006-6.
57. Zapatero A, Guerrero A, Maldonado X, et al. High-dose radiotherapy with short-term or long-term androgen deprivation in localised prostate cancer (DART01/05 GICOR): a randomised, controlled, phase 3 trial. *Lancet Oncol* 2015; 16: 320-327. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(15)70045-8.
58. Kunkler IH, Williams LJ, Jack WJ, et al. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. *Lancet Oncol* 2015; 16: 266-273. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)71221-5.
59. Bartelink H, Maingon P, Poortmans P, et al. Whole-breast irradiation with or without a boost for patients treated with breast-conserving surgery for early breast cancer: 20-year follow-up of a randomised phase 3 trial. *Lancet Oncol* 2015; 16: 47-56. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't DOI: https://dx.doi.org/10.1016/S1470-2045(14)71156-8.
60. Slotman BJ, van Tinteren H, Praag JO, et al. Use of thoracic radiotherapy for extensive stage small-cell lung cancer: a phase 3 randomised controlled trial. *Lancet* 2015; 385: 36-42. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(14)61085-0.
61. Middleton G, Silcocks P, Cox T, et al. Gemcitabine and capecitabine with or without telomerase peptide vaccine GV1001 in patients with locally advanced or metastatic pancreatic cancer (TeloVac): an open-label, randomised, phase 3 trial. *Lancet Oncol* 2014; 15: 829-840. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)70236-0.
62. Vora A, Goulden N, Mitchell C, et al. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. *Lancet Oncol* 2014; 15: 809-818. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)70243-8.
63. Corrie PG, Marshall A, Dunn JA, et al. Adjuvant bevacizumab in patients with melanoma at high risk of recurrence (AVAST-M): preplanned interim results from a multicentre, open-label, randomised controlled phase 3 study. *Lancet Oncol* 2014; 15: 620-630. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial,

Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)70110-X.
64. Dearnaley DP, Jovic G, Syndikus I, et al. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: long-term results from the MRC RT01 randomised controlled trial. *Lancet Oncol* 2014; 15: 464-473. Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(14)70040-3.
65. Judson I, Verweij J, Gelderblom H, et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. *Lancet Oncol* 2014; 15: 415-423. Clinical Trial, Phase III Comparative Study Randomized Controlled Trial Research Support, Non-U.S. Gov't DOI: https://dx.doi.org/10.1016/S1470-2045(14)70063-4.
66. Zuliani AC, Esteves SC, Teixeira LC, et al. Concomitant cisplatin plus radiotherapy and high-dose-rate brachytherapy versus radiotherapy alone for stage IIIB epidermoid cervical cancer: a randomized controlled trial. *J Clin Oncol* 2014; 32: 542-547. Randomized Controlled Trial. DOI: https://dx.doi.org/10.1200/JCO.2013.50.1205.
67. Reck M, Kaiser R, Mellemgaard A, et al. Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial. *Lancet Oncol* 2014; 15: 143-155. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(13)70586-2.
68. Ford HE, Marshall A, Bridgewater JA, et al. Docetaxel versus active symptom control for refractory oesophagogastric adenocarcinoma (COUGAR-02): an open-label, phase 3 randomised controlled trial. *Lancet Oncol* 2014; 15: 78-86. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(13)70549-7.
69. Poplin E, Wasan H, Rolfe L, et al. Randomized, multicenter, phase II study of CO-101 versus gemcitabine in patients with metastatic pancreatic ductal adenocarcinoma: including a prospective evaluation of the role of hENT1 in gemcitabine or CO-101 sensitivity. *J Clin Oncol* 2013; 31: 4453-4461. Clinical Trial, Phase II, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2013.51.0826.
70. Goldhirsch A, Gelber RD, Piccart-Gebhart MJ, et al. 2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. *Lancet* 2013; 382: 1021-1028. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(13)61094-6.
71. Yamamoto Y, Ishikawa T, Hozumi Y, et al. Randomized controlled trial of toremifene 120 mg compared with exemestane 25 mg after prior treatment with a non-steroidal aromatase inhibitor in postmenopausal women with hormone receptor-positive metastatic breast cancer. *BMC Cancer* 2013; 13: 239. Comparative Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1186/1471-2407-13-239.
72. Crosby T, Hurt CN, Falk S, et al. Chemoradiotherapy with or without cetuximab in patients with oesophageal cancer (SCOPE1): a multicentre, phase 2/3 randomised trial. *Lancet Oncol* 2013; 14: 627-637. Clinical Trial, Phase II, Clinical Trial, Phase III, Comparative Study, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(13)70136-0.
73. Veronesi U, Orecchia R, Maisonneuve P, et al. Intraoperative radiotherapy versus external radiotherapy for early breast cancer (ELIOT): a randomised controlled equivalence trial. *Lancet Oncol* 2013; 14: 1269-1277. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(13)70497-2.
74. Katsumata N, Yasuda M, Isonishi S, et al. Long-term results of dose-dense paclitaxel and carboplatin versus conventional paclitaxel and carboplatin for treatment of advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer (JGOG

3016): a randomised, controlled, open-label trial. *Lancet Oncol* 2013; 14: 1020-1026. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(13)70363-2.
75. Garassino MC, Martelli O, Broggini M, et al. Erlotinib versus docetaxel as second-line treatment of patients with advanced non-small-cell lung cancer and wild-type EGFR tumours (TAILOR): a randomised controlled trial. *Lancet Oncol* 2013; 14: 981-988. Comparative Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(13)70310-3.
76. Ye LC, Liu TS, Ren L, et al. Randomized controlled trial of cetuximab plus chemotherapy for patients with KRAS wild-type unresectable colorectal liver-limited metastases. *J Clin Oncol* 2013; 31: 1931-1938. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2012.44.8308.
77. Penichoux J, Michiels S, Bouche O, et al. Taking into account successive treatment lines in the analysis of a colorectal cancer randomised trial. *Eur J Cancer* 2013; 49: 1882-1888. Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.ejca.2013.02.006.
78. Peng ZW, Zhang YJ, Chen MS, et al. Radiofrequency ablation with or without transcatheter arterial chemoembolization in the treatment of hepatocellular carcinoma: a prospective randomized trial. *J Clin Oncol* 2013; 31: 426-432. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2012.42.9936.
79. Flotten O, Gronberg BH, Bremnes R, et al. Vinorelbine and gemcitabine vs vinorelbine and carboplatin as first-line treatment of advanced NSCLC. A phase III randomised controlled trial by the Norwegian Lung Cancer Study Group. *Br J Cancer* 2012; 107: 442-447. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1038/bjc.2012.284.
80. James ND, Sydes MR, Mason MD, et al. Celecoxib plus hormone therapy versus hormone therapy alone for hormone-sensitive prostate cancer: first results from the STAMPEDE multiarm, multistage, randomised controlled trial. *Lancet Oncol* 2012; 13: 549-558. Clinical Trial, Phase II, Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(12)70088-8.
81. de Gramont A, Van Cutsem E, Schmoll HJ, et al. Bevacizumab plus oxaliplatin-based chemotherapy as adjuvant treatment for colon cancer (AVANT): a phase 3 randomised controlled trial. *Lancet Oncol* 2012; 13: 1225-1233. Clinical Trial, Phase III, Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(12)70509-0.
82. Woll PJ, Reichardt P, Le Cesne A, et al. Adjuvant chemotherapy with doxorubicin, ifosfamide, and lenograstim for resected soft-tissue sarcoma (EORTC 62931): a multicentre randomised controlled trial. *Lancet Oncol* 2012; 13: 1045-1054. Multicenter Study Randomized Controlled Trial Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(12)70346-7.
83. Burmeister BH, Henderson MA, Ainslie J, et al. Adjuvant radiotherapy versus observation alone for patients at risk of lymph-node field relapse after therapeutic lymphadenectomy for melanoma: a randomised trial. *Lancet Oncol* 2012; 13: 589-597. Multicenter Study, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(12)70138-9.
84. Bang YJ, Kim YW, Yang HK, et al. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. *Lancet* 2012; 379: 315-321. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(11)61873-4.
85. Adams RA, Meade AM, Seymour MT, et al. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet*

Oncol 2011; 12: 642-653. Clinical Trial Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(11)70102-4.
86. Maughan TS, Adams RA, Smith CG, et al. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet* 2011; 377: 2103-2114. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(11)60613-2.
87. Morgan GJ, Davies FE, Gregory WM, et al. First-line treatment with zoledronic acid as compared with clodronic acid in multiple myeloma (MRC Myeloma IX): a randomised controlled trial. *Lancet* 2010; 376: 1989-1999. Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(10)62051-X.
88. Cuzick J, Sestak I, Baum M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol* 2010; 11: 1135-1141. Clinical Trial, Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(10)70257-6.
89. Rustin GJ, van der Burg ME, Griffin CL, et al. Early versus delayed treatment of relapsed ovarian cancer (MRC OV05/EORTC 55955): a randomised trial. *Lancet* 2010; 376: 1155-1163. Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(10)61268-8.
90. Issels RD, Lindner LH, Verweij J, et al. Neo-adjuvant chemotherapy alone or with regional hyperthermia for localised high-risk soft-tissue sarcoma: a randomised phase 3 multicentre study. *Lancet Oncol* 2010; 11: 561-570. Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(10)70071-1.
91. Zietman AL, Bae K, Slater JD, et al. Randomized trial comparing conventional-dose with high-dose conformal radiation therapy in early-stage adenocarcinoma of the prostate: long-term results from proton radiation oncology group/american college of radiology 95-09. *J Clin Oncol* 2010; 28: 1106-1111. Clinical Trial, Phase I, Clinical Trial, Phase II, Comparative Study, Randomized Controlled Trial, Research Support, N.I.H., Extramural. DOI: https://dx.doi.org/10.1200/JCO.2009.25.8475.
92. Kantoff PW, Schuetz TJ, Blumenstein BA, et al. Overall survival analysis of a phase II randomized controlled trial of a Poxviral-based PSA-targeted immunotherapy in metastatic castration-resistant prostate cancer. *J Clin Oncol* 2010; 28: 1099-1105. Clinical Trial, Phase II, Multicenter Study, Randomized Controlled Trial, Research Support, N.I.H., Extramural. DOI: https://dx.doi.org/10.1200/JCO.2009.25.0597.
93. Rajkumar SV, Jacobus S, Callander NS, et al. Lenalidomide plus high-dose dexamethasone versus lenalidomide plus low-dose dexamethasone as initial therapy for newly diagnosed multiple myeloma: an open-label randomised controlled trial. *Lancet Oncol* 2010; 11: 29-37. Comparative Study Multicenter Study Randomized Controlled Trial Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. DOI: https://dx.doi.org/10.1016/S1470-2045(09)70284-0.
94. Sutton L, Chevret S, Tournilhac O, et al. Autologous stem cell transplantation as a first-line treatment strategy for chronic lymphocytic leukemia: a multicenter, randomized, controlled trial from the SFGM-TC and GFLLC. *Blood* 2011; 117: 6109-6119. Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1182/blood-2010-11-317073.
95. Badwe R, Hawaldar R, Parmar V, et al. Single-injection depot progesterone before surgery and survival in women with operable breast cancer: a randomized controlled trial. *J Clin Oncol* 2011; 29: 2845-2851. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.2010.33.0738.

96. Gianni L, Dafni U, Gelber RD, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol* 2011; 12: 236-244. Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(11)70033-X.
97. Sharma A, Dwary AD, Mohanti BK, et al. Best supportive care compared with chemotherapy for unresectable gall bladder cancer: a randomized controlled study. *J Clin Oncol* 2010; 28: 4581-4586. Comparative Study, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1200/JCO.2010.29.3605.
98. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* 2010; 376: 687-697. Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(10)61121-X.
99. Giuliani J, Mantoan B and Bonetti A. The cost-effectiveness of atezolizumab in first-line for metastatic triple negative breast cancer is heavily linked to PD-L1 level. *J Oncol Pharm Pract* 2021; 27: 1245-1247. DOI: https://dx.doi.org/10.1177/10781552211019462.
100. Aziz MIA, Tan LE, Tan WHG, et al. Cost-effectiveness analysis of pembrolizumab monotherapy versus chemotherapy for previously untreated advanced non-small cell lung cancer. *J Med Econ* 2020; 23: 952-960. Randomized Controlled Trial. DOI: https://dx.doi.org/10.1080/13696998.2020.1775620.
101. Srivastava T, Prabhu VS, Li H, et al. Cost-effectiveness of Pembrolizumab as Second-line Therapy for the Treatment of Locally Advanced or Metastatic Urothelial Carcinoma in Sweden. *Eur Urol Oncol* 2020; 3: 663-670. Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.euo.2018.09.012.
102. Criss SD, Palazzo L, Watson TR, et al. Cost-effectiveness of pembrolizumab for advanced non-small cell lung cancer patients with varying comorbidity burden. *PLoS ONE* 2020; 15: e0228288. Research Support, N.I.H., Extramural. DOI: https://dx.doi.org/10.1371/journal.pone.0228288.
103. Bhadhuri A, Insinga R, Guggisberg P, et al. Cost effectiveness of pembrolizumab vs chemotherapy as first-line treatment for metastatic NSCLC that expresses high levels of PD-L1 in Switzerland. *Swiss Med Wkly* 2019; 149: w20170. Comparative Study. DOI: https://dx.doi.org/10.4414/smw.2019.20170.
104. Nam J, Milenkovski R, Yunger S, et al. Economic evaluation of rituximab in addition to standard of care chemotherapy for adult patients with acute lymphoblastic leukemia. *J Med Econ* 2018; 21: 47-59. Evaluation Study. DOI: https://dx.doi.org/10.1080/13696998.2017.1372230.
105. Capri S, Porta C and Delea TE. Cost-effectiveness of Pazopanib Versus Sunitinib as First-line Treatment for Locally Advanced or Metastatic Renal Cell Carcinoma from an Italian National Health Service Perspective. *Clin Ther* 2017; 39: 567-580.e562. Comparative Study. DOI: https://dx.doi.org/10.1016/j.clinthera.2017.01.017.
106. Miquel-Cases A, Retel VP, Lederer B, et al. Exploratory Cost-Effectiveness Analysis of Response-Guided Neoadjuvant Chemotherapy for Hormone Positive Breast Cancer Patients. *PLoS ONE* 2016; 11: e0154386. Randomized Controlled Trial, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0154386.
107. Hoyle M, Peters J, Crathorne L, et al. Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for KRAS wild-type patients with metastatic colorectal cancer. *Value Health* 2013; 16: 288-296. Comparative Study, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.jval.2012.11.001.

108. Yeh C and Bates SE. Two decades of research toward the treatment of locally advanced and metastatic pancreatic cancer: Remarkable effort and limited gain. *Semin Oncol* 2021; 48: 34-46. DOI: https://dx.doi.org/10.1053/j.seminoncol.2021.01.001.
109. Younus J, Verma S, Franek J, et al. Sunitinib malate for gastrointestinal stromal tumour in imatinib mesylate-resistant patients: recommendations and evidence. *Curr* 2010; 17: 4-10.
110. Cranmer H, Shields GE and 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: 1176-1185. DOI: https://dx.doi.org/10.1080/13696998.2020.1796360.
111. White RR, Murphy JD and Martin RCG. The Landmark Series: Locally Advanced Pancreatic Cancer and Ablative Therapy Options. *Ann Surg Oncol* 2021; 28: 4173-4180. Review. DOI: https://dx.doi.org/10.1245/s10434-021-09662-z.
112. Chen H, Horita N, Ito K, et al. Hazard ratio of progression-free survival is an excellent predictor of overall survival in phase III randomized controlled trials evaluating the first-line chemotherapy for extensive-disease small-cell lung cancer. *Transl* 2020; 9: 1333-1342. DOI: https://dx.doi.org/10.21037/tlcr-20-377.
113. Mlcoch T, Hrnciarova T, Tuzil J, et al. Propensity Score Weighting Using Overlap Weights: A New Method Applied to Regorafenib Clinical Data and a Cost-Effectiveness Analysis. *Value Health* 2019; 22: 1370-1377. Evaluation Study, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.jval.2019.06.010.
114. Narayan V, Kahlmeyer A, Dahm P, et al. Pembrolizumab monotherapy versus chemotherapy for treatment of advanced urothelial carcinoma with disease progression during or following platinum-containing chemotherapy. A Cochrane Rapid Review. *Cochrane Database Syst Rev* 2018; 7: CD012838. Research Support, Non-U.S. Gov't, Review, Systematic Review. DOI: https://dx.doi.org/10.1002/14651858.CD012838.pub2.
115. Nomura S, Hirakawa A and Hamada C. Sample size determination for the current strategy in oncology Phase 3 trials that tests progression-free survival and overall survival in a two-stage design framework. *J Biopharm Stat* 2018; 28: 589-611. Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1080/10543406.2017.1372775.
116. De Maeseneer DJ, Delafontaine B and Rottey S. Checkpoint inhibition: new treatment options in urologic cancer. *Acta Clin Belg* 2017; 72: 24-28. Review. DOI: https://dx.doi.org/10.1080/17843286.2016.1260890.
117. Poorthuis MHF, Vernooij RWM, van Moorselaar RJA and de Reijke TM. First-line non-cytotoxic therapy in chemotherapy-naive patients with metastatic castration-resistant prostate cancer: a systematic review of 10 randomised clinical trials. *BJU Int* 2017; 119: 831-845. Review, Systematic Review. DOI: https://dx.doi.org/10.1111/bju.13764.
118. Osanto S and van der Hulle T. Cabozantinib in the treatment of advanced renal cell carcinoma in adults following prior vascular endothelial growth factor targeted therapy: clinical trial evidence and experience. *Ther Adv Urol* 2018; 10: 109-123. Review. DOI: https://dx.doi.org/10.1177/1756287217748867.
119. Batson S, Greenall G and Hudson P. Review of the Reporting of Survival Analyses within Randomised Controlled Trials and the Implications for Meta-Analysis. *PLoS ONE* 2016; 11: e0154870. Review, Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0154870.
120. Zer A, Prince RM, Amir E and Abdul Razak A. Evolution of Randomized Trials in Advanced/Metastatic Soft Tissue Sarcoma: End Point Selection, Surrogacy, and Quality of Reporting. *J Clin Oncol* 2016; 34: 1469-1475. Review, Systematic Review. DOI: https://dx.doi.org/10.1200/JCO.2015.64.3437.
121. Le Saux O, Freyer G and Negrier S. First-Line Treatments for Poor-Prognosis Metastatic Renal Cell Carcinoma: Experts' Prescribing Practices and Systematic Literature Review. *Clin Drug Invest* 2016; 36: 389-399. Review, Systematic Review. DOI: https://dx.doi.org/10.1007/s40261-016-0384-0.

122. Galdy S, Cella CA, Spada F, et al. Systemic therapy beyond first-line in advanced gastric cancer: An overview of the main randomized clinical trials. *Crit Rev Oncol Hematol* 2016; 99: 1-12. Research Support, Non-U.S. Gov't, Review. DOI: https://dx.doi.org/10.1016/j.critrevonc.2015.09.004.
123. Hussain M, Tangen C, Higano C, et al. Evaluating Intermittent Androgen-Deprivation Therapy Phase III Clinical Trials: The Devil Is in the Details. *J Clin Oncol* 2016; 34: 280-285. Clinical Trial, Phase III, Randomized Controlled Trial. DOI: https://dx.doi.org/10.1200/JCO.2015.62.8065.
124. Garcia-Albeniz X, Maurel J and Hernan MA. Why post-progression survival and post-relapse survival are not appropriate measures of efficacy in cancer randomized clinical trials. *Int J Cancer* 2015; 136: 2444-2447. Research Support, N.I.H., Extramural. DOI: https://dx.doi.org/10.1002/ijc.29278.
125. Ge S and Huang D. Systemic therapies for hepatocellular carcinoma. *Drug discov* 2015; 9: 352-362. Review. DOI: https://dx.doi.org/10.5582/ddt.2015.01047.
126. Bai X, Liu J, Li L and Faries D. Adaptive truncated weighting for improving marginal structural model estimation of treatment effects informally censored by subsequent therapy. *Pharm* 2015; 14: 448-454. DOI: https://dx.doi.org/10.1002/pst.1719.
127. Qi XS, Guo XZ, Han GH, et al. MET inhibitors for treatment of advanced hepatocellular carcinoma: A review. *World J Gastroenterol* 2015; 21: 5445-5453. Review. DOI: https://dx.doi.org/10.3748/wjg.v21.i18.5445.
128. Matulonis UA, Oza AM, Ho TW and Ledermann JA. Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. *Cancer* 2015; 121: 1737-1746. Research Support, Non-U.S. Gov't, Review. DOI: https://dx.doi.org/10.1002/cncr.29082.
129. Bria E, Massari F, Maines F, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma. Correlation with overall survival, benchmarking and power analysis. *Crit Rev Oncol Hematol* 2015; 93: 50-59. Research Support, Non-U.S. Gov't, Review. DOI: https://dx.doi.org/10.1016/j.critrevonc.2014.08.001.
130. Cucchetti A, Piscaglia F, Cescon M, et al. Systematic review of surgical resection vs radiofrequency ablation for hepatocellular carcinoma. *World J Gastroenterol* 2013; 19: 4106-4118. Research Support, Non-U.S. Gov't, Review, Systematic Review. DOI: https://dx.doi.org/10.3748/wjg.v19.i26.4106.
131. Gupta AA, Yao X, Verma S, et al. Systematic chemotherapy for inoperable, locally advanced, recurrent, or metastatic uterine leiomyosarcoma: a systematic review. *Clin Oncol (R Coll Radiol)* 2013; 25: 346-355. Research Support, Non-U.S. Gov't Review Systematic Review. DOI: https://dx.doi.org/10.1016/j.clon.2012.11.008.
132. Fojo T, Amiri-Kordestani L and Bates SE. Potential pitfalls of crossover and thoughts on iniparib in triple-negative breast cancer. *J Natl Cancer Inst* 2011; 103: 1738-1740. Research Support, N.I.H., Intramural, Review. DOI: https://dx.doi.org/10.1093/jnci/djr386.
133. Korn EL, Freidlin B and Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol* 2011; 29: 2439-2442. Review. DOI: https://dx.doi.org/10.1200/JCO.2011.34.6056.
134. Gardner GJ and Jewell EL. Current and future directions of clinical trials for ovarian cancer. *Cancer Control* 2011; 18: 44-51. Review.

Appendix C Chapter 3 Supplementary Material

How should overall survival be analysed in randomised clinical trials in cancer if participants receive subsequent treatment lines? A stakeholder consultation - Supplementary Material

Appendix Table C.1: Checklist for Reporting Results of Internet E-Surveys (CHERRIES)

Checklist Item	***Explanation***	***Page Number***
Describe survey design	Describe target population, sample frame. Is the sample a convenience sample? (In “open” surveys this is most likely.)	51
IRB approval	Mention whether the study has been approved by an IRB.	53
Informed consent	Describe the informed consent process. Where were the participants told the length of time of the survey, which data were stored and where and for how long, who the investigator was, and the purpose of the study?	53
Data protection	If any personal information was collected or stored, describe what mechanisms were used to protect unauthorized access.	53
Development and testing	State how the survey was developed, including whether the usability and technical functionality of the electronic questionnaire had been tested before fielding the questionnaire.	51-53
Open survey versus closed survey	An “open survey” is a survey open for each visitor of a site, while a closed survey is only open to a sample which the investigator knows (password-protected survey).	56
Contact mode	Indicate whether or not the initial contact with the potential participants was made on the Internet. (Investigators may also send out questionnaires by mail and allow for Web-based data entry.)	56
Advertising the survey	How/where was the survey announced or advertised? Some examples are offline media (newspapers), or online (mailing lists – If yes, which ones?) or banner ads (Where were these banner ads posted and what did they look like?). It is important to know the wording of the announcement as it will heavily influence who chooses to participate. Ideally the survey announcement should be published as an appendix.	56
Web/E-mail	State the type of e-survey (eg, one posted on a Web site, or one sent out through e-mail). If it is an e-mail survey, were the responses entered manually into a database, or was there an automatic method for capturing responses?	51
Context	Describe the Web site (for mailing list/newsgroup) in which the survey was posted. What is the Web site about, who is visiting it, what are visitors normally looking for? Discuss to what degree the content of the	N/A

	Web site could pre-select the sample or influence the results. For example, a survey about vaccination on a anti-immunization Web site will have different results from a Web survey conducted on a government Web site	
Mandatory/voluntary	Was it a mandatory survey to be filled in by every visitor who wanted to enter the Web site, or was it a voluntary survey?	56
Incentives	Were any incentives offered (eg, monetary, prizes, or non-monetary incentives such as an offer to provide the survey results)?	53
Time/Date	In what timeframe were the data collected?	58
Randomization of items or questionnaires	To prevent biases items can be randomized or alternated.	52
Adaptive questioning	Use adaptive questioning (certain items, or only conditionally displayed based on responses to other items) to reduce number and complexity of the questions.	52
Number of Items	What was the number of questionnaire items per page? The number of items is an important factor for the completion rate.	52
Number of screens (pages)	Over how many pages was the questionnaire distributed? The number of items is an important factor for the completion rate.	52
Completeness check	It is technically possible to do consistency or completeness checks before the questionnaire is submitted. Was this done, and if “yes”, how (usually JAVAScript)? An alternative is to check for completeness after the questionnaire has been submitted (and highlight mandatory items). If this has been done, it should be reported. All items should provide a non-response option such as “not applicable” or “rather not say”, and selection of one response option should be enforced.	52
Review step	State whether respondents were able to review and change their answers (eg, through a Back button or a Review step which displays a summary of the responses and asks the respondents if they are correct).	52
Unique site visitor	If you provide view rates or participation rates, you need to define how you determined a unique visitor. There are different techniques available, based on IP addresses or cookies or both.	56
View rate (Ratio of unique survey visitors/unique site visitors)	Requires counting unique visitors to the first page of the survey, divided by the number of unique site visitors (not page views!). It is not unusual to have view rates of less than 0.1 % if the survey is voluntary.	N/A – Unable to obtain unique site visitors
Participation rate (Ratio of unique visitors who agreed to	Count the unique number of people who filled in the first survey page (or agreed to participate, for example by checking a checkbox), divided by visitors who visit the first page of the survey (or the informed consents page, if present). This can also be called “recruitment” rate.	N/A – Unable to obtain unique site visitors

participate/unique first survey page visitors)		
Completion rate (Ratio of users who finished the survey/users who agreed to participate)	The number of people submitting the last questionnaire page, divided by the number of people who agreed to participate (or submitted the first survey page). This is only relevant if there is a separate “informed consent” page or if the survey goes over several pages. This is a measure for attrition. Note that “completion” can involve leaving questionnaire items blank. This is not a measure for how completely questionnaires were filled in. (If you need a measure for this, use the word “completeness rate”.)	58
Cookies used	Indicate whether cookies were used to assign a unique user identifier to each client computer. If so, mention the page on which the cookie was set and read, and how long the cookie was valid. Were duplicate entries avoided by preventing users access to the survey twice; or were duplicate database entries having the same user ID eliminated before analysis? In the latter case, which entries were kept for analysis (eg, the first entry or the most recent)?	56
IP check	Indicate whether the IP address of the client computer was used to identify potential duplicate entries from the same user. If so, mention the period of time for which no two entries from the same IP address were allowed (eg, 24 hours). Were duplicate entries avoided by preventing users with the same IP address access to the survey twice; or were duplicate database entries having the same IP address within a given period of time eliminated before analysis? If the latter, which entries were kept for analysis (eg, the first entry or the most recent)?	56
Log file analysis	Indicate whether other techniques to analyze the log file for identification of multiple entries were used. If so, please describe.	56
Registration	In “closed” (non-open) surveys, users need to login first and it is easier to prevent duplicate entries from the same user. Describe how this was done. For example, was the survey never displayed a second time once the user had filled it in, or was the username stored together with the survey results and later eliminated? If the latter, which entries were kept for analysis (eg, the first entry or the most recent)?	N/A
Handling of incomplete questionnaires	Were only completed questionnaires analyzed? Were questionnaires which terminated early (where, for example, users did not go through all questionnaire pages) also analyzed?	56
Questionnaires submitted with an atypical timestamp	Some investigators may measure the time people needed to fill in a questionnaire and exclude questionnaires that were submitted too soon. Specify the timeframe that was used as a cut-off point, and describe how this point was determined.	N/A
Statistical correction	Indicate whether any methods such as weighting of items or propensity scores have been used to adjust for the non-representative sample; if so, please describe the methods.	56

Appendix Table C.2: COREQ (Consolidated criteria for REporting Qualitative research) Checklist

A checklist of items that should be included in reports of qualitative research. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Topic	Item No.	Guide Questions/Description	Reported on Page No.
Domain 1: Research team and reflexivity			
Personal characteristics			
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	58
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	58
Occupation	3	What was their occupation at the time of the study?	58
Gender	4	Was the researcher male or female?	58
Experience and training	5	What experience or training did the researcher have?	58
Relationship with participants			
Relationship established	6	Was a relationship established prior to study commencement?	58
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	58
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	58
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	58
Participant selection			

Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	58
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	58
Sample size	12	How many participants were in the study?	68
Non-participation	13	How many people refused to participate or dropped out? Reasons?	68
Setting			
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	58
Presence of nonparticipants	15	Was anyone else present besides the participants and researchers?	58
Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	68
Data collection			
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	N/A - Not an interview
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	N/A - Not an interview
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	58
Field notes	20	Were field notes made during and/or after the inter view or focus group?	58
Duration	21	What was the duration of the inter views or focus group?	58
Data saturation	22	Was data saturation discussed?	70
Transcripts returned	23	Were transcripts returned to participants for comment correction?	N/A – Not transcribed
Domain 3: analysis and findings			
Data analysis			

Number of data coders	24	How many data coders coded the data?	N/A - Themes identified prior to questionnaire. Only free text coded
Description of the coding tree	25	Did authors provide a description of the coding tree?	N/A - Themes identified prior to questionnaire. Only free text coded
Derivation of themes	26	Were themes identified in advance or derived from the data?	N/A - Themes identified prior to questionnaire. Only free text coded
Software	27	What software, if applicable, was used to manage the data?	51 & 56
Participant checking	28	Did participants provide feedback on the findings?	57
Reporting			
Quotations presented	29	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	68-69
Data and findings consistent	30	Was there consistency between the data presented and the findings?	N/A - Themes identified prior to questionnaire. Only free text coded

Clarity of major themes	31	Were major themes clearly presented in the findings?	N/A - Themes identified prior to questionnaire. Only free text coded.
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	N/A - Themes identified prior to questionnaire. Only free text coded

Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care. 2007. Volume 19, Number 6: pp. 349 - 357
Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

Appendix Table C.3: Recipients of Targeted Emails

Target Stakeholder Group Organization / Mailing List	A member of the public / patient or carer	Health Professional	Industry Partner / Payer	Statistician or other data analyst
Allstat A mailing list for statisticians and data scientists				✓
AMRC (Association of Medical Research Charities) A membership organisation of the leading medical and health research charities in the UK		✓		
APBI (Association of the British Pharmaceutical Industry) An organisation which represents companies who invest in drug development.			✓	
Association of Lead Research Nurses (Research and Innovation Bulletin) A network of research active nurses in the UK.		✓		
Cancer Patient and Public Involvement Group Leeds A patient and public involvement group based in Leeds, UK.	✓			
Cancer Research UK Clinical Trial Units A network of clinical trials units across the UK who receive infrastructure funding from cancer research UK.	✓	✓		✓
Health Data Research UK A charity which crosscuts industry and academia focusing on matters relating to health data science.	✓	✓		✓
Health Research Board Trials Methodology Research Network (Ireland) A network of researchers based in Ireland focusing on trials methodology.				✓
HESG (Health Economists' Study Group) Email List A mailing list of all members signed up to the HESG.				✓
HRA (Health Research Agency)			✓	

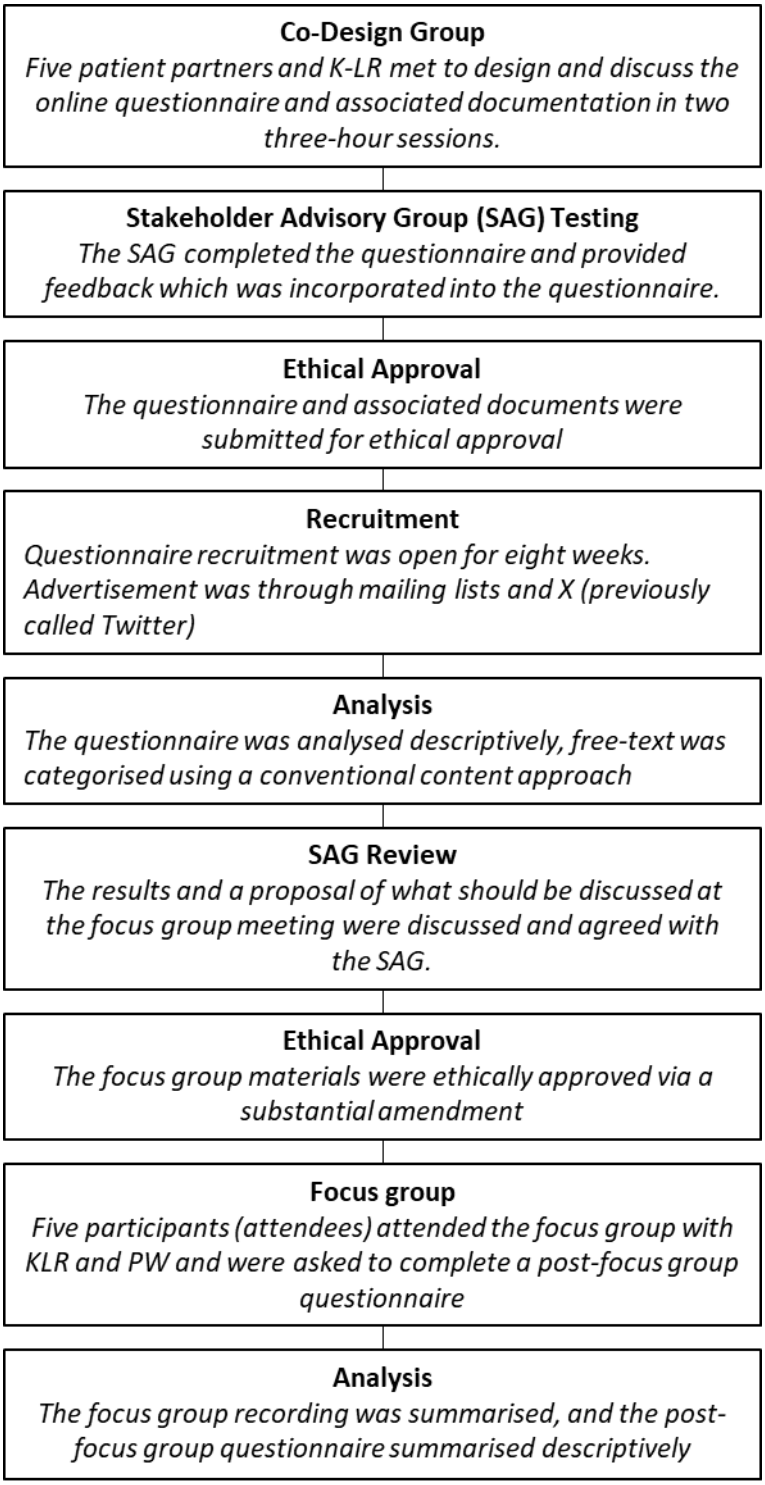
Target Stakeholder Group Organization / Mailing List	A member of the public / patient or carer	Health Professional	Industry Partner / Payer	Statistician or other data analyst
An organisation ensuring that health and social care research protects the interests of patients and the public.				
Independent Cancer Patient Voice A patient advocate group supporting involvement in research,	✓			
Lymphoma Action A UK charity dedicated to supporting patients with lymphoma	✓			
Marie Curie A UK charity dedicated to supporting patients at end of life care.	✓			
Mednet A University of Leeds mailing list for everyone in medicine and health.	✓	✓	✓	✓
Myeloma UK A UK charity dedicated to supporting patients with Myeloma	✓			
NCRI (National Cancer Research Institute) A network of researchers and patients to support the advancement of cancer research.	✓	✓	✓	✓
NHS Digital An organisation which focuses on the sharing of health systems data.			✓	
SOLVE PhD Supervisors A group of three statisticians and an health economist spanning industry and academia.				✓
PSI (Statisticians in the Pharmaceutical Industry) A network of statisticians and other data analysts working in medical research both in industry and academia.				✓
RADAR Patient Advisory Group A group of six patients and carers who provide PPI for the Myeloma trial RADAR run by Leeds CTRU.	✓			

Target Stakeholder Group Organization / Mailing List	A member of the public / patient or carer	Health Professional	Industry Partner / Payer	Statistician or other data analyst
Registered Clinical Trials Unit (CTU) Network A network of all of the registered CTUs in the UK	✓	✓	✓	✓
Royal Colleges A membership organisation for the 23 medical royal colleges and faculties focusing on the different specialities in the UK.		✓		
Royal Statistical Society A membership organisation for statisticians				✓
SOLVE Stakeholder Advisory Group A group of 2 patient representatives, 2 health economists and 2 clinicians.	✓	✓		✓
Trial Managers' Network A network of trial managers who work in clinical trials in the UK.	✓			
Trials Methodology Research Partnership (UK) A network of institutions and organisations who work in trial methodology.	✓	✓	✓	✓
Yorkshire Cancer Community A charity focused on supporting people in Yorkshire who have cancer.	✓			
Cancer Drugs Fund Team An organisation focused on facilitating early access to effective treatments in the UK		✓	✓	

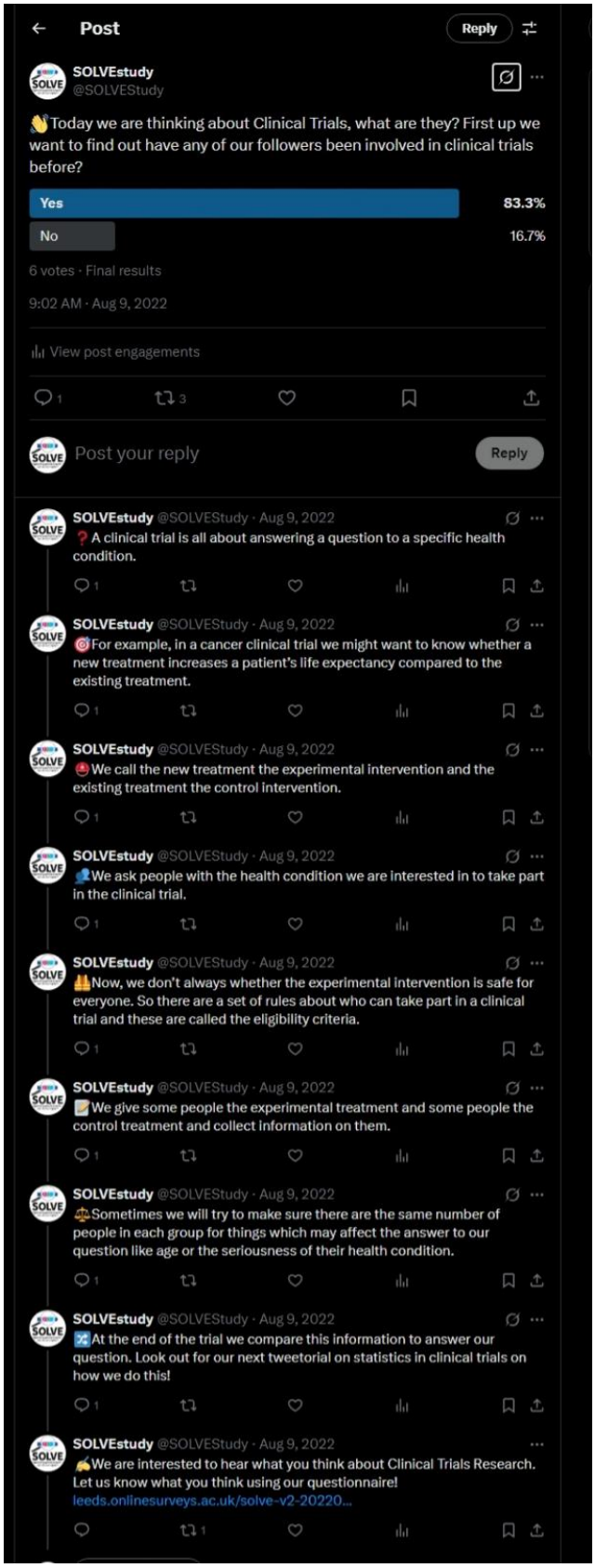
Appendix Table C.4: Summary of Invited Respondents

	Attended (n=6)	Declined (n=8)	Did Not Attend (n=5)	Not Invited (n=14)	Total (n=33)
How old are you?					
Median (IQR)	58.00 (37.00, 62.00)	46.50 (39.00, 54.00)	55.00 (50.00, 55.00)	60.50 (48.00, 68.00)	55.00 (45.00, 64.00)
What is your ethnic group?					
English/Welsh/Scottish/Northern Irish/British	5 (83.3%)	6 (75.0%)	2 (40.0%)	13 (92.9%)	26 (78.8%)
Irish	1 (16.7%)	0 (0.0%)	1 (20.0%)	0 (0.0%)	2 (6.1%)
Other, please describe	0 (0.0%)	1 (12.5%)	1 (20.0%)	0 (0.0%)	2 (6.1%)
Prefer not to say	0 (0.0%)	1 (12.5%)	1 (20.0%)	0 (0.0%)	2 (6.1%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.1%)	1 (3.0%)
What sex are you?					
Male	5 (83.3%)	3 (37.5%)	1 (20.0%)	3 (21.4%)	12 (36.4%)
Female	1 (16.7%)	5 (62.5%)	4 (80.0%)	10 (71.4%)	20 (60.6%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.1%)	1 (3.0%)
Where in the UK are you based?					
Scotland	0 (0.0%)	0 (0.0%)	2 (40.0%)	1 (7.1%)	3 (9.1%)
Northern Ireland	1 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.0%)
North East England	1 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.0%)
Yorkshire and The Humber	2 (33.3%)	3 (37.5%)	2 (40.0%)	6 (42.9%)	13 (39.4%)
East Midlands	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.1%)	1 (3.0%)
West Midlands	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	1 (3.0%)
London	0 (0.0%)	2 (25.0%)	1 (20.0%)	2 (14.3%)	5 (15.2%)
South East England	1 (16.7%)	1 (12.5%)	0 (0.0%)	2 (14.3%)	4 (12.1%)
South West England	1 (16.7%)	0 (0.0%)	0 (0.0%)	2 (14.3%)	3 (9.1%)
Other	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	1 (3.0%)
Are You?					
A Member of the Public	1 (16.7%)	1 (12.5%)	1 (20.0%)	3 (21.4%)	6 (18.2%)
A Patient or Carer	1 (16.7%)	1 (12.5%)	1 (20.0%)	8 (57.1%)	11 (33.3%)
A Health Professional	1 (16.7%)	2 (25.0%)	1 (20.0%)	0 (0.0%)	4 (12.1%)
An Industry Partner	0 (0.0%)	0 (0.0%)	1 (20.0%)	0 (0.0%)	1 (3.0%)
A Payer	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	1 (3.0%)
A Statistician or Other Data Analyst	3 (50.0%)	3 (37.5%)	1 (20.0%)	3 (21.4%)	10 (30.3%)
Have you had any prior experience of clinical trials?					
Yes	5 (83.3%)	7 (87.5%)	4 (80.0%)	9 (64.3%)	25 (75.8%)
No	1 (16.7%)	1 (12.5%)	1 (20.0%)	5 (35.7%)	8 (24.2%)
Do you think it is important to consider the effect that treatment given after trial treatment has on overall survival?					
Yes	5 (83.3%)	5 (62.5%)	4 (80.0%)	13 (92.9%)	27 (81.8%)
Depends	1 (16.7%)	3 (37.5%)	1 (20.0%)	1 (7.1%)	6 (18.2%)
Before today have you ever thought about how after trial treatment might affect overall survival?					
Yes	4 (66.7%)	7 (87.5%)	2 (40.0%)	8 (57.1%)	21 (63.6%)
No	2 (33.3%)	1 (12.5%)	3 (60.0%)	5 (35.7%)	11 (33.3%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.1%)	1 (3.0%)

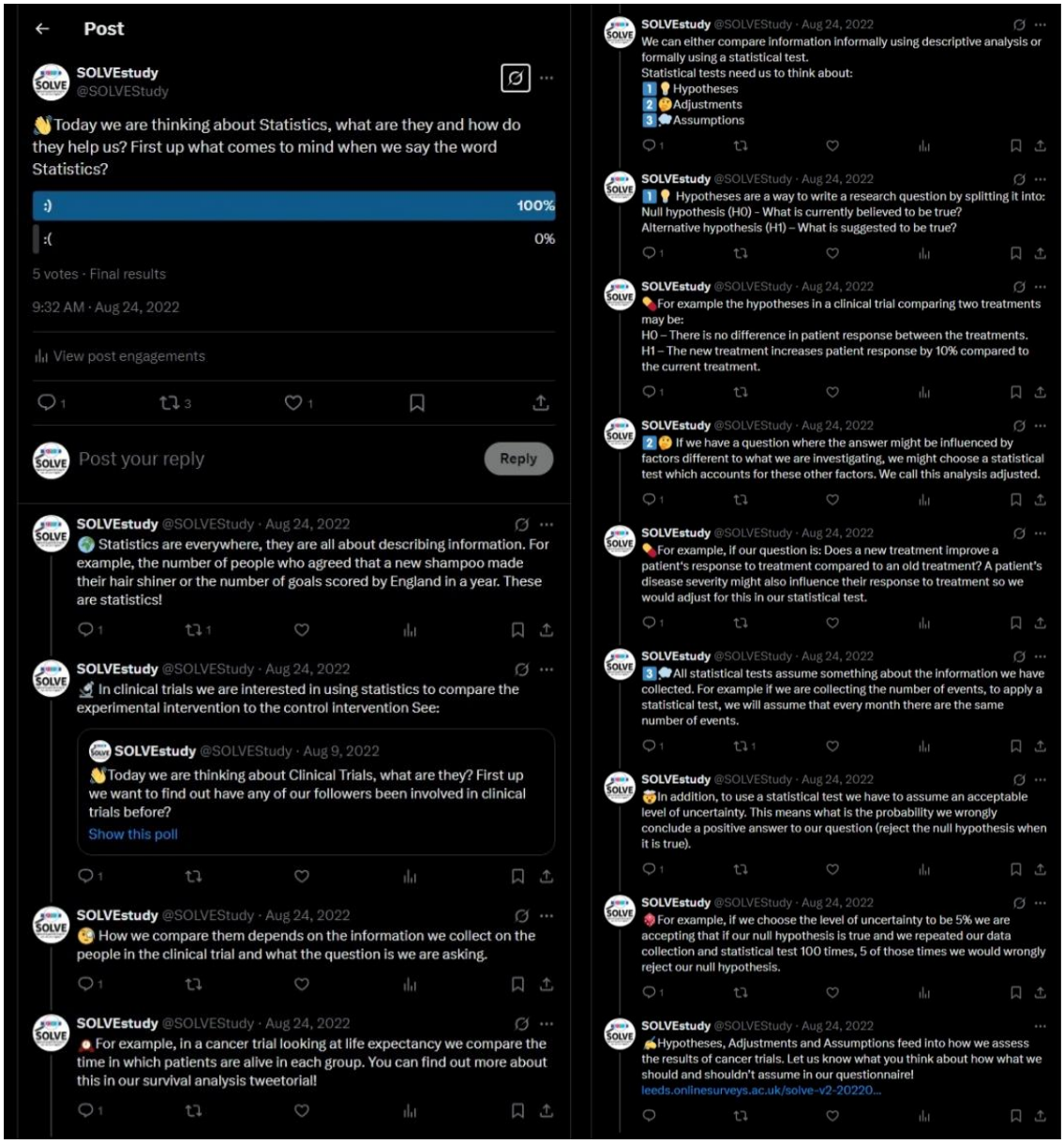
Appendix Figure C.1: Flow Chart of Implemented Methods



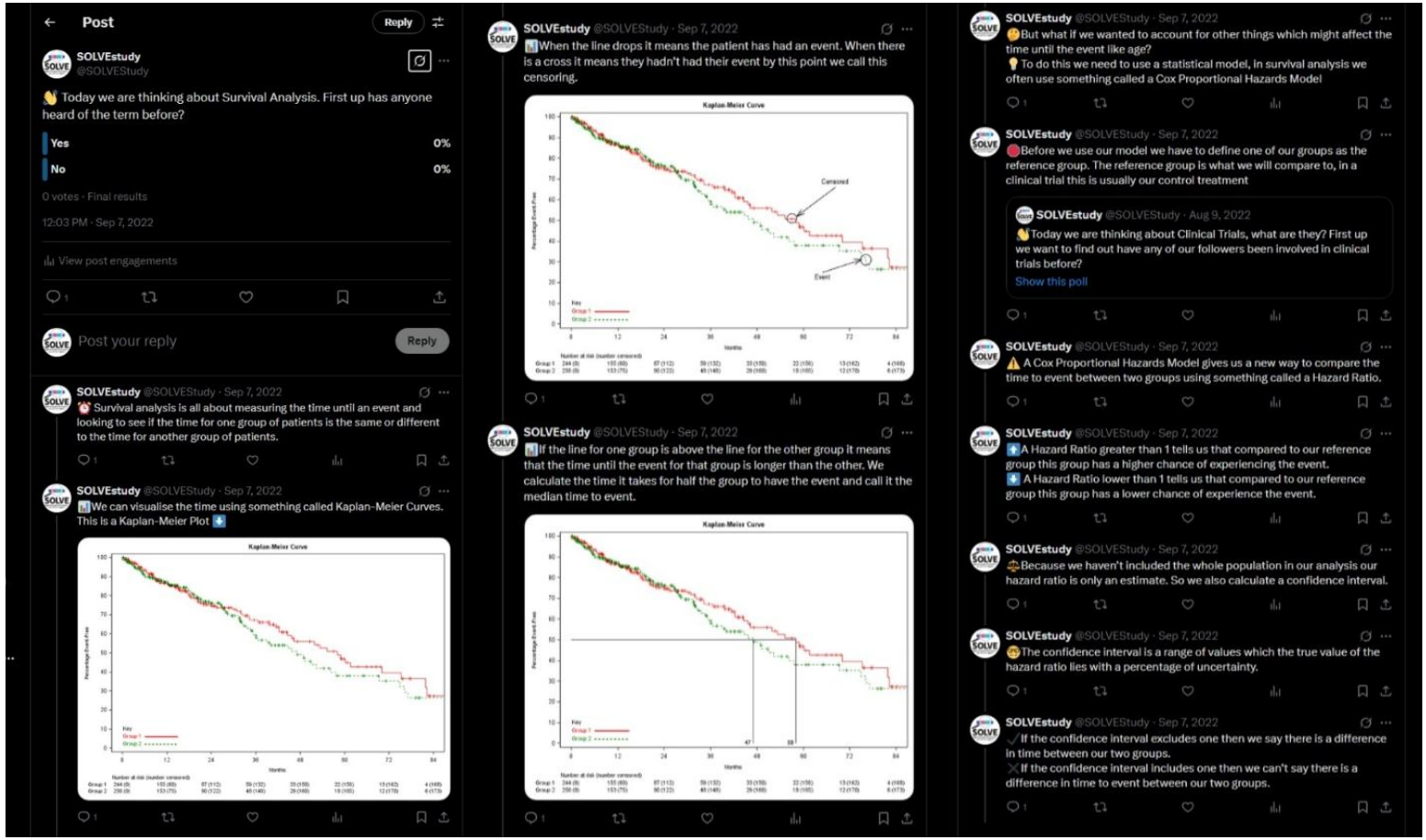
Appendix Figure C.2: Clinical Trials Tweetorial



Appendix Figure C.3: Statistics Tweetorial

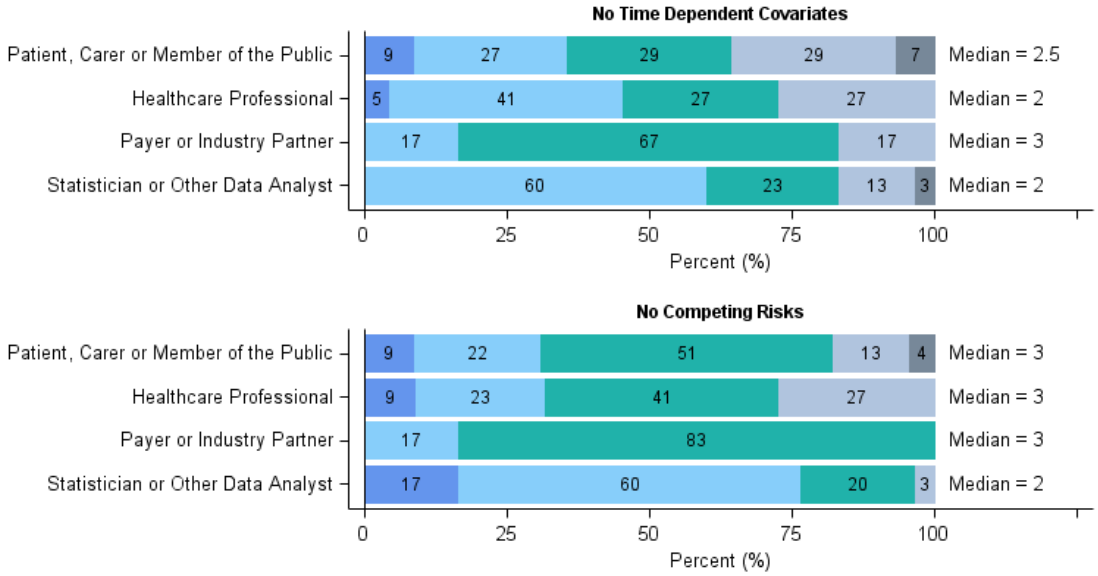


Appendix Figure C.4: Survival Analysis Tweetorial



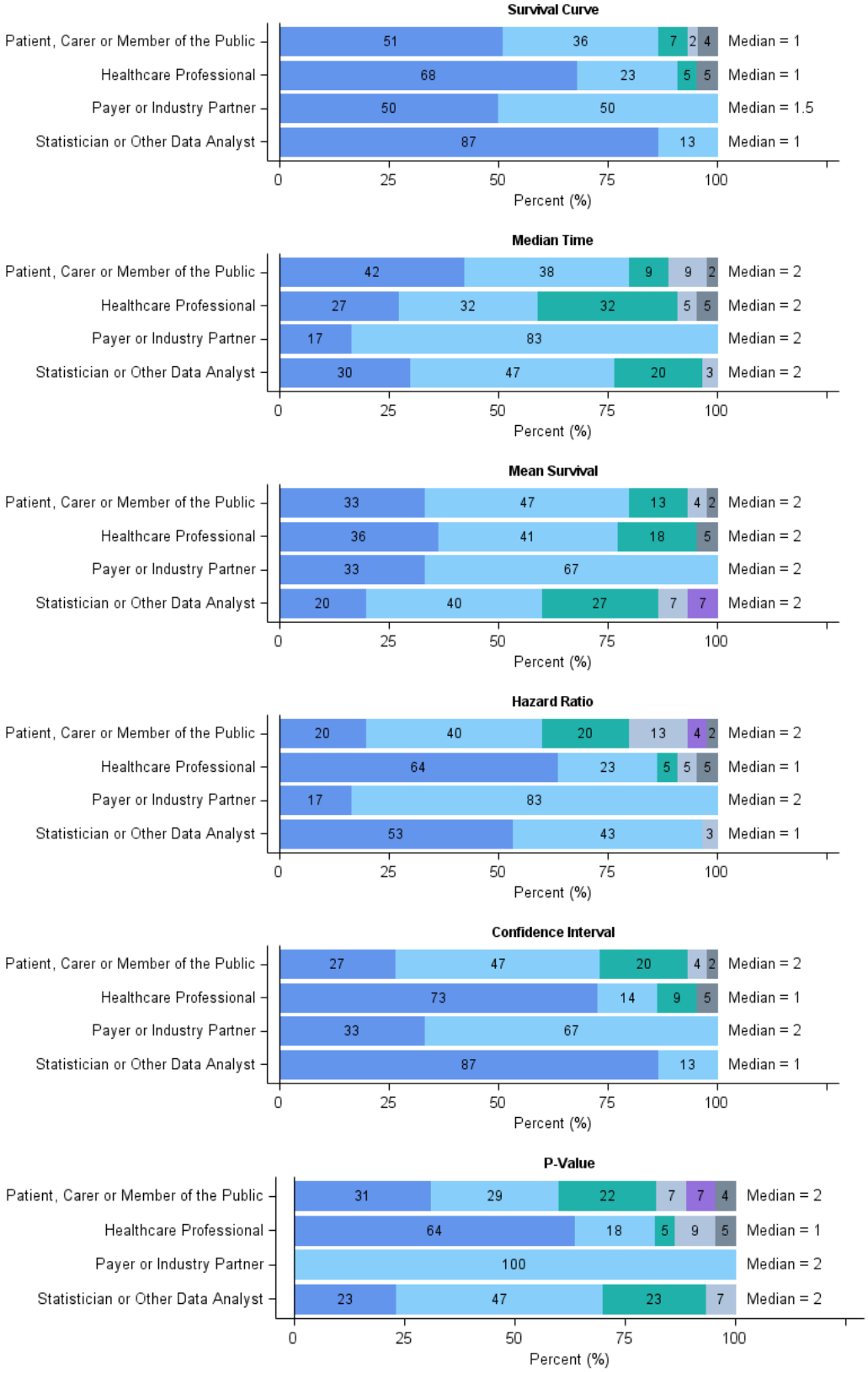
Appendix Figure C.5: Summary of Statistical Assumptions Results





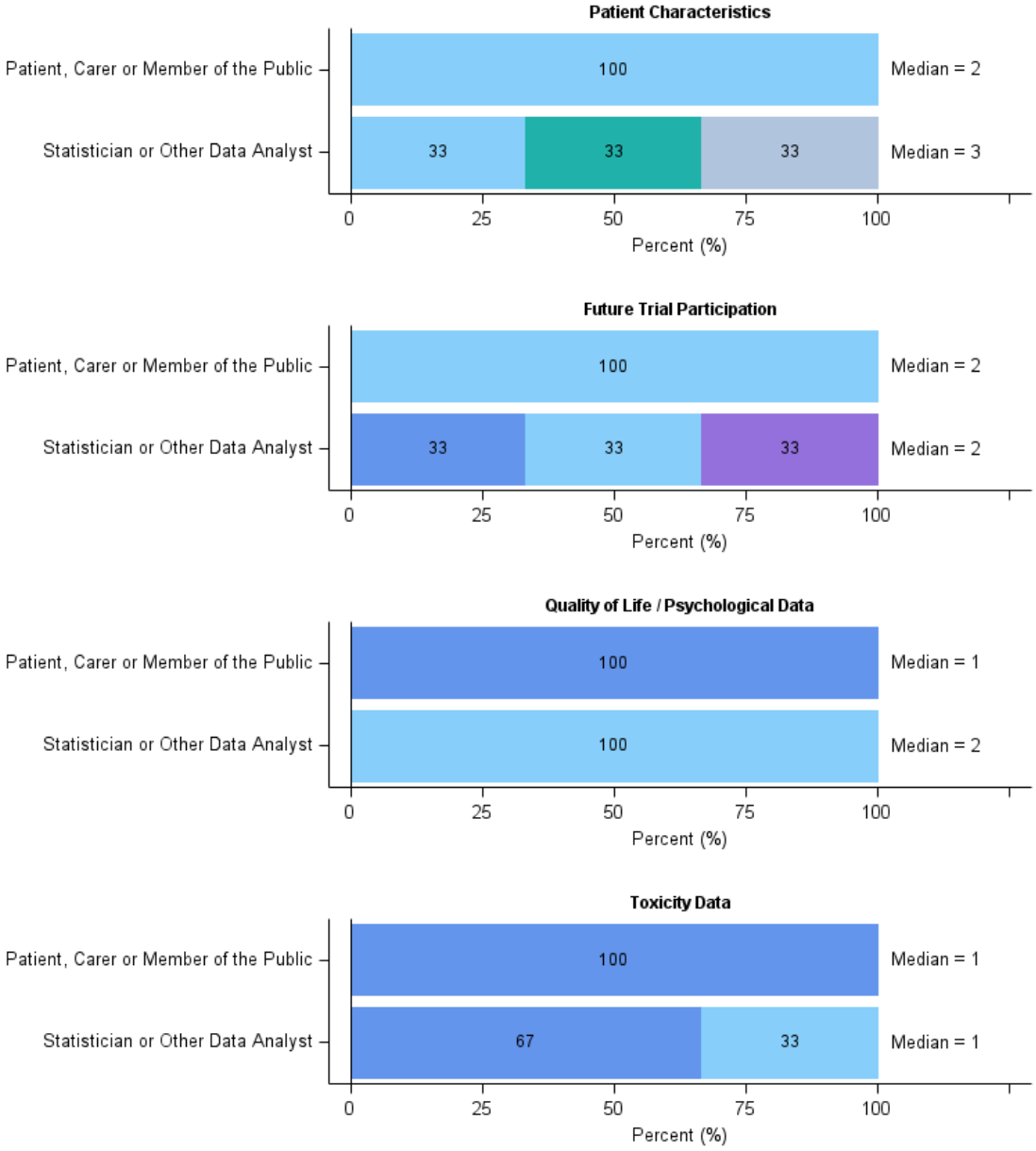
Key	All Scenarios	Some Scenarios	No Scenarios	Unsure	Missing

Appendix Figure C.6: Summary of Presentation of Results



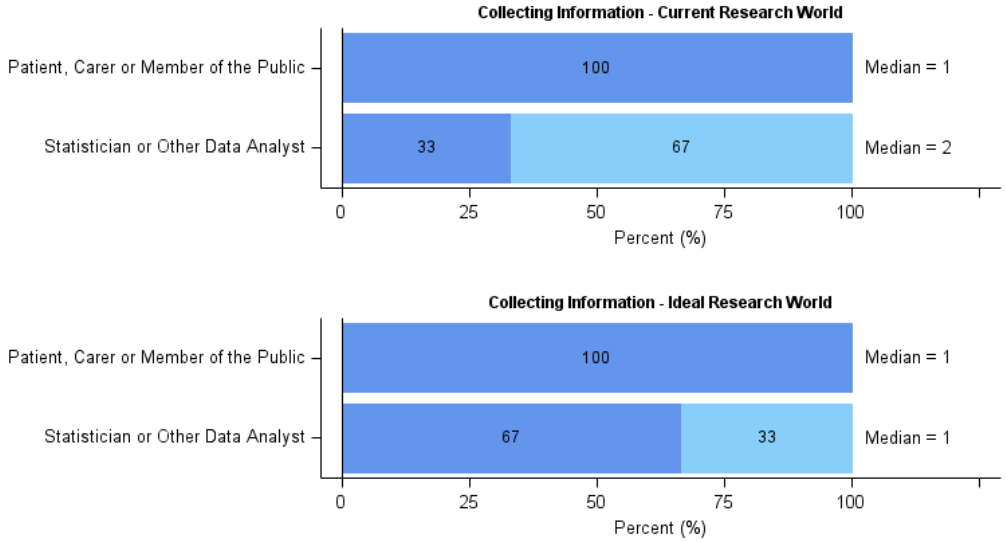
Key	Very Helpful	Helpful	Neither Helpful or Unhelpful	Unhelpful	Very Unhelpful	Missing

Appendix Figure C.7: Post-Focus Group Questionnaire Results – Data Items



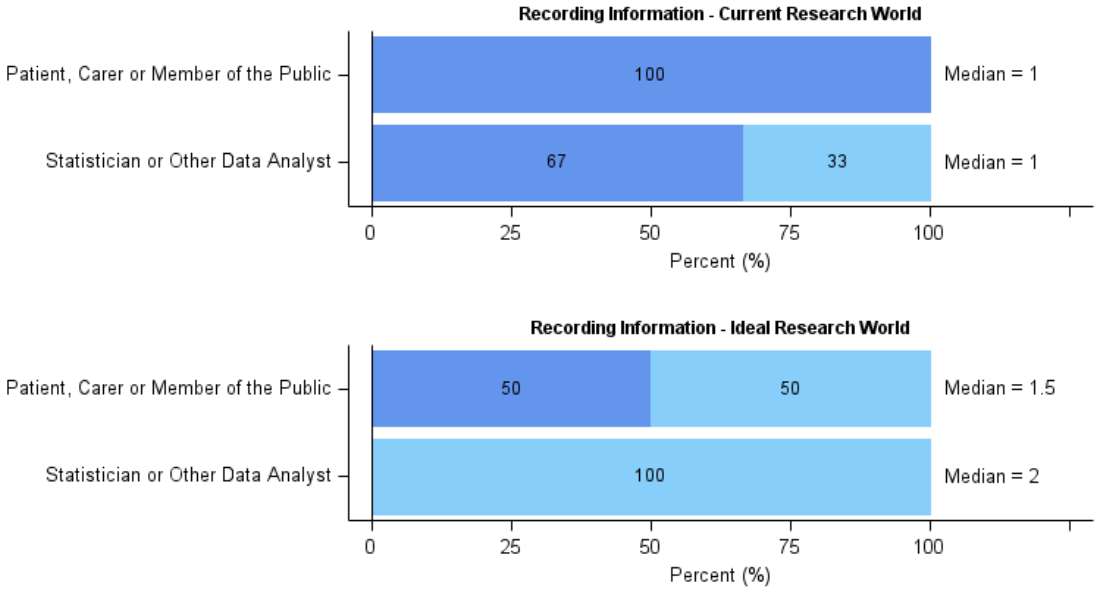
Key					
Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree	Missing

Appendix Figure C.8: Post-Focus Group Questionnaire Results – Collecting Information



Key	
Colour	**Score**
	At a trial follow-up appointment.
	At a routine appointment.
	Missing

Appendix Figure C.9: Post-Focus Group Questionnaire Results – Recording Information



Key	
Colour	**Choice**
	From a database which was made specifically for the trial.
	From a database which is completed normally as part of standard practice.
	Missing

Appendix D Chapter 4 Supplementary Material

This document provides the supplementary material for: What Would Happen if Everyone Received the Same Subsequent Therapy? A Hypothetical Estimand for Overall Survival in the Presence of Treatment Discontinuation

D.1 Additional Methods Information

This section provides additional information on the methods implemented within the simulation study.

D.1.1 Inputted Parameters

Appendix Table D.1 provides full descriptions for the parameters and inputted values for the base scenario.

Appendix Table D.1: Simulation Study Parameter and Variable Descriptions for the Base Scenario

Notation	Description	Parameter value
N	Number of patients	700
Variables and parameters required for simulation of baseline variables		
bo_1	Observed covariate 1 at baseline (binary) • $bo_1 \sim Bernoulli(bop_1)$	
bop_1	Prevalence of observed covariate 1 at baseline	10%
bo_2	Observed covariate 2 at baseline (binary) • $bo_2 \sim Bernoulli(bop_2)$	
bop_2	Prevalence of observed covariate 2 at baseline	60%
bo_3	Observed covariate 3 at baseline (categorical) • $bo_3 \sim Binomial(2, bop_3)$	
bop_3	Probability of a success (maximum 2) for observed covariate 3 at baseline	20%
bu_1	Unobserved covariate 1 at baseline (binary) • $bu_1 \sim Bernoulli(bup_1)$	
bup_1	Prevalence of unobserved covariate 1 at baseline	40%
bu_2	Unobserved covariate 2 at baseline (continuous) • $bu_2 \sim Normal(bum_2, bus_2^2)$	
bum_2	Mean of unobserved covariate 2 at baseline	0
bus_2	Standard deviation of unobserved covariate 2 at baseline	1
$randtrt$	Randomisation allocation (binary) stratified by bo_1, bo_2, bo_3.	
Variables and parameters required for simulation of follow-up variables		
$fo_{2[i]}$	Observed covariate 2 at follow-up timepoint i (binary). The logistic regression model included: $bo_1, bo_2, bo_3, bu_1, bu_2, randtrt$ and fo_2, fo_3, fu_1, fu_2 at the previous timepoints.	
$\beta_{[fo2,bo1]}$	Presence of observed covariate 1 at baseline relative effect on observed covariate 2 at follow-up compared to absence of observed covariate 1 at baseline	1.2

Notation	Description	Parameter value
$\beta_{[fo2,bo2]}$	Presence of observed covariate 2 at baseline relative effect on observed covariate 2 at follow-up compared to absence of observed covariate 2 at baseline	1.8
$\beta_{[fo2,bo3D1]}$	Observed covariate 3 group 2 relative effect on observed covariate 2 at follow-up compared to group 1 at baseline	1.2
$\beta_{[fo2,bo3D2]}$	Observed covariate 3 group 3 relative effect on observed covariate 2 at follow-up compared to group 1 at baseline	1.4
$\beta_{[fo2,bu1]}$	Presence of unobserved covariate 1 at baseline relative effect on observed covariate 2 at follow-up compared to absence of unobserved covariate 1 at baseline	1.2
$\beta_{[fo2,bu2]}$	One unit increase of unobserved covariate 2 at baseline effect on observed covariate 2 at follow-up	1.2
$\beta_{[fo2,trt]}$	Allocation to experimental relative effect on observed covariate 2 at follow-up compared to allocation to control	0.8
$\beta_{[fo2,fo2]}$	Presence of observed covariate 2 at previous follow-up relative effect on observed covariate 2 at current follow-up compared to absence of observed covariate 2 at previous follow-up	1.8
$\beta_{[fo2,fo3D1]}$	Observed covariate 3 group 2 at previous follow-up relative effect on effect on observed covariate 2 at current follow-up compared to group 1 at previous follow-up	1.2
$\beta_{[fo2,fo3D2]}$	Observed covariate 3 group 3 at previous follow-up relative effect on observed covariate 2 at current follow-up compared to group 1 at follow-up at previous follow-up	1.2
$\beta_{[fo2,fu1]}$	Presence of unobserved covariate 1 at previous follow-up relative effect on observed covariate 2 at current follow-up compared to absence of unobserved covariate 1 at previous follow-up	1.2
$\beta_{[fo2,fu2]}$	One unit increase of unobserved covariate 2 at follow-up effect on observed covariate 2 at follow-up	1.2
$fo_{3[i]}$	Observed covariate 3 at follow-up timepoint i (categorical). The logistic regression model included: $bo_1, bo_2, bo_3, bu_1, bu_2, randtrt$ and fo_2, fo_3, fu_1, fu_2 at the previous timepoints.	
$\beta_{[fo3,bo1]}$	Presence of observed covariate 1 at baseline relative effect on observed covariate 3 at follow-up compared to absence of observed covariate 1 at baseline	1.2
$\beta_{[fo3,bo2]}$	Presence of observed covariate 2 at baseline relative effect on observed covariate 3 at follow-up compared to absence of observed covariate 2 at baseline	1.2
$\beta_{[fo3,bo3D1]}$	Observed covariate 3 group 2 relative effect on observed covariate 3 at follow-up compared to group 1 at baseline	1.5
$\beta_{[fo3,bo3D2]}$	Observed covariate 3 group 3 relative effect on observed covariate 3 at follow-up compared to group 1 at baseline	1.8

Notation	Description	Parameter value
$\beta_{[fo3,bu1]}$	Presence of unobserved covariate 1 at baseline relative effect on observed covariate 3 at follow-up compared to absence of unobserved covariate 1 at baseline	1.2
$\beta_{[fo3,bu2]}$	One unit increase of unobserved covariate 2 at baseline effect on observed covariate 3 at follow-up	1.2
$\beta_{[fo3,trt]}$	Allocation to experimental relative effect on observed covariate 3 at follow-up compared to allocation to control	0.8
$\beta_{[fo3,fo2]}$	Presence of observed covariate 2 at previous follow-up relative effect on observed covariate 3 at current follow-up compared to absence of observed covariate 2 at previous follow-up	1.2
$\beta_{[fo3,fo3D1]}$	Observed covariate 3 group 2 at previous follow-up relative effect on effect on observed covariate 3 at current follow-up compared to group 1 at previous follow-up	1.5
$\beta_{[fo3,fo3D2]}$	Observed covariate 3 group 3 at previous follow-up relative effect on observed covariate 3 at current follow-up compared to group 1 at follow-up at previous follow-up	1.8
$\beta_{[fo3,fu1]}$	Presence of unobserved covariate 1 at previous follow-up relative effect on observed covariate 3 at current follow-up compared to absence of unobserved covariate 1 at previous follow-up	1.2
$\beta_{[fo3,fu2]}$	One unit increase of unobserved covariate 2 at previous follow-up effect on observed covariate 3 at current follow-up	1.2
$fu_{1[i]}$	Unobserved covariate 1 at follow-up timepoint i (binary). The logistic regression model included: $bo_1, bo_2, bo_3, bu_1, bu_2, randtrt$ and fo_2, fo_3, fu_1, fu_2 at the previous timepoints.	
$\beta_{[fu1,bo1]}$	Presence of observed covariate 1 at baseline relative effect on unobserved covariate 1 at follow-up compared to absence of observed covariate 1 at baseline	1.2
$\beta_{[fu1,bo2]}$	Presence of observed covariate 2 at baseline relative effect on unobserved covariate 1 at follow-up compared to absence of observed covariate 2 at baseline	1.2
$\beta_{[fu1,bo3D1]}$	Observed covariate 3 group 2 relative effect on unobserved covariate 1 at follow-up compared to group 1 at baseline	1.2
$\beta_{[fu1,bo3D2]}$	Observed covariate 3 group 3 relative effect on unobserved covariate 1 at follow-up compared to group 1 at baseline	1.4
$\beta_{[fu1,bu1]}$	Presence of unobserved covariate 1 at baseline relative effect on unobserved covariate 1 at follow-up compared to absence of unobserved covariate 1 at baseline	1.8
$\beta_{[fu1,bu2]}$	One unit increase of unobserved covariate 2 at baseline effect on unobserved covariate 1 at follow-up	1.2
$\beta_{[fu1,trt]}$	Allocation to experimental relative effect on unobserved covariate 1 at follow-up compared to allocation to control	0.8

Notation	Description	Parameter value
$\beta_{[fu1,fo2]}$	Presence of observed covariate 2 at previous follow-up relative effect on unobserved covariate 1 at current follow-up compared to absence of observed covariate 2 at previous follow-up	1.2
$\beta_{[fu1,fo3D1]}$	Observed covariate 3 group 2 at previous follow-up relative effect on effect on unobserved covariate 1 at current follow-up compared to group 1 at previous follow-up	1.2
$\beta_{[fu1,fo3D2]}$	Observed covariate 3 group 3 at previous follow-up relative effect on unobserved covariate 1 at current follow-up compared to group 1 at follow-up at previous follow-up	1.2
$\beta_{[fu1,fu1]}$	Presence of unobserved covariate 1 at previous follow-up relative effect on unobserved covariate 1 at current follow-up compared to absence of unobserved covariate 1 at previous follow-up	1.8
$\beta_{[fu1,fu2]}$	One unit increase of unobserved covariate 2 at previous follow-up effect on unobserved covariate 1 at current follow-up	1.2
$fu_{2[i]}$	Unobserved covariate 2 at follow-up timepoint i (continuous). The linear regression model included: $bo_{1,}bo_2, bo_3, bu_1, bu_2, randtrt$ and fo_2, fo_3, fu_1, fu_2 at the previous timepoints.	
$\beta_{[fu2,bo1]}$	Presence of observed covariate 1 at baseline relative effect on unobserved covariate 2 at follow-up compared to absence of observed covariate 1 at baseline	1.2
$\beta_{[fu2,bo2]}$	Presence of observed covariate 2 at baseline relative effect on unobserved covariate 2 at follow-up compared to absence of observed covariate 2 at baseline	1.2
$\beta_{[fu2,bo3D1]}$	Observed covariate 3 group 2 relative effect on unobserved covariate 2 at follow-up compared to group 1 at baseline	1.2
$\beta_{[fu2,bo3D2]}$	Observed covariate 3 group 3 relative effect on unobserved covariate 2 at follow-up compared to group 1 at baseline	1.4
$\beta_{[fu2,bu1]}$	Presence of unobserved covariate 1 at baseline relative effect on unobserved covariate 2 at follow-up compared to absence of unobserved covariate 1 at baseline	1.2
$\beta_{[fu2,bu2]}$	One unit increase of unobserved covariate 2 at baseline effect on unobserved covariate 2 at follow-up	1.8
$\beta_{[fu2,trt]}$	Allocation to experimental relative effect on unobserved covariate 2 at follow-up compared to allocation to control	0.8
$\beta_{[fu2,fo2]}$	Presence of observed covariate 2 at previous follow-up relative effect on unobserved covariate 2 at current follow-up compared to absence of observed covariate 2 at previous follow-up	1.2
$\beta_{[fu2,fo3D1]}$	Observed covariate 3 group 2 at previous follow-up relative effect on effect on unobserved covariate 2 at current follow-up compared to group 1 at previous follow-up	1.2
$\beta_{[fu2,fo3D2]}$	Observed covariate 3 group 3 at previous follow-up relative effect on unobserved covariate 2 at current	1.2

Notation	Description	Parameter value
	follow-up compared to group 1 at follow-up at previous follow-up	
$\beta_{[fu2,fu1]}$	Presence of unobserved covariate 1 at previous follow-up relative effect on unobserved covariate 2 at current follow-up compared to absence of unobserved covariate 1 at previous follow-up	1.2
$\beta_{[fu2,fu2]}$	One unit increase of unobserved covariate 2 at previous follow-up effect on unobserved covariate 2 at current follow-up	1.8
α_{fu2}	Intercept for linear regression predicting unobserved covariate 2	0.05
Variables and parameters required for simulation of discontinuation		
$Discontinue_{[i]}$	Discontinuation at timepoint i (binary). The logistic regression model included $bo_1, bo_2, bo_3, bu_1, bu_2, randtrt$ and fo_2, fo_3, fu_1, fu_2 at timepoint i.	
α_D	Intercept for logistic regression model predicting discontinuation	0.005
$\beta_{[D,bo1]}$	Presence of observed covariate 1 at baseline relative effect on discontinuation compared to absence of observed covariate 1 at baseline	1.05
$\beta_{[D,bo2]}$	Presence of observed covariate 2 at baseline relative effect on discontinuation compared to absence of observed covariate 2 at baseline	1.02
$\beta_{[D,bo3D1]}$	Observed covariate 3 group 2 relative effect on discontinuation compared to group 1 at baseline	1.10
$\beta_{[D,bo3D2]}$	Observed covariate 3 group 3 relative effect on discontinuation compared to group 1 at baseline	1.20
$\beta_{[D,bu1]}$	Presence of unobserved covariate 1 at baseline relative effect on discontinuation compared to absence of unobserved covariate 1 at baseline	1.05
$\beta_{[D,bu2]}$	One unit increase of unobserved covariate 2 at baseline effect on discontinuation	1.02
$\beta_{[D,fo2]}$	Presence of observed covariate 2 at follow-up relative effect on discontinuation compared to absence of observed covariate 2 at follow-up	1.10
$\beta_{[D,fo3D1]}$	Observed covariate 3 group 2 relative effect on discontinuation compared to group 1 at follow-up	1.20
$\beta_{[D,fo3D2]}$	Observed covariate 3 group 3 relative effect on discontinuation compared to group 1 at follow-up	1.30
$\beta_{[D,fu1]}$	Presence of unobserved covariate 1 at follow-up relative effect on discontinuation compared to absence of unobserved covariate 1 at follow-up	1.15
$\beta_{[D,fu2]}$	One unit increase of unobserved covariate 2 at follow-up effect on discontinuation	1.25
$\beta_{[D,trt]}$	Allocation to experimental relative effect on discontinuation compared to allocation to control	0.65
Variables and parameters required for simulation of subsequent treatment		
$subtrt$	Subsequent anti-cancer therapy (Binary) The logistic regression model included $randtrt$ and fo_2, fo_3, fu_1, fu_2 at the point of discontinuation. Note the model intercept was used to fix the proportion of participants receiving each subsequent anti-cancer therapy to a user specified proportion ($subtreatp$).	
$subtreatp$	Prevalence of subsequent anti-cancer therapy A	30%

Notation	Description	Parameter value
$\beta_{[A,fo2]}$	Presence of observed covariate 2 at follow-up relative effect on receiving subsequent anti-cancer therapy A compared to absence of observed covariate 2 at follow-up	0.95
$\beta_{[A,fo3D1]}$	Observed covariate 3 group 2 at follow-up relative effect on receiving subsequent anti-cancer therapy A compared to group 1 at follow-up	0.8
$\beta_{[A,fo3D2]}$	Observed covariate 3 group 3 at follow-up relative effect on receiving subsequent anti-cancer therapy A compared to group 1 at follow-up	0.2
$\beta_{[A,fu1]}$	Presence of unobserved covariate 1 at follow-up relative effect on receiving subsequent anti-cancer therapy A compared to absence of unobserved covariate 1 at follow-up	0.9
$\beta_{[A,fu2]}$	One unit increase of unobserved covariate 2 at follow-up effect on receiving subsequent anti-cancer therapy A	0.85
$\beta_{[A,trt]}$	Allocation to experimental relative effect on receiving subsequent anti-cancer therapy A compared to allocation to control	1.45
Variables and parameters required for simulation of survival		
T_s, E	Survival Time and Event. The log-hazard function included $bo_1, bo_2, bo_3, bu_1, bu_2,$ $randtrt, fo_2, fo_3, fu_1, fu_2$ and $subtrt.$ Note indicator functions were included so the instance of each follow-up variable and subsequent treatment only influenced survival in the appropriate interval.	
t_{max}	Maximum follow-up time	5 years
k	Number of follow-up timepoints	4
λ	Baseline hazard of death	0.2
$\beta_{[S,bo1]}$	Presence of observed covariate 1 at baseline relative effect on death compared to absence of observed covariate 1 at baseline	1.15
$\beta_{[S,bo2]}$	Presence of observed covariate 2 at baseline relative effect on death compared to absence of observed covariate 2 at baseline	1.05
$\beta_{[S,bo3D1]}$	Observed covariate 3 group 2 at baseline relative effect on death compared to group 1 at baseline	1.05
$\beta_{[S,bo3D2]}$	Observed covariate 3 group 3 at baseline relative effect on death compared to group 1 at baseline	1.10
$\beta_{[S,bu1]}$	Presence of unobserved covariate 1 at baseline relative effect on death compared to absence of unobserved covariate 1 at baseline	1.10
$\beta_{[S,bu2]}$	One unit increase of unobserved covariate 2 at baseline effect on death	1.05
$\beta_{[S,fo2]}$	Presence of observed covariate 2 at follow-up relative effect on death compared to absence of observed covariate 2 at follow-up	1.05
$\beta_{[S,fo3D1]}$	Observed covariate 3 group 2 at follow-up relative effect on death compared to group 1 at follow-up	1.05
$\beta_{[S,fo3D2]}$	Observed covariate 3 group 3 at follow-up relative effect on death compared to group 1 at follow-up	1.15
$\beta_{[S,fu1]}$	Presence of unobserved covariate 1 at follow-up relative effect on death compared to absence of unobserved covariate 1 at follow-up	1.05

Notation	Description	Parameter value
$\beta_{[S,fu2]}$	One unit increase of unobserved covariate 2 at follow-up effect on death	1.15
$\beta_{[S,trt]}$	Allocation to experimental relative effect on death compared to allocation to control	0.725
$\beta_{[S,A]}$	Receipt of subsequent anti-cancer therapy A relative effect on death compared to receipt of subsequent anti-cancer therapy B	0.8

Note that variables are shown in shaded rows above their related parameters.

D.1.2 Estimation and Interpretation of the Cox Proportional Hazards Model

This section describes how the estimates of the parameters in the Cox proportional hazards model are estimated using a partial likelihood (Equation 5.1). Note that the partial likelihood only depends on the order in which survival events occur rather the times or events themselves.

Equation 5.1: Partial Likelihood Function for Cox Proportional Hazards Model

Suppose N individuals in a clinical trial are followed up from time t_o to time T with the last follow-up time for each individual denoted as $t_1, \dots, t_N$ in chronological order. For every t_i define an $\boldsymbol{x}_i$, δ_i and $R(t_i)$ such that:

1. $\boldsymbol{x}_i$ represents a vector of explanatory variables for the individual with the i^{th} follow-up time.
2. δ_i defines whether the follow-up time t_i corresponds to an individual experiencing the event of interest ($\delta_i = 1$) or a right censored observation ($\delta_i = 0$).
3. $R(t_i)$ corresponds to a subset of the N individuals who are event-free and at therefore at risk of the event of interest just before time t_i.

Then the partial likelihood function for the Cox proportional hazards model is given by:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left\{ \frac{\exp(\boldsymbol{\beta x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta x}_l)} \right\}^{\delta_i}$$

Where, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$ is the vector of parameters to be estimated for a model with k explanatory variables.

From the partial likelihood function, a partial log-likelihood function can be derived (Equation 5.2) and estimates for the $\boldsymbol{\beta}$ parameters ($\widehat{\boldsymbol{\beta}}$) and the associated standard errors ($se(\widehat{\boldsymbol{\beta}})$) are obtained using numerical methods to maximise the partial log-likelihood function. Note that the lack of assumptions on the functional form of the baseline hazard makes the method semi-parametric.

Equation 5.2: Partial Log-Likelihood Function for Cox Proportional Hazards Model

Consider the partial likelihood function derived in Equation 5.1. Then the log-likelihood is given by:

$$\log\big(L(\boldsymbol{\beta})\big) = \log\left(\prod_{i=1}^{N}\left\{\frac{exp(\boldsymbol{\beta x}_i)}{\sum_{l\in R(t_i)} exp(\boldsymbol{\beta x}_l)}\right\}^{\delta_i}\right)$$

$$= \sum_{i=1}^{N} \log\left\{\frac{exp(\boldsymbol{\beta x}_i)}{\sum_{l\in R(t_i)} exp(\boldsymbol{\beta x}_l)}\right\}^{\delta_i}$$

$$= \sum_{i=1}^{N} \delta_i \log\left\{\frac{exp(\boldsymbol{\beta x}_i)}{\sum_{l\in R(t_i)} exp(\boldsymbol{\beta x}_l)}\right\}$$

$$= \sum_{i=1}^{N} \delta_i\{\,\boldsymbol{\beta x}_i - \log \sum_{l\in R(t_i)} exp(\boldsymbol{\beta x}_l)\}$$

The parameter estimates ($\widehat{\boldsymbol{\beta}}$), and their associated standard errors ($se(\widehat{\boldsymbol{\beta}})$) can be used to derive a test statistic to test the null hypothesis that $\beta_p = 0$ in the presence of the other $k-1$ explanatory variables. The test statistic is calculated and then, under the assumption it is normally distributed, compared to an appropriate critical value from the standard normal distribution (Equation 5.3). To implement a two-sided test at the α% significance level, the test statistic must be above the 1-α/2 upper-tail critical value or below the α/2 lower-tail critical value for a conclusion of statistical significance. For example, a two-sided test at the 5% level of significance would require a test statistic above 1.96 or below -1.96 to be concluded as statistically significant. However, in practice usually the duality of confidence intervals is used and for a two-sided test at the 5% level of significance, if 0 is not included in a two-sided 95% confidence interval for the parameter estimate (Equation 5.4) a conclusion of statistical significance is reached.

Equation 5.3: Cox Proportional Hazards Model Test Statistic

Let $\widehat{\beta_p}$ represent the parameter estimate of β_p from a Cox proportional hazards model with associated standard error $se(\widehat{\beta_p})$. Then the test statistic for the null hypothesis that $\beta_p = 0$ is given by:

$$\frac{\widehat{\beta_p}}{se(\widehat{\beta_p})} \sim N(0,1)$$

where $N(0,1)$ represents the standard normal distribution with mean zero and variance one.

Equation 5.4: 95% Confidence Interval for a Cox Proportional Hazards Parameter Estimate

Let $\widehat{\beta_p}$ represent the parameter estimate of β_p from a Cox proportional hazards model with associated standard error $se(\widehat{\beta_p})$. Then the 95% confidence interval is given by:

$$\widehat{\beta_p} \pm 1.96 \times se(\widehat{\beta_p})$$

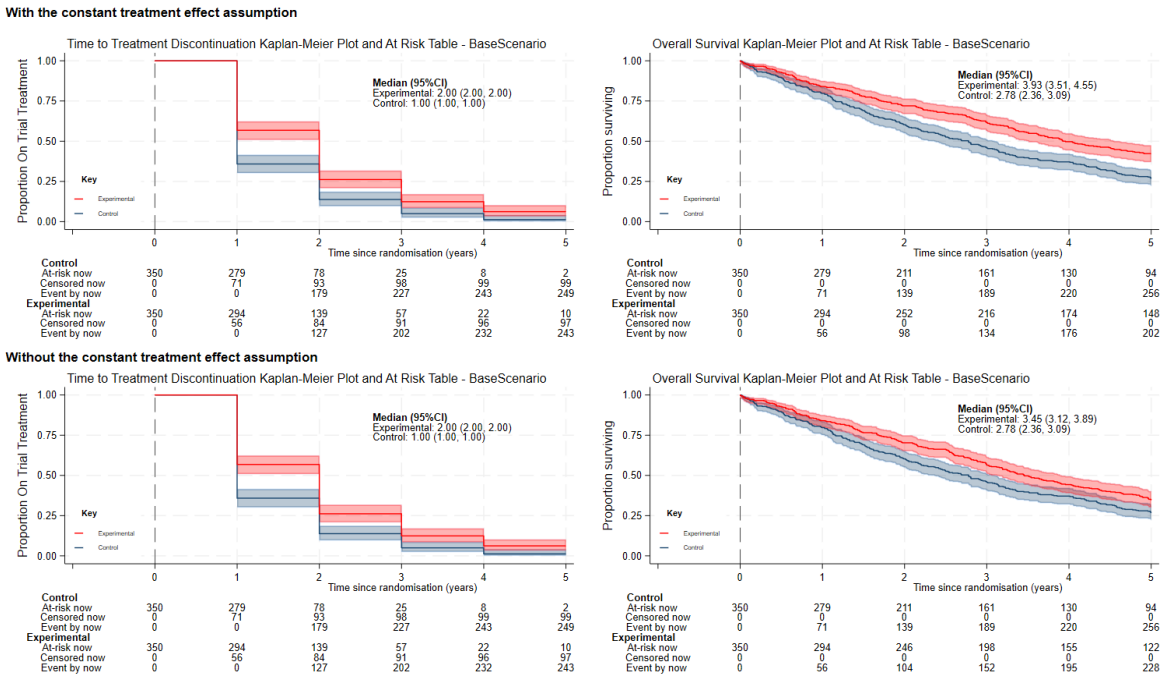
For parameter estimate interpretation, the estimates are exponentiated to the hazard scale from the log-hazard scale and interpreted as adjusted hazard ratios. For example, for a one unit increase in explanatory variable x_p the hazard of a participant experiencing the event of interest multiplies by $\exp(\widehat{\beta_p})$, provided all other explanatory variables remain constant.

D.2 Preliminary Investigations

D.2.1 Single Dataset

The Kaplan-Meier curves for overall survival and time to discontinuation resulting from the BaseScenario are shown in Appendix Figure D.1. As expected from the inputted parameters, discontinuation occurs at yearly intervals and was higher in the control compared to the experimental arm. Similarly, survival is better in the experimental arm and continues throughout follow-up. As expected, when the constant treatment assumption is relaxed, the improvement in survival lessens.

Appendix Figure D.1: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - BaseScenario with and without the constant treatment effect assumption



Following the implementation of a Cox model for OS from time zero to one year on one million participants it was concluded that, as all the inputted parameters are within the estimated confidence intervals, the DGM was working as expected (Appendix Table D.2).

Appendix Table D.2: Cox Model for OS from Time 0 to the First Follow-up Timepoint – One Million Participants Base Scenario with the constant treatment effect assumption.

	Coefficient*	Std. error	95% CI
Baseline Observed Covariate 1 - Binary			
Yes	1.154	0.009	(1.137, 1.172)
Baseline Observed Covariate 2 - Binary			
Yes	1.044	0.005	(1.034, 1.054)
Baseline Observed Covariate 3 - Categorical			
Level 2	1.046	0.005	(1.036, 1.057)
Level 3	1.095	0.013	(1.070, 1.121)
Baseline Unobserved Covariate 1 - Binary			
Yes	1.097	0.005	(1.087, 1.108)
Baseline Unobserved Covariate 2 - Continuous			
	1.053	0.003	(1.047, 1.057)
Randomisation Allocation			
Experimental	0.729	0.004	(0.722, 0.736)

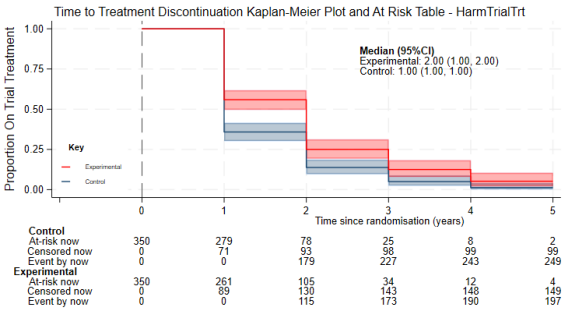
** Note this is exp(beta)*

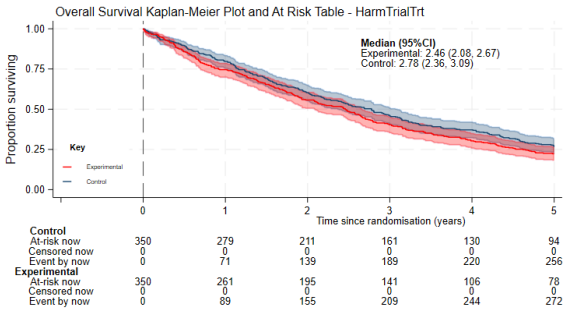
Similar curves to Appendix Figure D.1 for the other scenarios are presented in Appendix Figure D.2 to Appendix Figure D.11. Each of the descriptive statistics were reviewed and it was concluded the scenarios were being implemented as expected. For instance, in scenarios where the trial or subsequent treatment effect is changed (HarmTrialTrt - Appendix Figure D.2, NoEffectTrialTrt – Appendix Figure D.3, VeryBenTrialTrt – Appendix Figure D.4, VeryBenSubTrt – Appendix Figure D.11) as expected, more or less participants in the experimental arm die dependent on the scenario, which impacts the number of participants in the experimental arm at the end of follow-up and discontinuing treatment prior to death. Similarly, in scenarios where the time or proportion of discontinuation or number of follow-ups are changed (EarlyDiscontinuation – Appendix Figure D.8, LateDiscontinuation – Appendix Figure D.9, ThreeMonthlyFup – Appendix Figure D.5, MajorityExpTrtB – Appendix Figure D.6, MajorityDiscontExp – Appendix Figure D.7, MajorityDiscontA – Appendix Figure D.10) the time to discontinuation graph changes appropriately compared to the BaseScenario which impacts the number of participants alive at the end of follow-up.

In all scenarios apart HarmTrialTrt (Appendix Figure D.2) when the constant treatment assumption is relaxed it results in more participants dying. This is expected as the treatment effect is beneficial in all scenarios but HarmTrialTrt, and its impact has been reduced. Note that the constant treatment effect assumption has no impact on NoEffectTrialTrt so the summaries are the same and further results relaxing the assumption are not included.

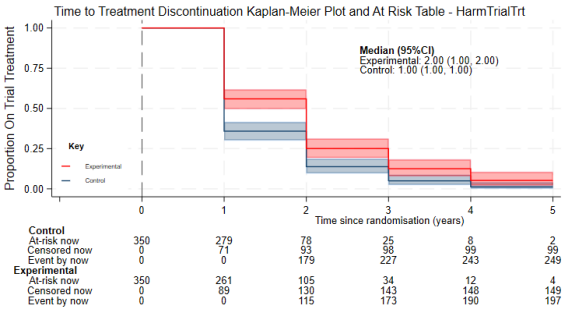
Appendix Figure D.2: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - HarmTrialTrt with and without the constant treatment effect assumption

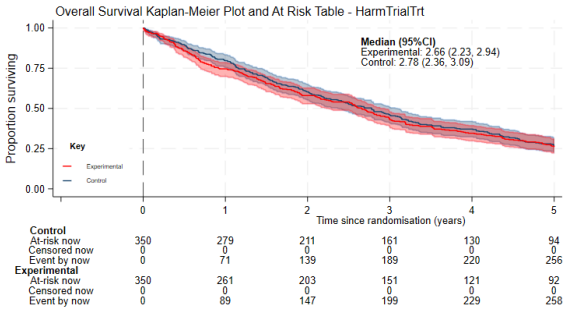
With the constant treatment effect assumption





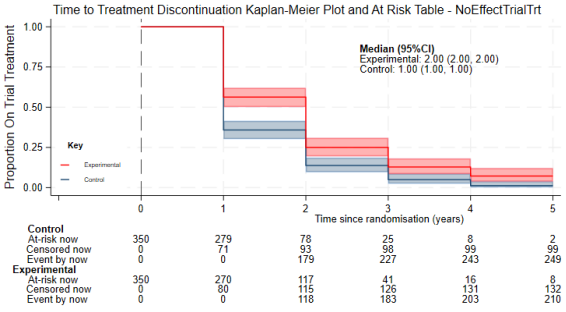
Without the constant treatment effect assumption

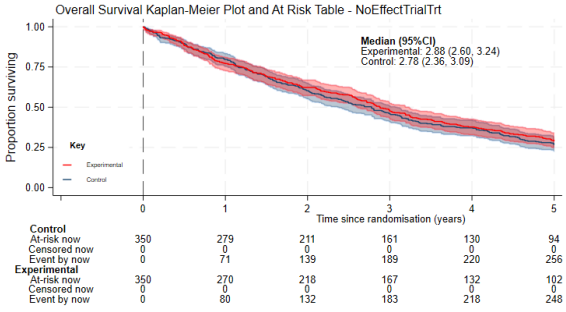




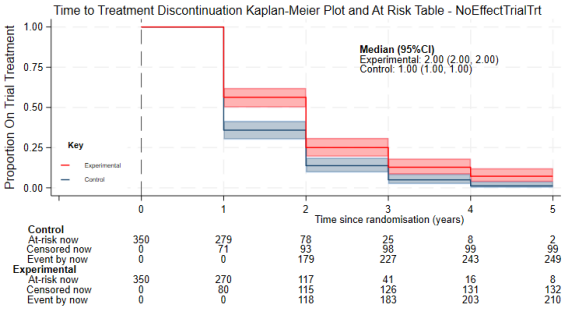
Appendix Figure D.3:Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - NoEffectTrialTrt with and without the constant treatment effect assumption

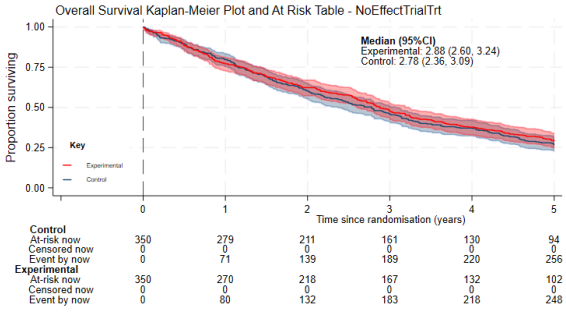
With the constant treatment effect assumption





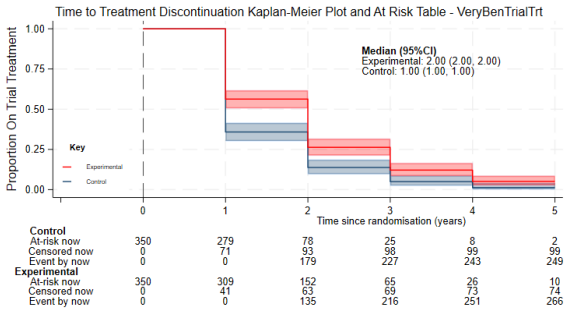
Without the constant treatment effect assumption

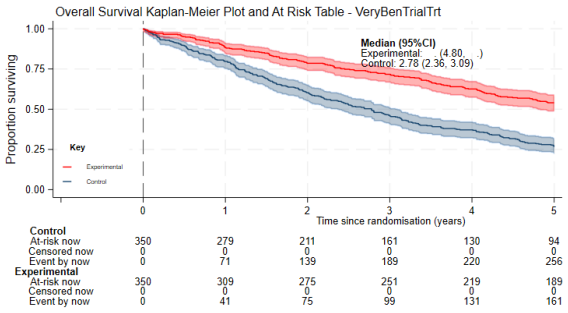




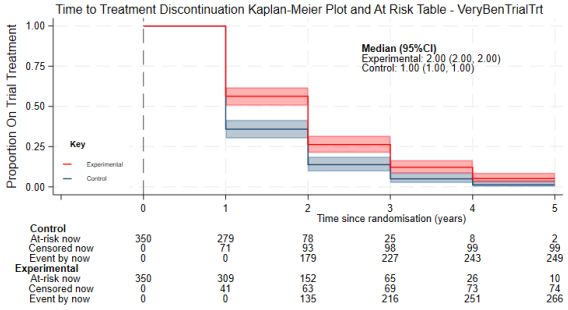
Appendix Figure D.4: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - VeryBenTrialTrt with and without the constant treatment effect assumption

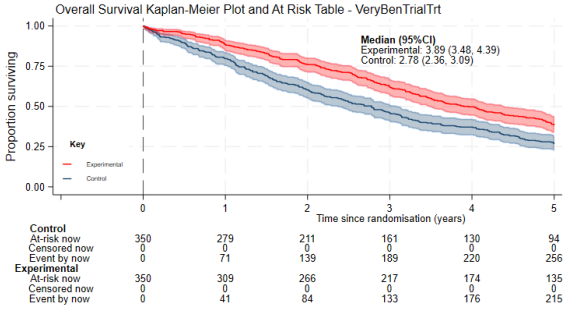
With the constant treatment effect assumption





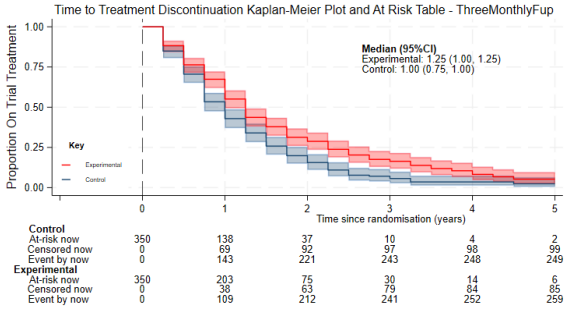
Without the constant treatment effect assumption

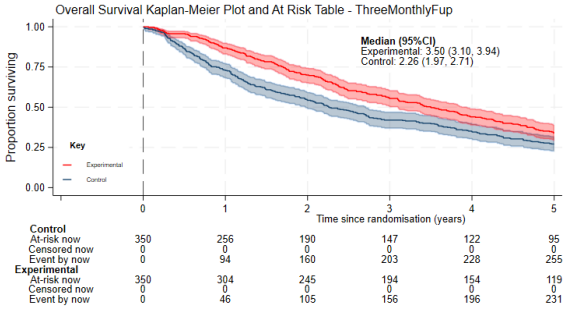




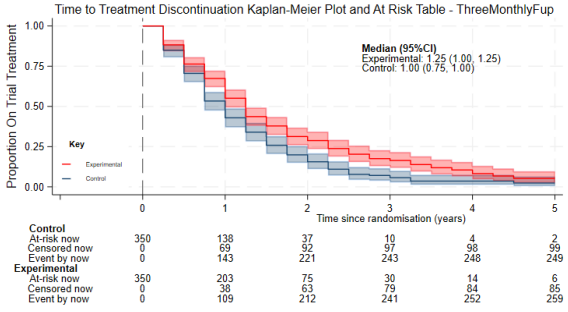
Appendix Figure D.5: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment -ThreeMonthlyFup with and without the constant treatment effect assumption

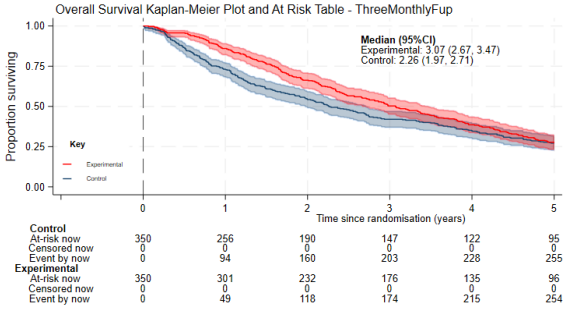
With the constant treatment effect assumption





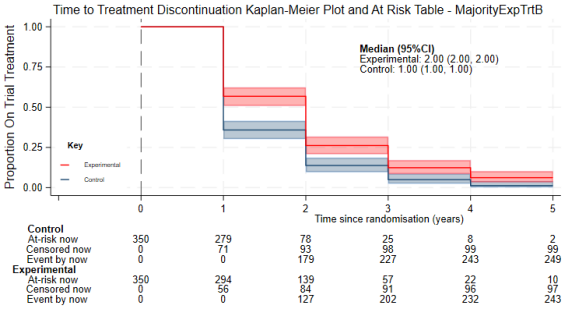
Without the constant treatment effect assumption

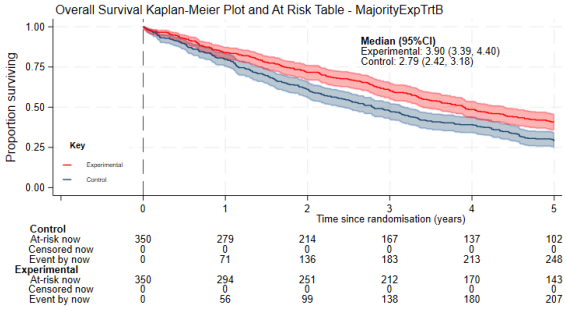




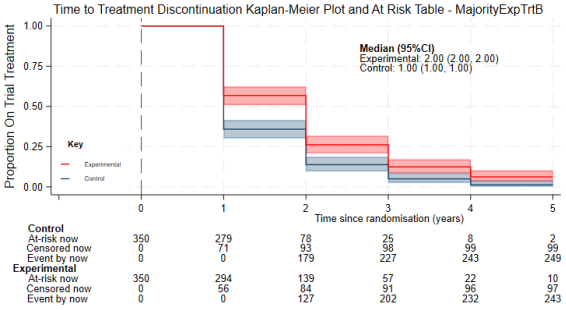
Appendix Figure D.6: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - MajorityExpTrtB with and without the constant treatment effect assumption

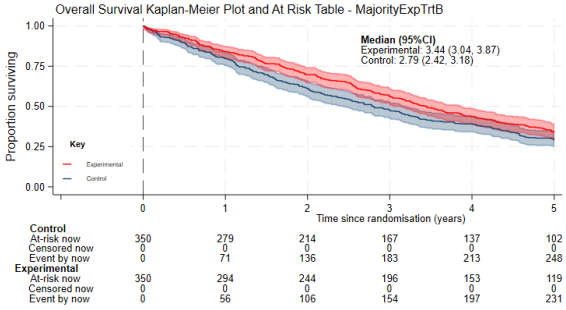
With the constant treatment effect assumption





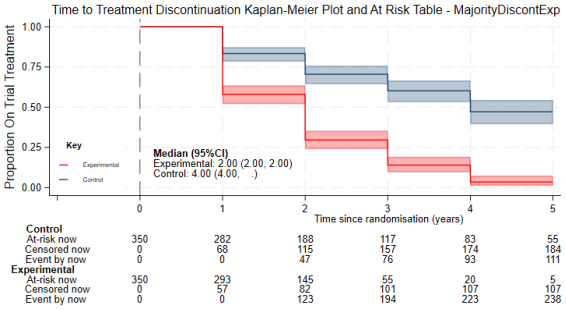
Without the constant treatment effect assumption

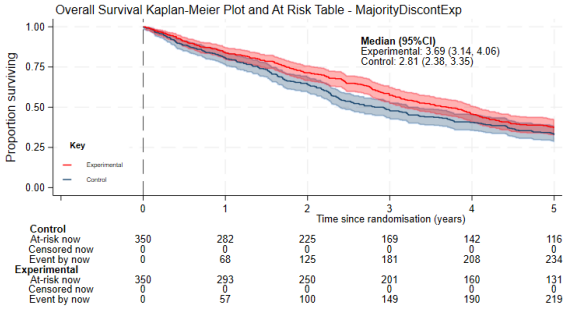




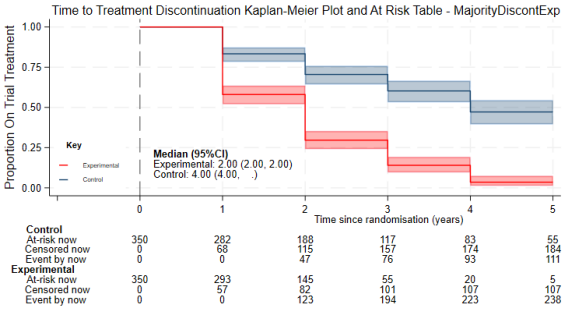
Appendix Figure D.7: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - MajorityDiscontExp with and without the constant treatment effect assumption

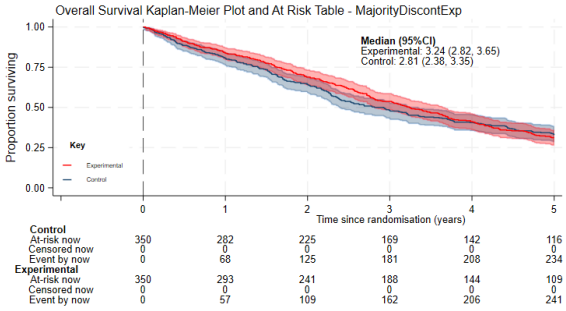
With the constant treatment effect assumption





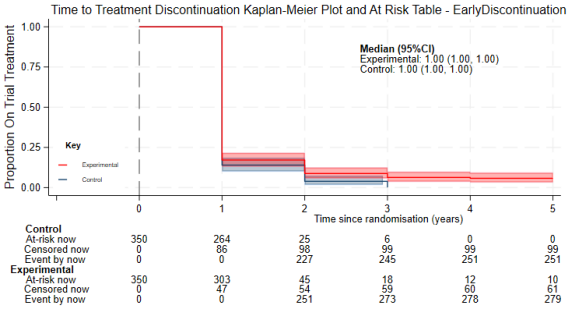
Without the constant treatment effect assumption

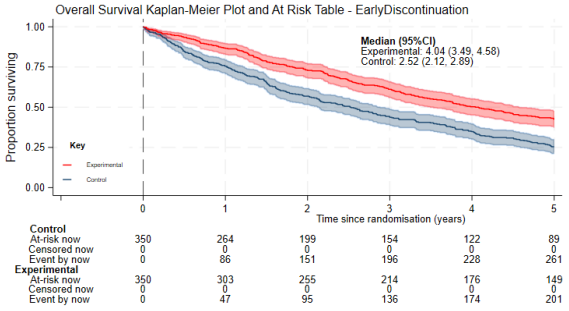




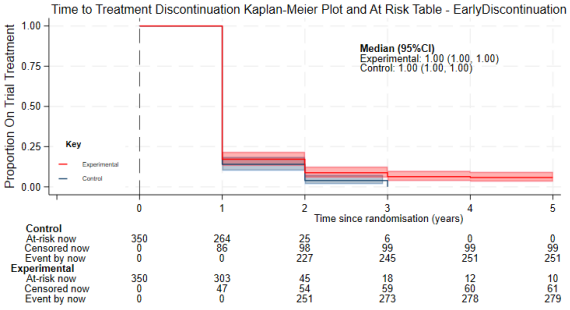
Appendix Figure D.8: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - EarlyDiscontinuation with and without the constant treatment effect assumption

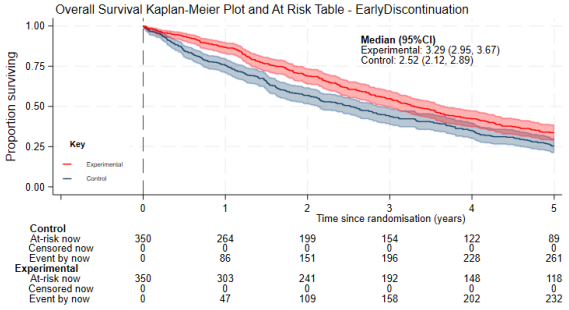
With the constant treatment effect assumption





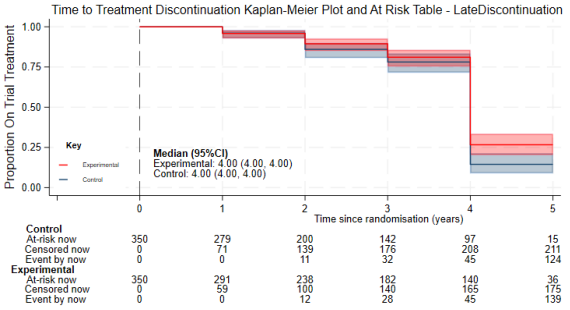
Without the constant treatment effect assumption

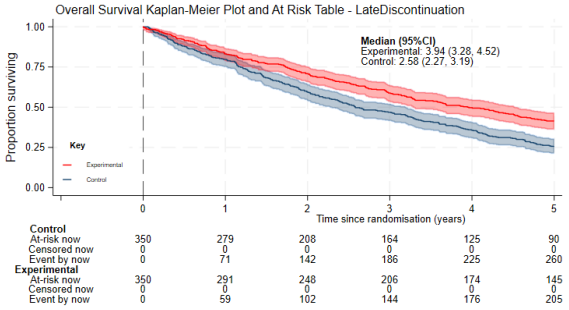




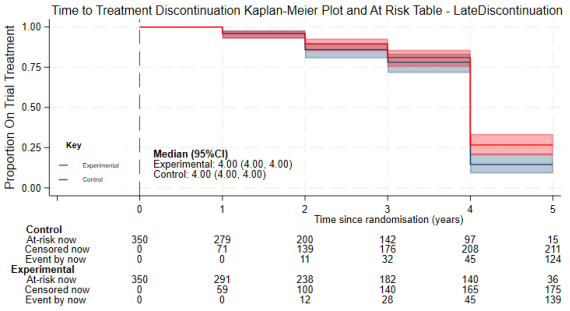
Appendix Figure D.9: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - LateDiscontinuation with and without the constant treatment effect assumption

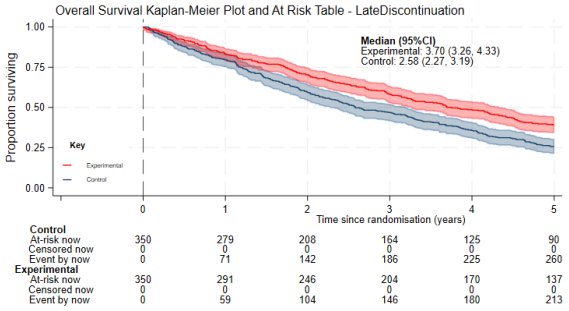
With the constant treatment effect assumption





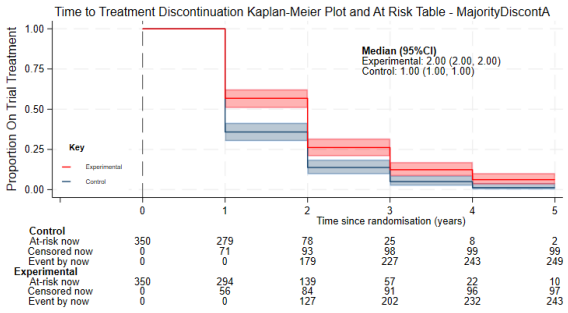
Without the constant treatment effect assumption

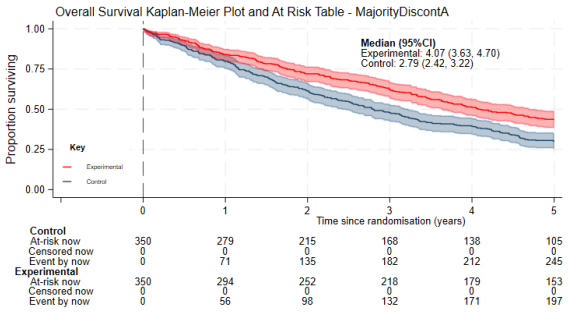




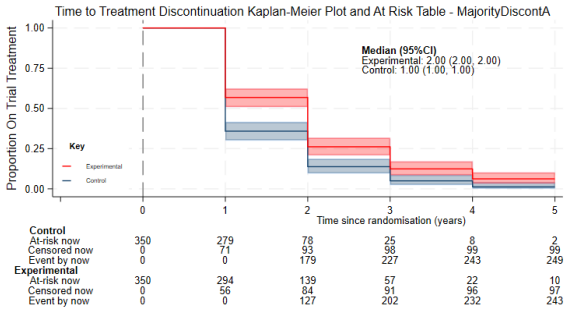
Appendix Figure D.10: Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - MajorityDiscontA with and without the constant treatment effect assumption

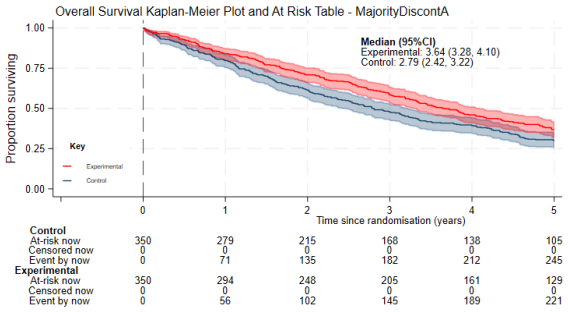
With the constant treatment effect assumption





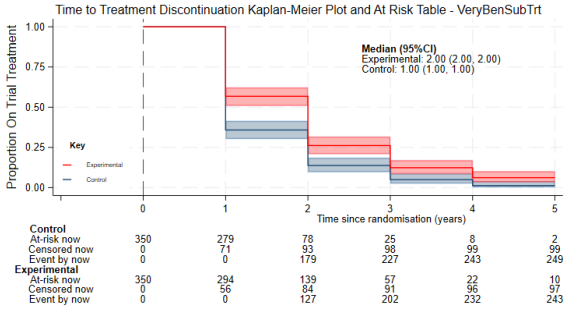
Without the constant treatment effect assumption

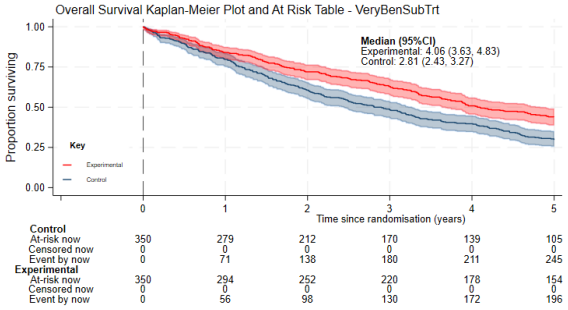




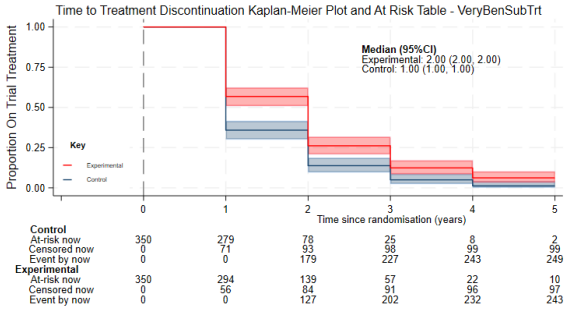
Appendix Figure D.11:Kaplan-Meier for Time to Treatment Discontinuation and Overall Survival by Randomised Treatment - VeryBenSubTrt with and without the constant treatment effect assumption

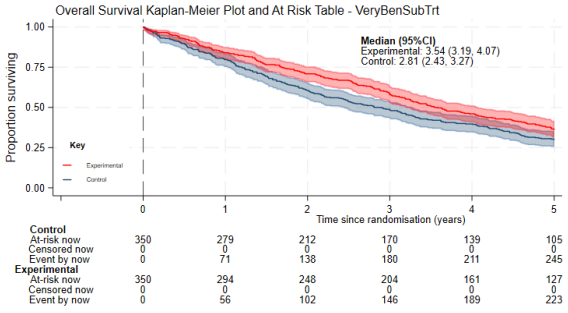
With the constant treatment effect assumption





Without the constant treatment effect assumption





D.2.2 True Hazard Ratio

The true hazard ratios for each scenario with and without the constant treatment effect assumption are shown in Appendix Table D.3. Recall these were derived using a numerical approach.

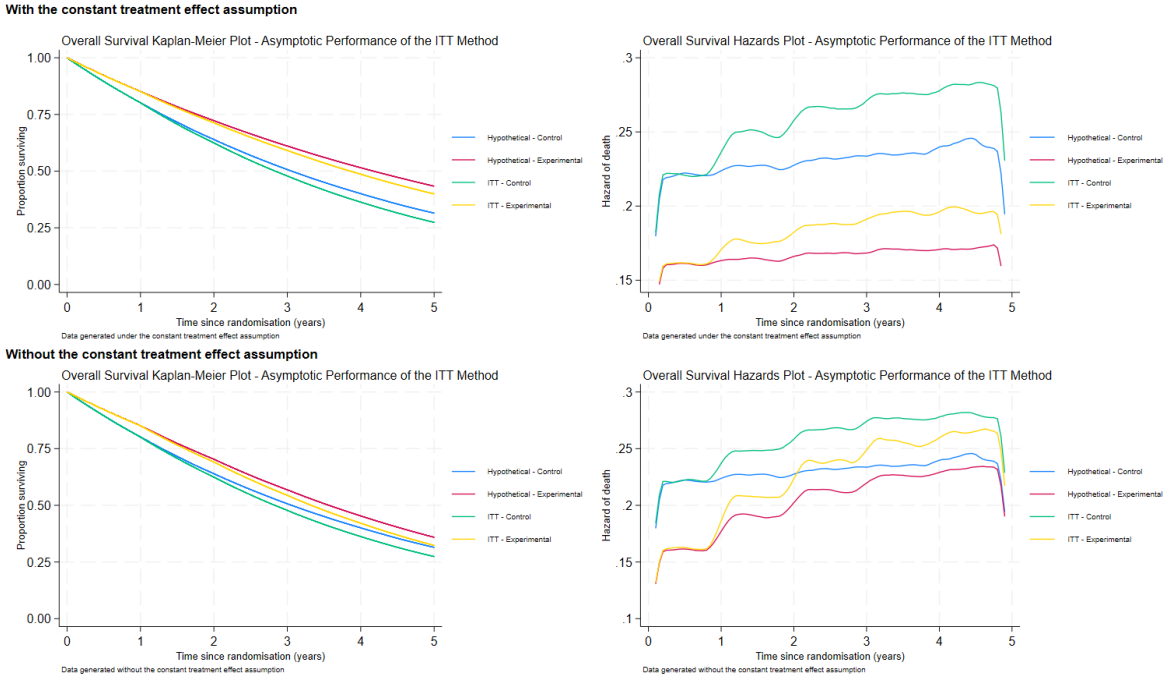
Appendix Table D.3: True Hazard Ratio for Each Scenario - With and without the constant treatment effect assumption.

	With the Constant Treatment Effect	Without the Constant Treatment Effect
BaseScenario	0.72	0.86
HarmTrialTrt	1.24	1.12
NoEffectTrialTrt	1.00	1.00
VeryBenTrialTrt	0.50	0.75
ThreeMonthlyFup	0.73	0.87
MajorityExpTrtB	0.72	0.86
MajorityDiscontExp	0.67	0.80
EarlyDiscontinuation	0.72	0.89
LateDiscontinuation	0.72	0.75
MajorityDiscontA	0.72	0.86
VeryBenSubTrt	0.74	0.86

D.2.3 Investigation into Asymptotic Performance

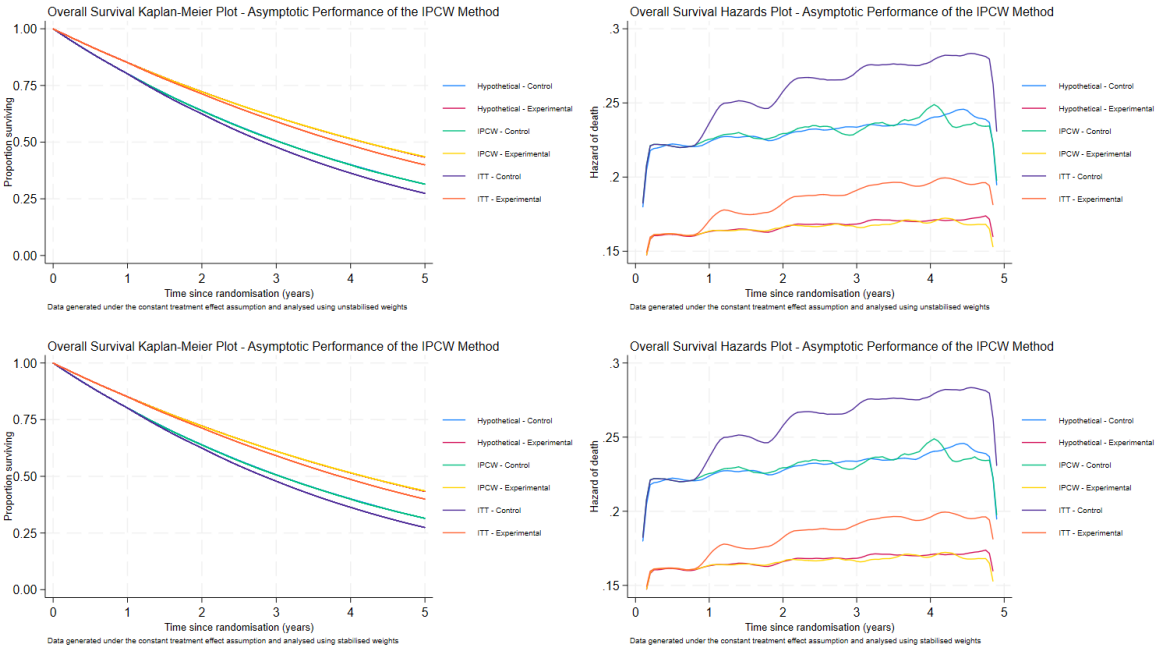
The investigation into asymptotic performance showed that the ITT method is biased following the first follow-up timepoint with and without the constant treatment effect assumption (Appendix Figure D.12). However, all methods (TSE, IPCW) corrected the bias in the asymptotic scenario with and without the constant treatment effect assumption (Appendix Figure D.13 - Appendix Figure D.15). Note that the stabilisation of weights made little difference to the performance of the IPCW method (Appendix Figure D.13).

Appendix Figure D.12: Asymptotic Performance of the ITT Method – With and without the constant treatment effect assumption

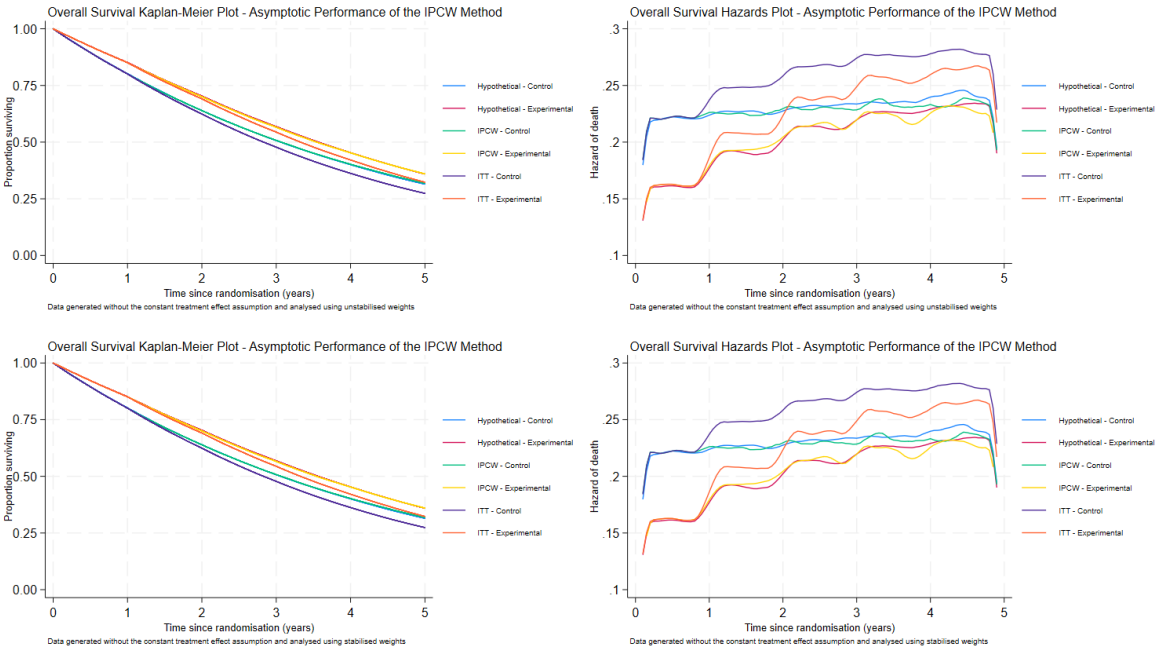


Appendix Figure D.13: Asymptotic Performance of the IPCW Method – With and without the constant treatment effect assumption

With the constant treatment effect assumption

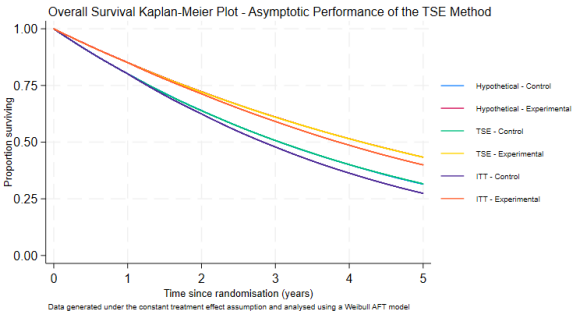


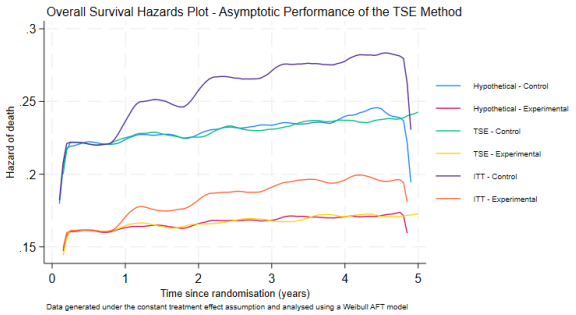
Without the constant treatment effect assumption

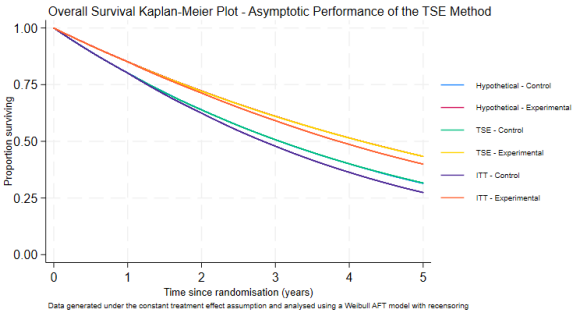


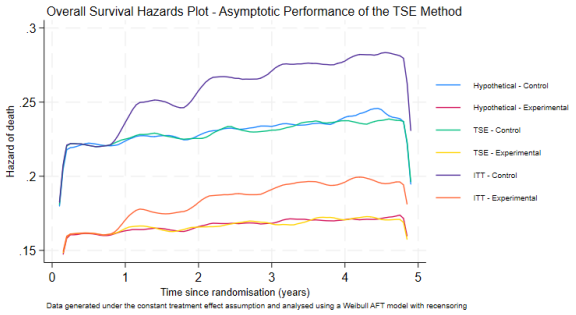
Appendix Figure D.14: Asymptotic Performance of the Two Stage Method using a Weibull AFT Model – With and without the constant treatment effect assumption

With the constant treatment effect assumption

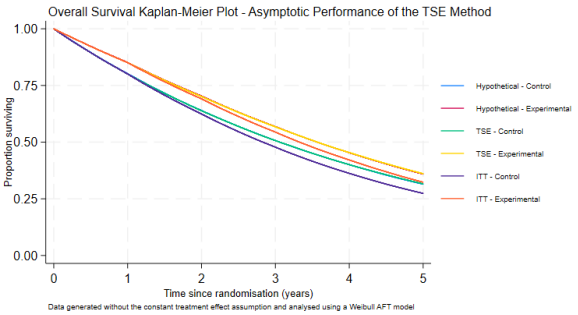


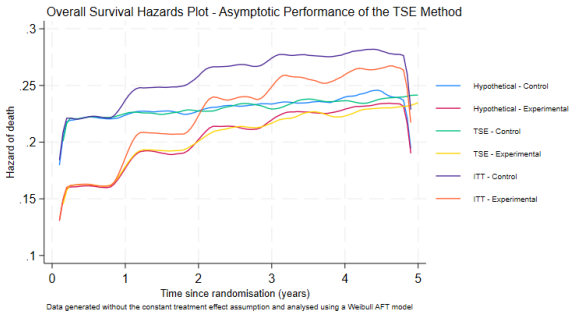


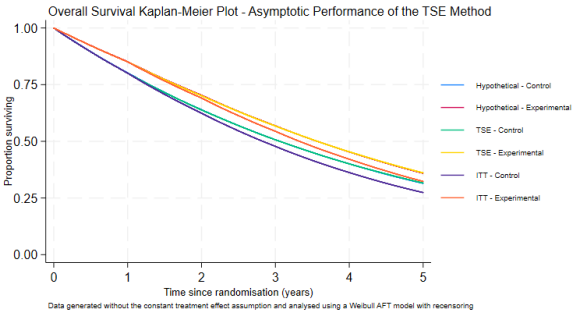


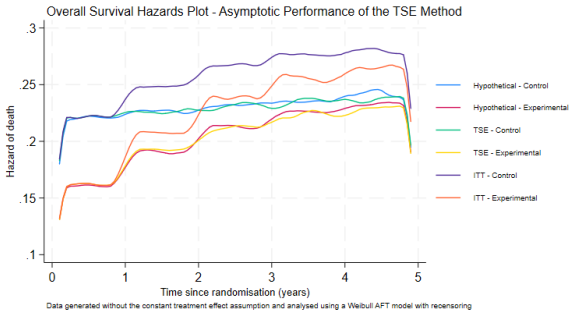


Without the constant treatment effect assumption



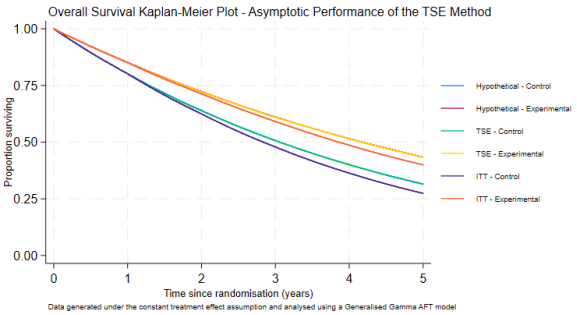


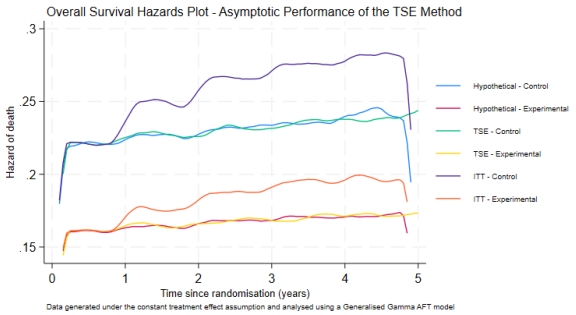


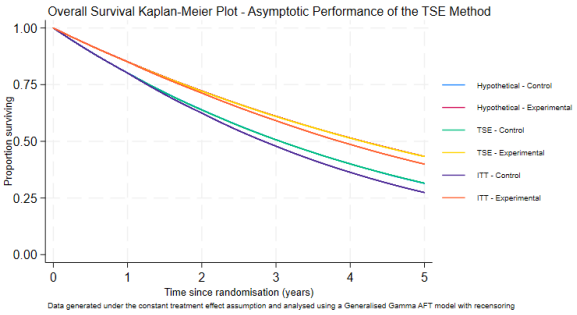


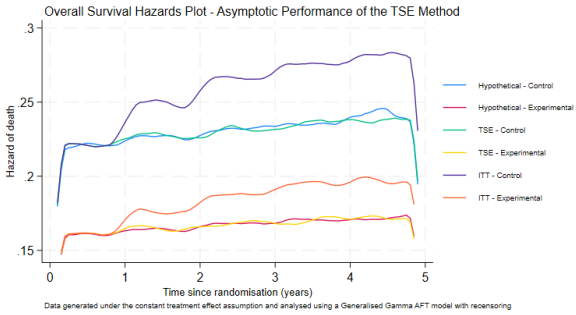
Appendix Figure D.15: Asymptotic Performance of the TSE Method using a Generalised Gamma AFT Model – With and without the constant treatment effect assumption

With the constant treatment effect assumption

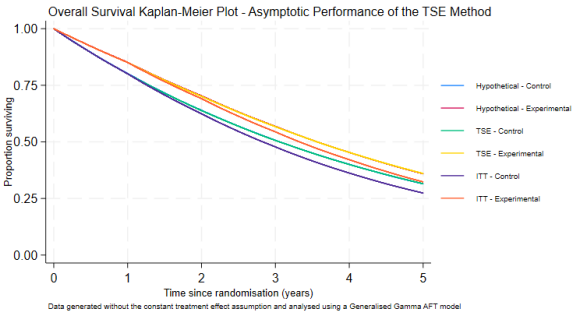


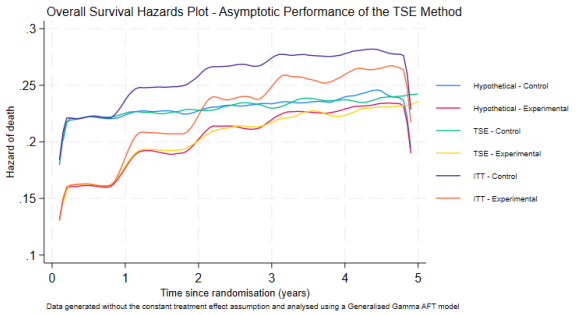


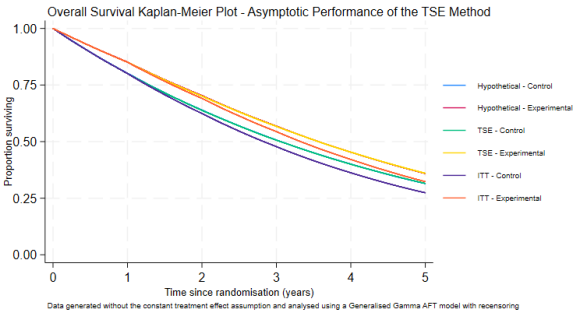


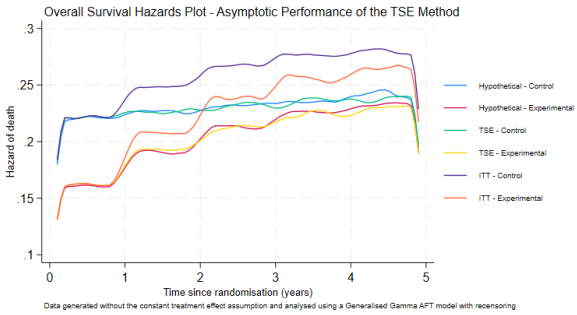


Without the constant treatment effect assumption







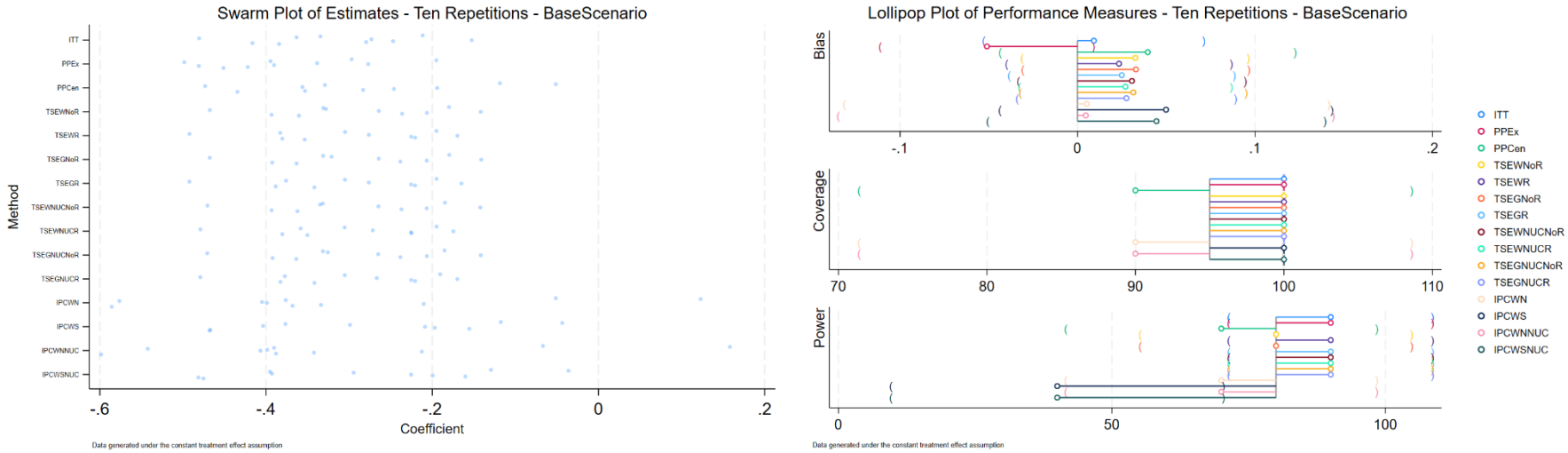


D.2.4 Small Simulation Study

This section provides the results of the steps conducted to check the implementation and replication of the simulation study for the base scenario. Note this was conducted under the constant treatment effect assumption.

The swarm plot of the estimates across the ten repetitions and lollipop plot of the associated performance parameters are shown in Appendix Figure D.16. From these plots it was concluded the simulation study was implementing correctly.

Appendix Figure D.16: Plots Investigating the Simulation Study Implementation - Base Scenario



The replication was tested considering repetition number 3. The parameter estimates from the simulation post-file are shown in Appendix Table D.4.

Appendix Table D.4: Parameter Estimates from Repetition 3/10 Base Scenario DGM

dgm	rep	method	coeff	se	lb	ub	N	events	php	aic	theta	mean	sd	min	max
BaseScenario	3	ITT	-0.363	0.094	-0.547	-0.179	700	461	0.433						
BaseScenario	3	PPEx	-0.395	0.122	-0.633	-0.156	375	274	0.450						
BaseScenario	3	PPCen	-0.329	0.122	-0.568	-0.090	700	274	0.754						
BaseScenario	3	TSEWNoR	-0.328	0.094	-0.512	-0.144	700	461	0.393	1275.970	1.499				
BaseScenario	3	TSEWR	-0.383	0.099	-0.577	-0.190	700	419	0.870						
BaseScenario	3	TSEGNoR	-0.321	0.094	-0.505	-0.137	700	461	0.386	1270.449	1.608				
BaseScenario	3	TSEGR	-0.376	0.100	-0.572	-0.181	700	410	0.850						
BaseScenario	3	TSEWNUCNoR	-0.335	0.094	-0.519	-0.151	700	461	0.406	1268.555	1.402				
BaseScenario	3	TSEWNUCR	-0.359	0.097	-0.549	-0.168	700	429	0.626						
BaseScenario	3	TSEGNUCNoR	-0.326	0.094	-0.510	-0.142	700	461	0.390	1263.432	1.535				
BaseScenario	3	TSEGNUCR	-0.378	0.099	-0.572	-0.184	700	416	0.918						
BaseScenario	3	IPCWN	-0.376	0.155	-0.680	-0.073	2539	410	0.018			1.498	1.373	1.000	11.705
BaseScenario	3	IPCWS	-0.377	0.137	-0.645	-0.109	1445	241	0.073			0.734	0.615	0.119	8.578
BaseScenario	3	IPCWNNUC	-0.391	0.154	-0.693	-0.088	2527	411	0.064			1.496	1.408	1.000	16.936
BaseScenario	3	IPCWSNUC	-0.395	0.137	-0.664	-0.127	1441	243	0.128			0.734	0.629	0.119	12.411

***Dgm**: Data generating mechanism; **rep**: repetition number; **method**: implemented method; **coeff**: randomisation allocation coefficient from cox model adjusted by stratification factors; **se**: standard error for coeff; **lb**: 95% lower bound confidence interval limit for coeff; **ub**: 95% upper bound confidence interval limit for coeff; **N**: Number of participants; **events**: Number of events; **php**: p-value for proportional hazards assessment of randomisation allocation in the cox model; **aic**: Akaike information criterion from the AFT model; **theta**: coefficient of subsequent anti-cancer therapy from the AFT model; **mean**: mean of weights used in the cox model; **sd**: standard deviation of the weights used in the cox model; **min**: minimum value of the weights used in the cox model; **max**: maximum value of weights used in the cox model.*

The parameter estimates from the replication check are shown in Appendix Table D.5. As the results are identical the replication was concluded to be successful.

Appendix Table D.5: Parameter Estimates from Base Scenario Replication Check

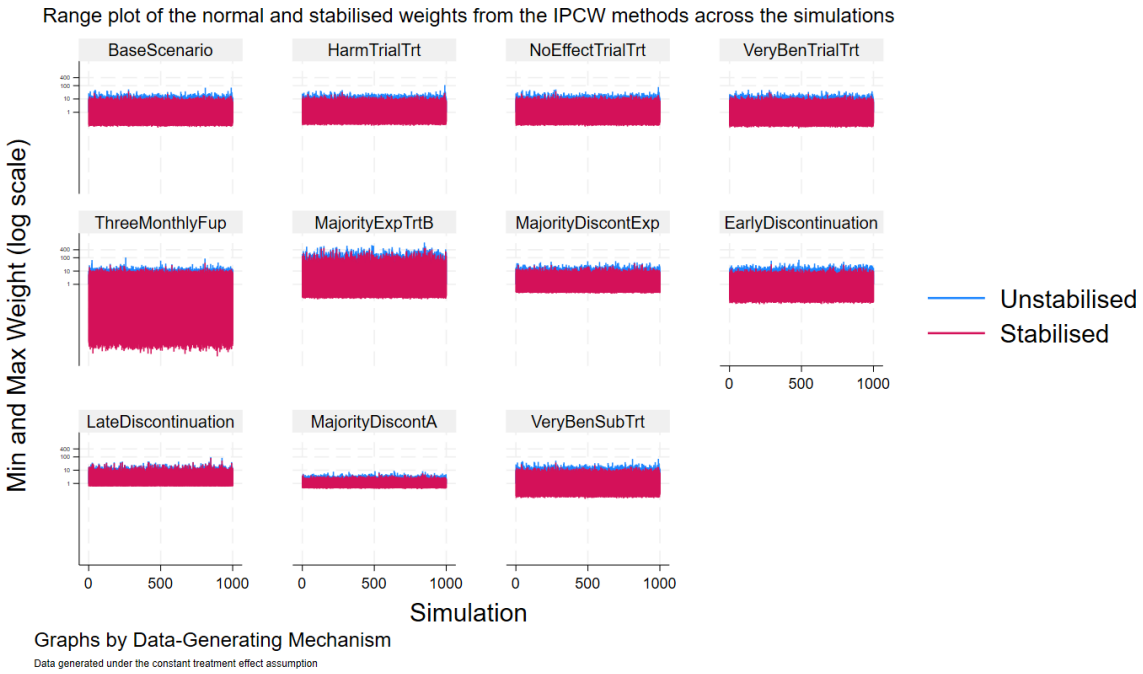
dgm	rep	method	coeff	se	lb	ub	N	events	php	aic	theta	mean	sd	min	max
BaseScenario_Chk	3	ITT	-0.363	0.094	-0.547	-0.179	700	461	0.433						
BaseScenario_Chk	3	PPEx	-0.395	0.122	-0.633	-0.156	375	274	0.450						
BaseScenario_Chk	3	PPCen	-0.329	0.122	-0.568	-0.090	700	274	0.754						
BaseScenario_Chk	3	TSEWNoR	-0.328	0.094	-0.512	-0.144	700	461	0.393	1275.970	1.499				
BaseScenario_Chk	3	TSEWR	-0.383	0.099	-0.577	-0.190	700	419	0.870						
BaseScenario_Chk	3	TSEGNoR	-0.321	0.094	-0.505	-0.137	700	461	0.386	1270.449	1.608				
BaseScenario_Chk	3	TSEGR	-0.376	0.100	-0.572	-0.181	700	410	0.850						
BaseScenario_Chk	3	TSEWNUCNoR	-0.335	0.094	-0.519	-0.151	700	461	0.406	1268.555	1.402				
BaseScenario_Chk	3	TSEWNUCR	-0.359	0.097	-0.549	-0.168	700	429	0.626						
BaseScenario_Chk	3	TSEGNUCNoR	-0.326	0.094	-0.510	-0.142	700	461	0.390	1263.432	1.535				
BaseScenario_Chk	3	TSEGNUCR	-0.378	0.099	-0.572	-0.184	700	416	0.918						
BaseScenario_Chk	3	IPCWN	-0.376	0.155	-0.680	-0.073	2539	410	0.018			1.498	1.373	1.000	11.705
BaseScenario_Chk	3	IPCWS	-0.377	0.137	-0.645	-0.109	1445	241	0.073			0.734	0.615	0.119	8.578
BaseScenario_Chk	3	IPCWNNUC	-0.391	0.154	-0.693	-0.088	2527	411	0.064			1.496	1.408	1.000	16.936
BaseScenario_Chk	3	IPCWSNUC	-0.395	0.137	-0.664	-0.127	1441	243	0.128			0.734	0.629	0.119	12.411

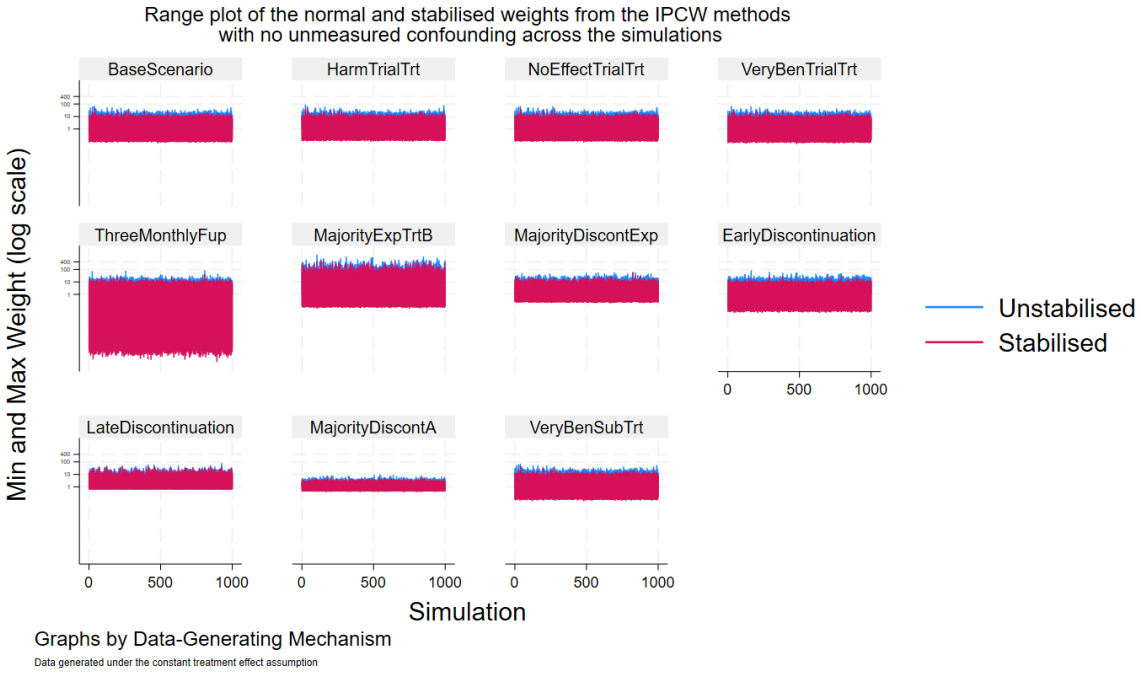
***Dgm**: Data generating mechanism; **rep**: repetition number; **method**: implemented method; **coeff**: randomisation allocation coefficient from cox model adjusted by stratification factors; **se**: standard error for coeff; **lb**: 95% lower bound confidence interval limit for coeff; **ub**: 95% upper bound confidence interval limit for coeff; **N**: Number of participants; **events**: Number of events; **php**: p-value for proportional hazards assessment of randomisation allocation in the cox model; **aic**: Akaike information criterion from the AFT model; **theta**: coefficient of subsequent anti-cancer therapy from the AFT model; **mean**: mean of weights used in the cox model; **sd**: standard deviation of the weights used in the cox model; **min**: minimum value of the weights used in the cox model; **max**: maximum value of weights used in the cox model.*

D.3 Additional Simulation Results – Under the Constant Treatment Effect Assumption

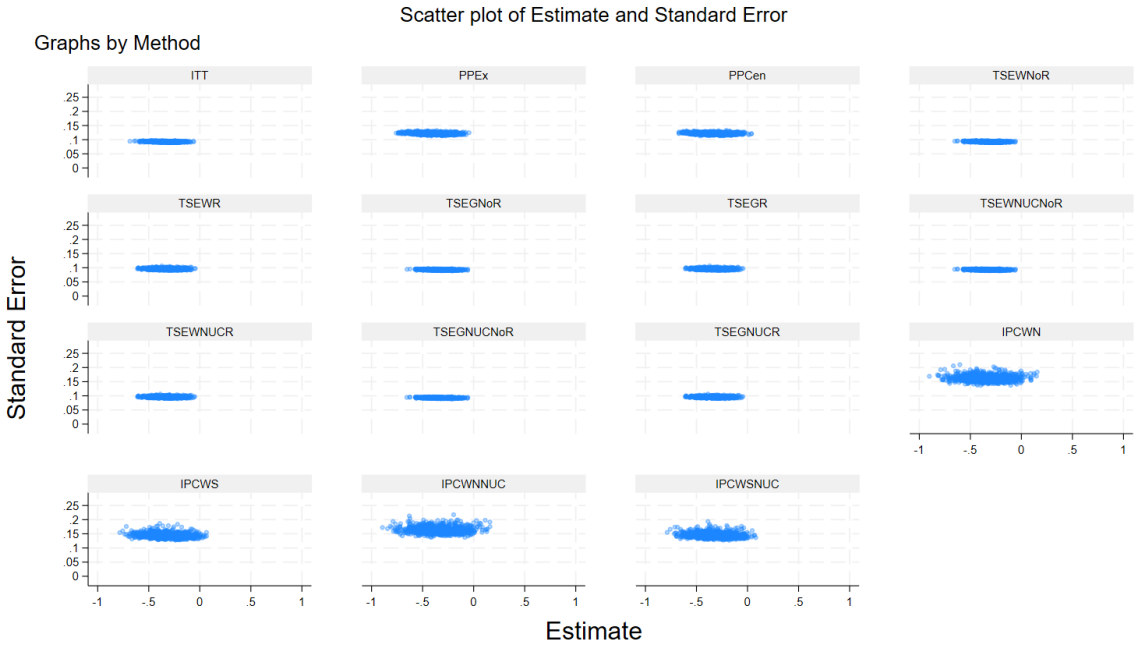
This section provides the additional figures and tables from the simulation study scenarios generated under the constant treatment effect assumption.

Appendix Figure D.17: Range Plot of the Unstabilised and Stabilised Weights from the IPCW Methods Across the Simulations and Scenarios - With the constant treatment effect assumption

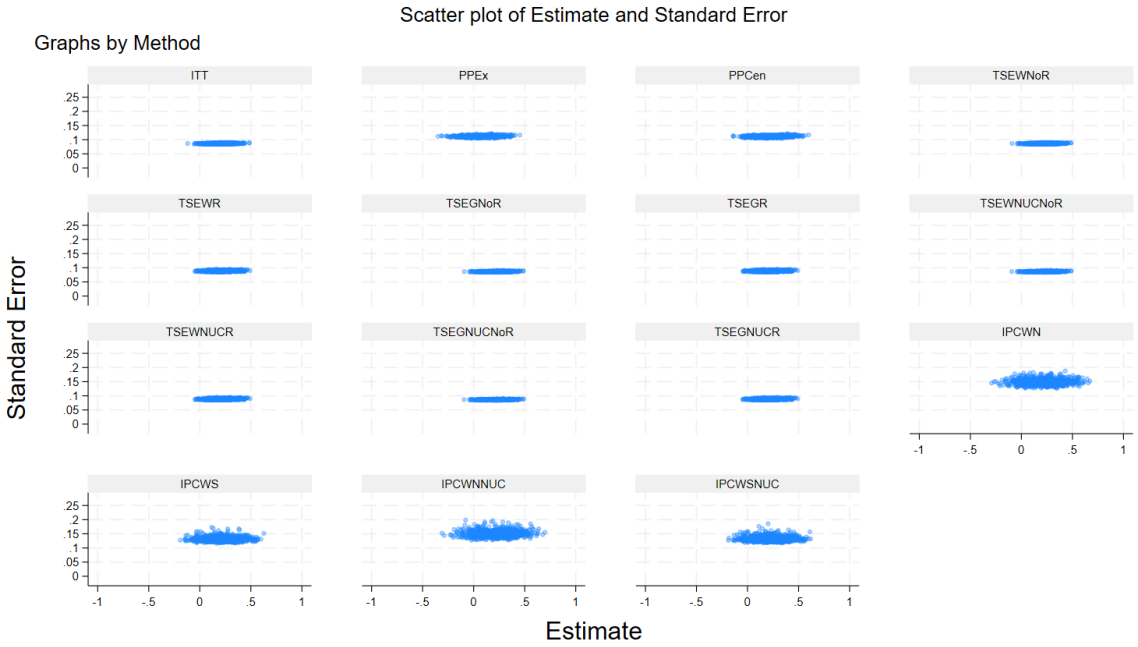



Range plot of the normal and stabilised weights from the IPCW methods
with no unmeasured confounding across the simulations
BaseScenario
HarmTrialTrt
NoEffectTrialTrt
VeryBenTrialTrt
ThreeMonthlyFup
MajorityExpTrtB
MajorityDiscontExp
EarlyDiscontinuation
LateDiscontinuation
MajorityDiscontA
VeryBenSubTrt
Min and Max Weight (log scale)
400
100
10
1
0
500
1000
Simulation
Unstabilised
Stabilised
Graphs by Data-Generating Mechanism
Data generated under the constant treatment effect assumption

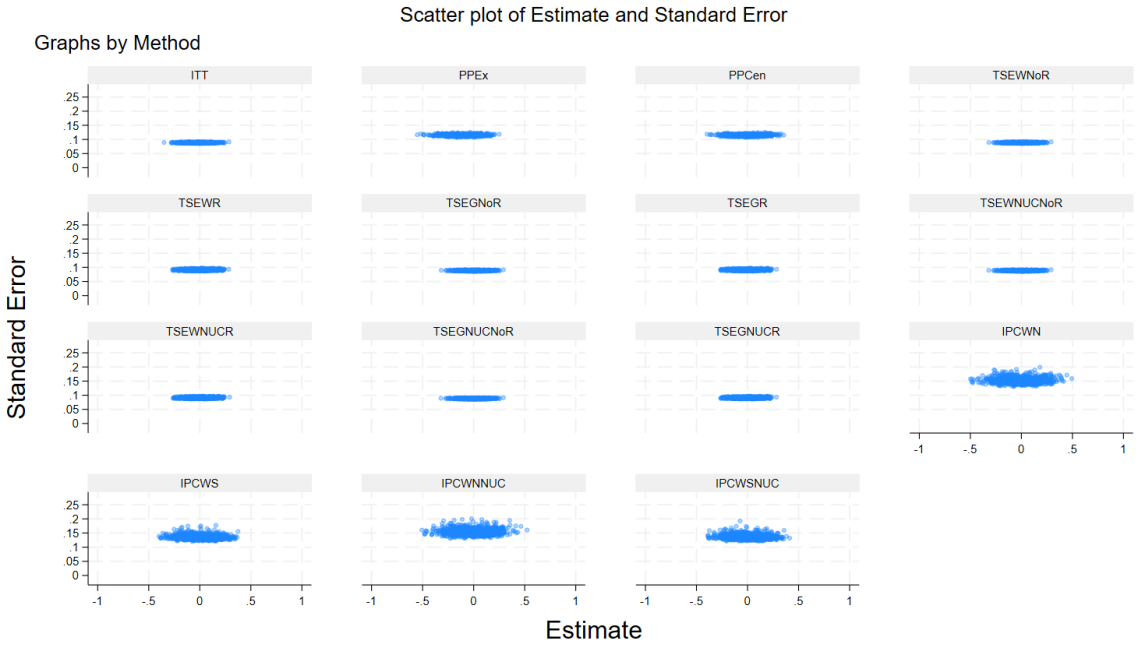
Appendix Figure D.18: Scatter Plots of the Estimates and their Standard Errors Across All Scenarios - With the constant treatment effect assumption



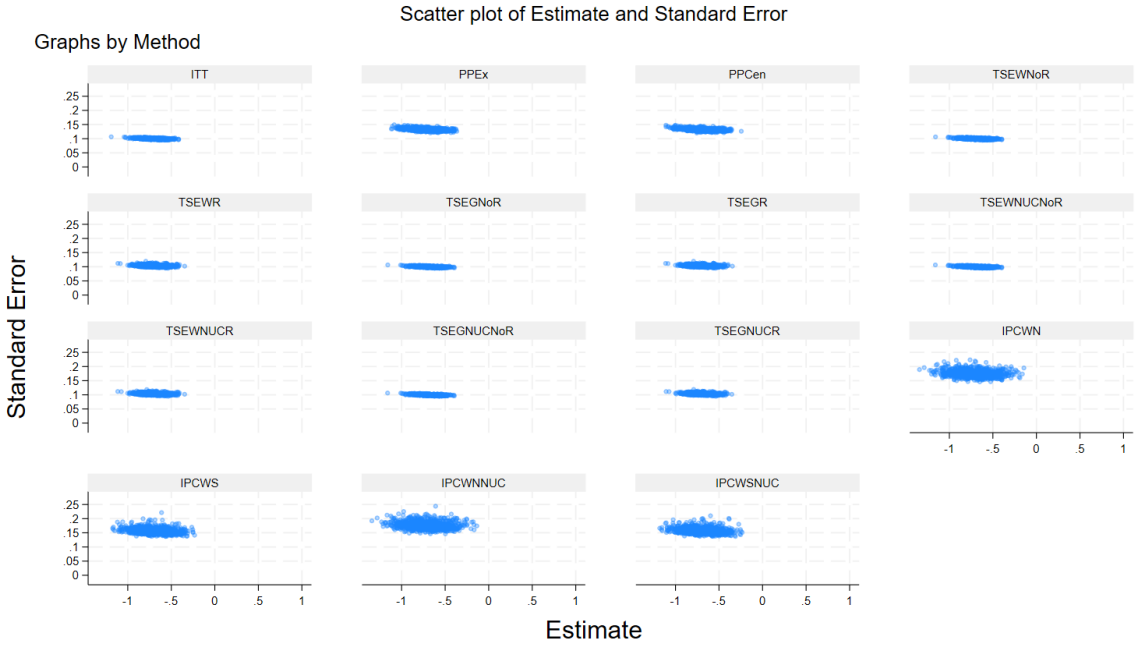
Data generated under the constant treatment effect assumption: BaseScenario



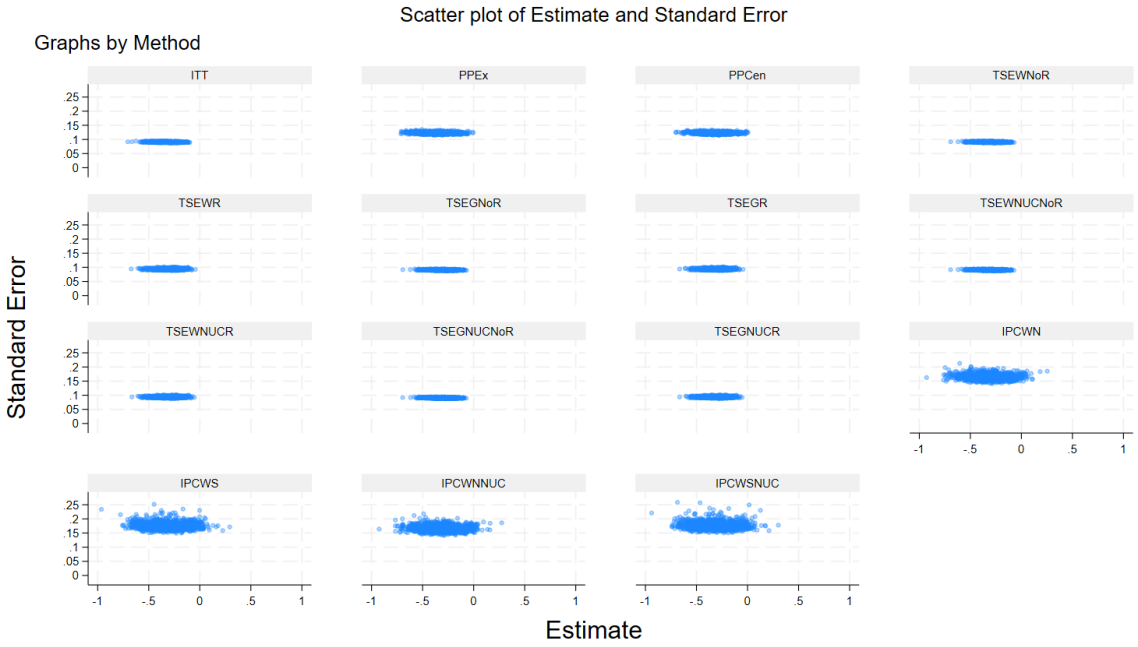
Data generated under the constant treatment effect assumption: HarmTrialTrt



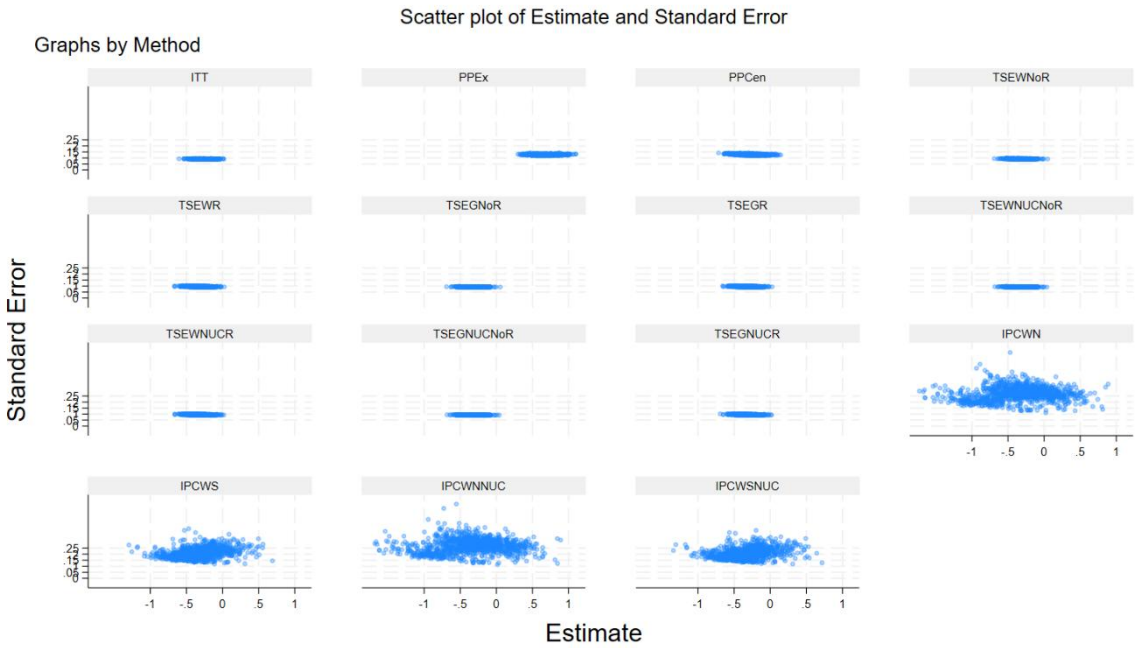
Data generated under the constant treatment effect assumption: NoEffectTrialTrt



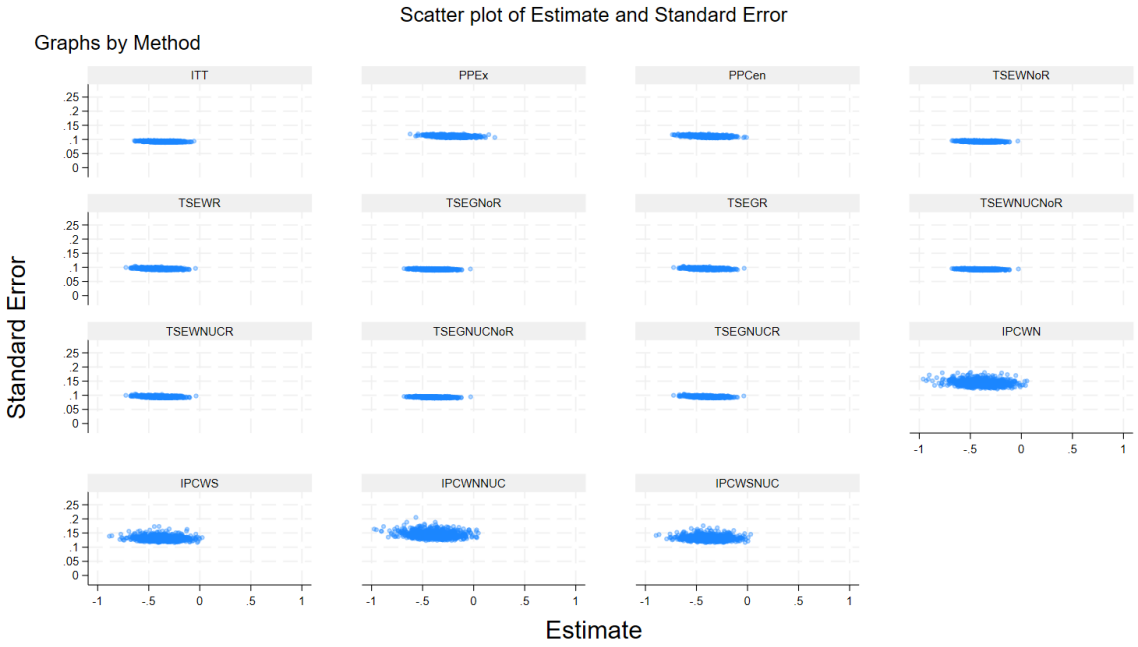
Data generated under the constant treatment effect assumption: VeryBenTrialTrt



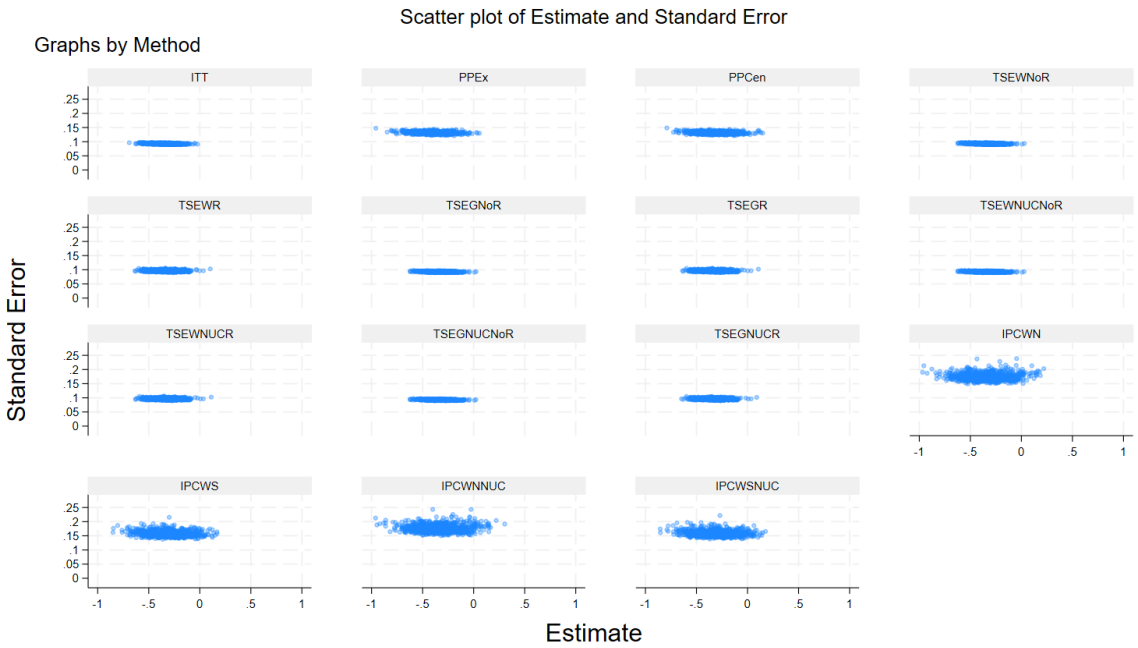
Data generated under the constant treatment effect assumption: ThreeMonthlyFup



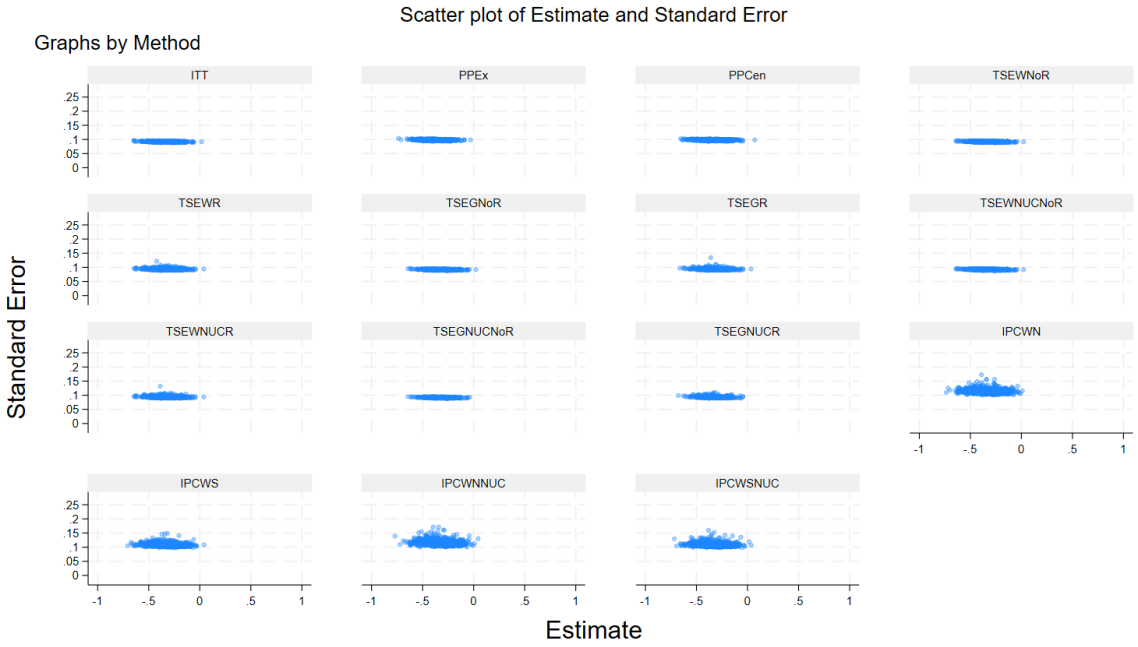
Data generated under the constant treatment effect assumption: MajorityExpTrtB



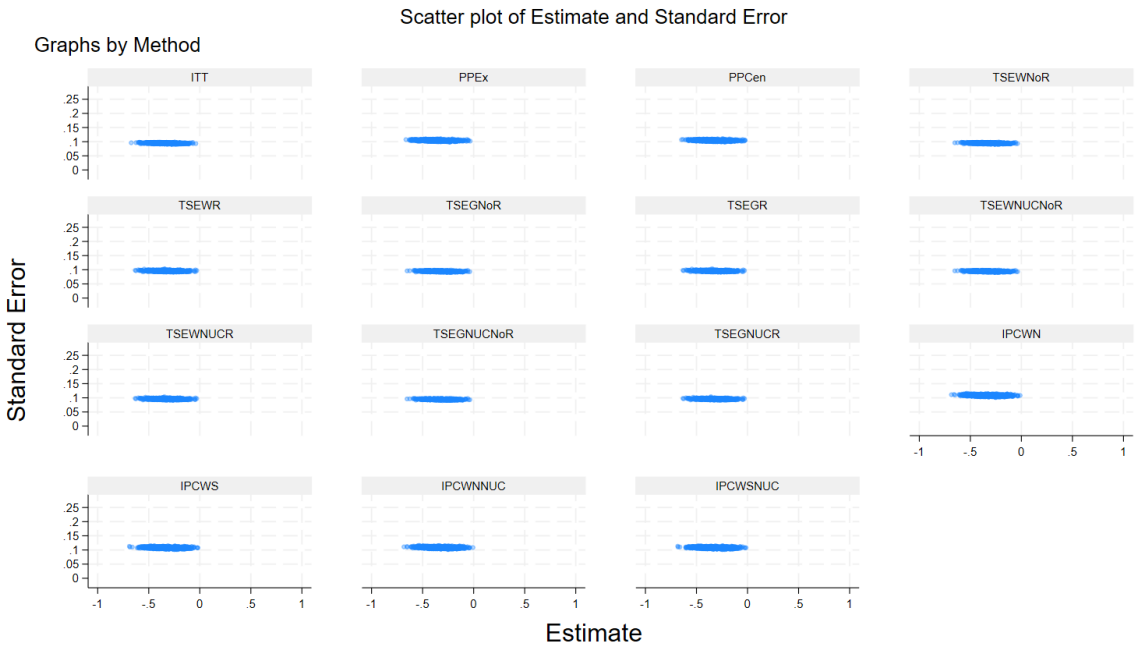
Data generated under the constant treatment effect assumption: MajorityDiscontExp



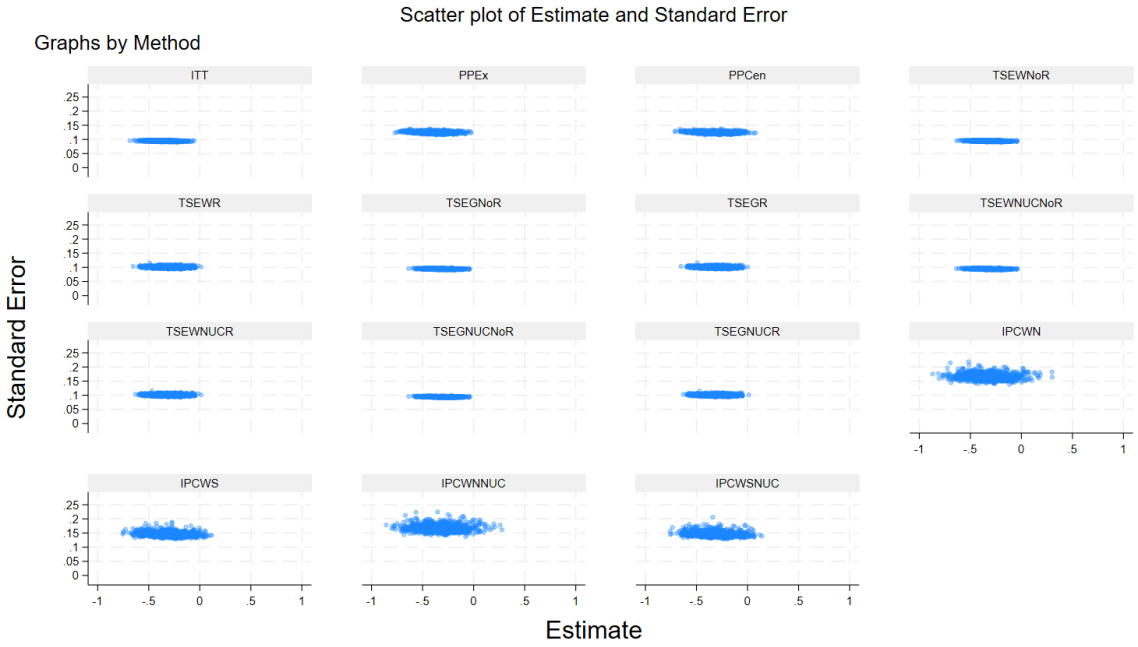
Data generated under the constant treatment effect assumption: EarlyDiscontinuation



Data generated under the constant treatment effect assumption: LateDiscontinuation



Data generated under the constant treatment effect assumption: MajorityDiscontA



Data generated under the constant treatment effect assumption: VeryBenSubTrt

Appendix Table D.6: AIC Goodness of Fit Comparisons for TSE AFT Models

	TSE: Unmeasured Confounding			TSE: No Unmeasured Confounding		
	Gamma < Weibull	**Gamma ≥ Weibull**	**N/A**	**Gamma < Weibull**	**Gamma ≥ Weibull**	**N/A**
BaseScenario	179 (17.90%)	774 (77.40%)	47 (4.70%)	163 (16.30%)	748 (74.80%)	89 (8.90%)
HarmTrialTrt	199 (19.90%)	780 (78.00%)	21 (2.10%)	201 (20.10%)	748 (74.80%)	51 (5.10%)
NoEffectTrialTrt	196 (19.60%)	780 (78.00%)	24 (2.40%)	188 (18.80%)	750 (75.00%)	62 (6.20%)
VeryBenTrialTrt	151 (15.10%)	774 (77.40%)	75 (7.50%)	139 (13.90%)	738 (73.80%)	123 (12.30%)
ThreeMonthlyFup	186 (18.60%)	793 (79.30%)	21 (2.10%)	180 (18.00%)	788 (78.80%)	32 (3.20%)
MajorityExpTrtB	185 (18.50%)	778 (77.80%)	37 (3.70%)	175 (17.50%)	744 (74.40%)	81 (8.10%)
MajorityDiscontExp	152 (15.20%)	760 (76.00%)	88 (8.80%)	141 (14.10%)	681 (68.10%)	178 (17.80%)
EarlyDiscontinuation	157 (15.70%)	805 (80.50%)	38 (3.80%)	158 (15.80%)	773 (77.30%)	69 (6.90%)
LateDiscontinuation	117 (11.70%)	579 (57.90%)	304 (30.40%)	94 (9.40%)	442 (44.20%)	464 (46.40%)
MajorityDiscontA	172 (17.20%)	774 (77.40%)	54 (5.40%)	144 (14.40%)	739 (73.90%)	117 (11.70%)
VeryBenSubTrt	170 (17.00%)	775 (77.50%)	55 (5.50%)	155 (15.50%)	734 (73.40%)	111 (11.10%)

N/A: Indicates that the generalised gamma model did not converge so no comparison can be made. Percentages have been calculated out of the 1000 repetitions for each method within in scenario.

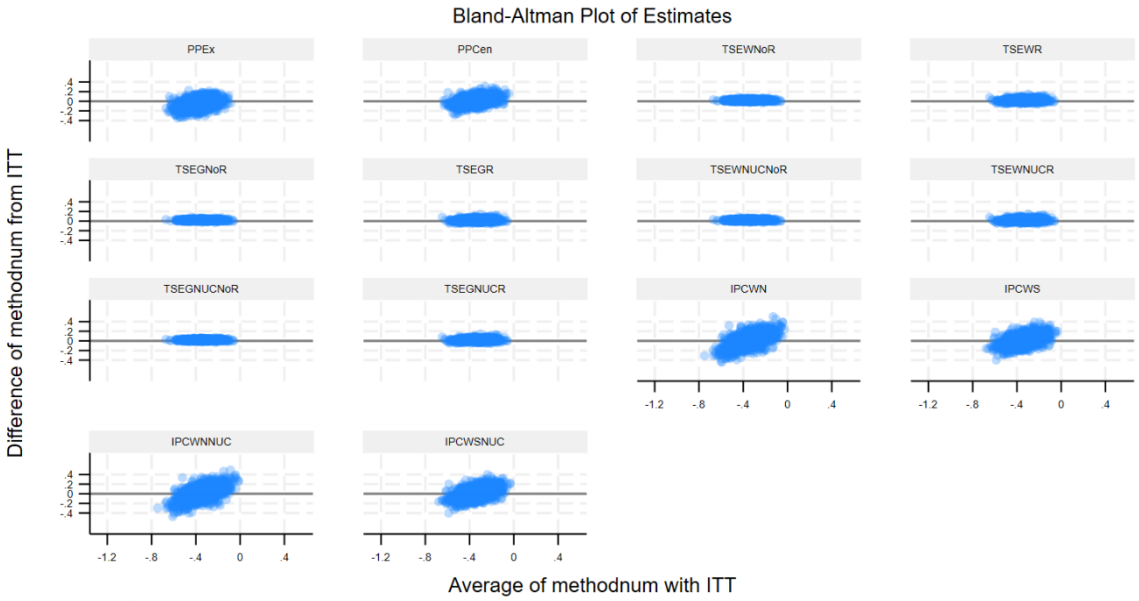
Appendix Table D.7: Mean (SD) of Theta from the TSE Methods Calculated Across the Simulations

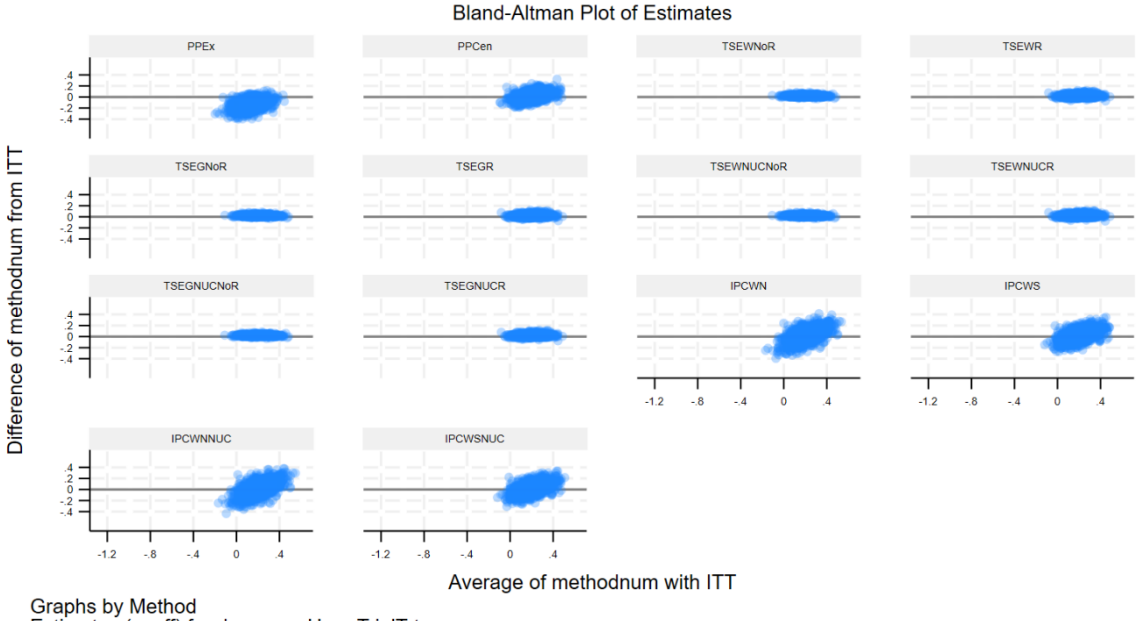
	TSEGNUCNoR	TSEGNoR	TSEWNUCNoR	TSEWNoR
BaseScenario	1.25 (0.18)	1.26 (0.19)	1.26 (0.19)	1.27 (0.19)
HarmTrialTrt	1.25 (0.18)	1.26 (0.18)	1.25 (0.18)	1.26 (0.18)
NoEffectTrialTrt	1.25 (0.18)	1.26 (0.18)	1.26 (0.18)	1.27 (0.18)
VeryBenTrialTrt	1.25 (0.19)	1.26 (0.20)	1.26 (0.20)	1.27 (0.20)
ThreeMonthlyFup	1.26 (0.18)	1.27 (0.18)	1.26 (0.18)	1.26 (0.18)
MajorityExpTrtB	1.25 (0.23)	1.25 (0.22)	1.25 (0.23)	1.26 (0.22)
MajorityDiscontExp	1.26 (0.22)	1.27 (0.22)	1.26 (0.22)	1.27 (0.22)
EarlyDiscontinuation	1.23 (0.16)	1.24 (0.16)	1.24 (0.16)	1.25 (0.16)
LateDiscontinuation	1.31 (0.41)	1.31 (0.39)	1.31 (0.40)	1.32 (0.39)
MajorityDiscontA	1.25 (0.19)	1.25 (0.19)	1.25 (0.19)	1.26 (0.19)
VeryBenSubTrt	1.66 (0.27)	1.67 (0.28)	1.67 (0.28)	1.68 (0.28)

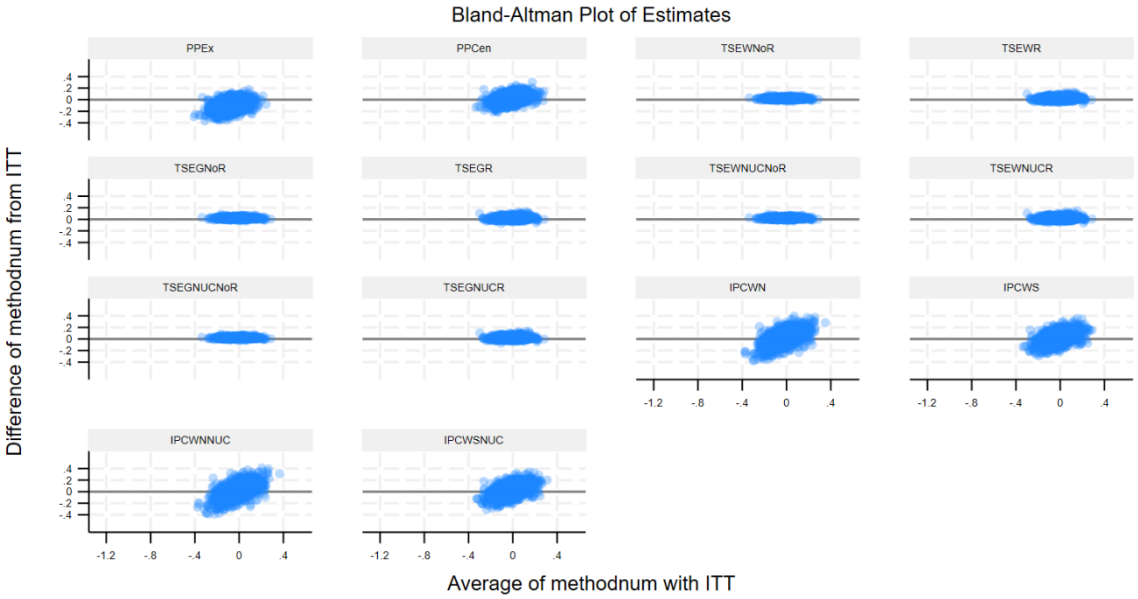
Appendix Table D.8: Percentage of Simulations Randomisation Allocation Violates Proportional Hazards Within Each Implemented Method

	ITT	PPCen	PPEx	TSEG NUCNoR	TSEG NUCR	TSEG NoR	TSEG R	TSEW NUCNoR	TSEW NUCR	TSE WNoR	TSE WR	IPCW N	IPCW NNUC	IPCW S	IPCW SNUC
BaseScenario	5%	6%	7%	4%	4%	5%	5%	5%	5%	5%	5%	33%	33%	27%	27%
HarmTrialTrt	5%	6%	14%	5%	6%	5%	5%	5%	5%	5%	5%	37%	37%	28%	29%
NoEffectTrialTrt	5%	6%	11%	4%	5%	4%	5%	5%	5%	5%	5%	34%	35%	27%	27%
VeryBenTrialTrt	6%	6%	7%	4%	5%	5%	5%	5%	5%	5%	5%	32%	33%	26%	26%
ThreeMonthlyFup	6%	5%	8%	5%	5%	5%	5%	5%	5%	5%	6%	35%	35%	40%	39%
MajorityExpTrtB	8%	4%	11%	5%	5%	6%	6%	6%	6%	7%	6%	86%	84%	83%	83%
MajorityDiscontExp	6%	7%	41%	5%	5%	5%	6%	5%	6%	6%	6%	61%	61%	57%	56%
EarlyDiscontinuation	7%	5%	7%	5%	5%	6%	5%	6%	5%	6%	5%	28%	30%	26%	26%
LateDiscontinuation	6%	5%	6%	3%	3%	4%	4%	6%	5%	6%	5%	17%	18%	15%	16%
MajorityDiscontA	5%	6%	6%	5%	5%	5%	5%	5%	5%	5%	5%	10%	9%	9%	9%
VeryBenSubTrt	6%	6%	8%	4%	4%	5%	4%	5%	5%	5%	5%	33%	34%	26%	27%

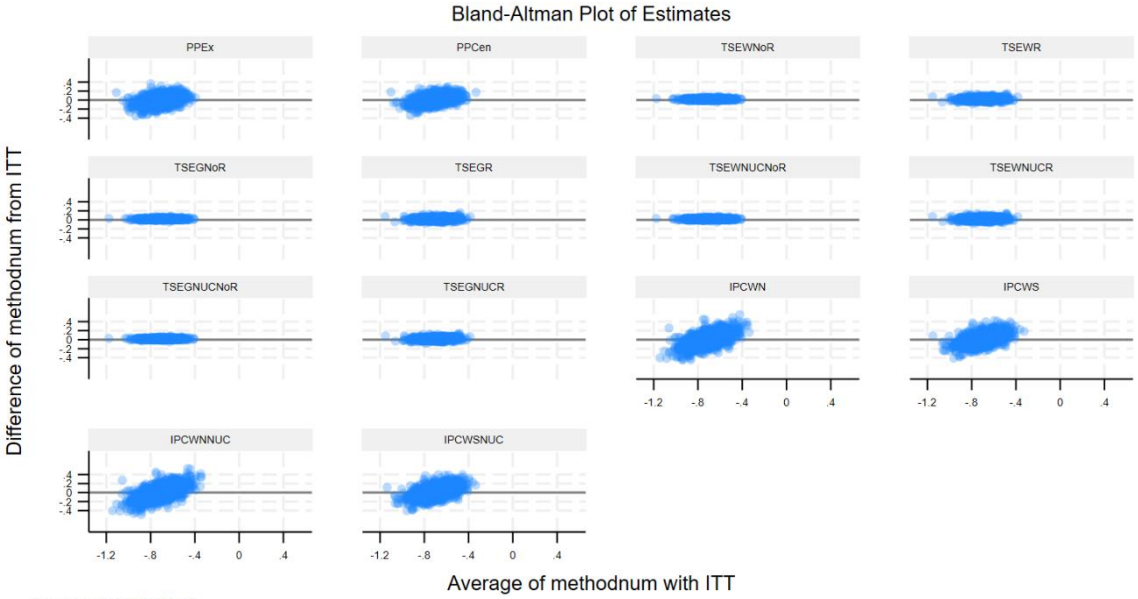
Appendix Figure D.19: Bland-Altman Plots of Estimates Across the Simulations - With the constant treatment effect assumption



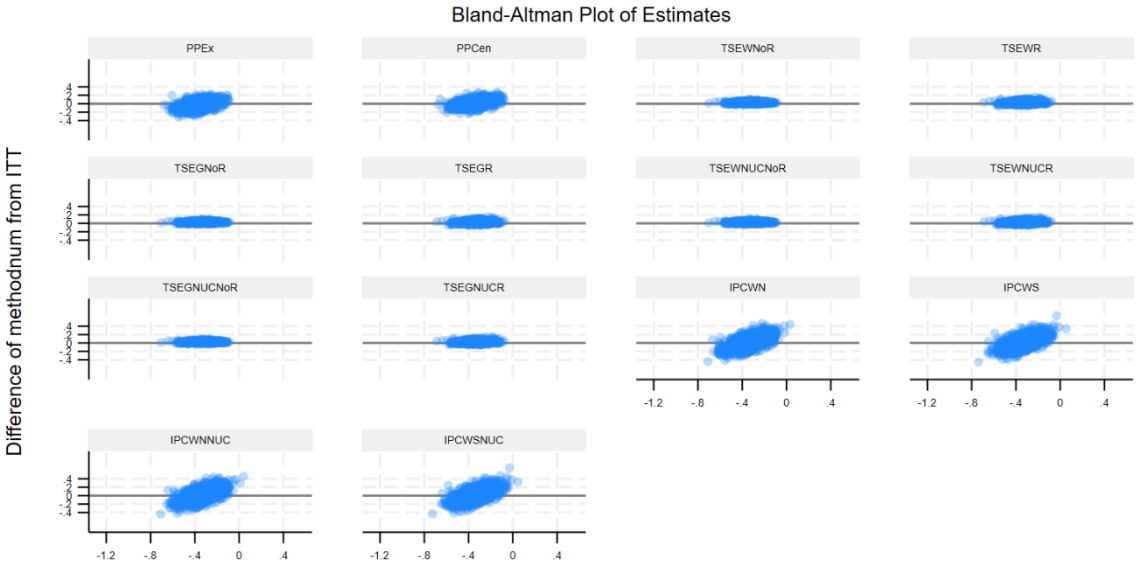




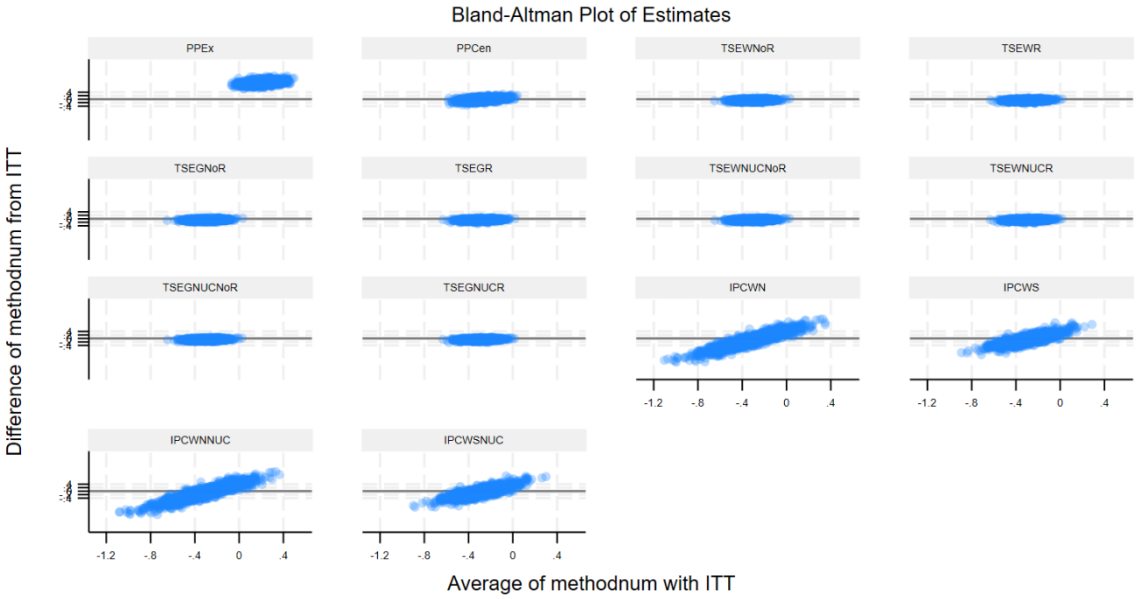
Graphs by Method
Estimates (coeff) for dgmnum=NoEffectTrialTrt
Data generated under the constant treatment effect assumption



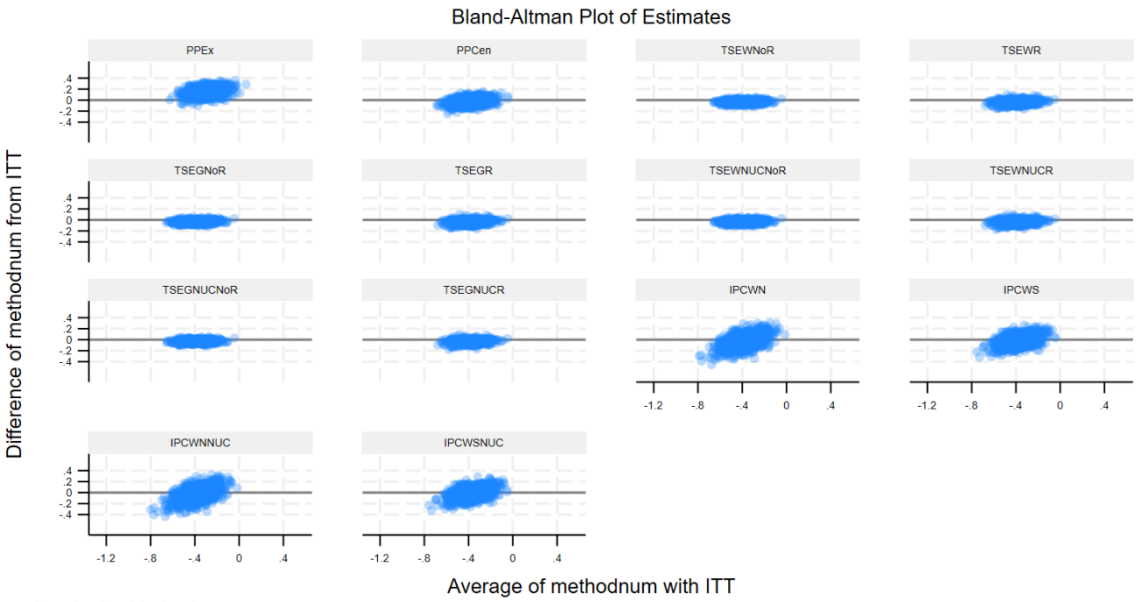
Graphs by Method
Estimates (coeff) for dgmnum=VeryBenTrialTrt
Data generated under the constant treatment effect assumption



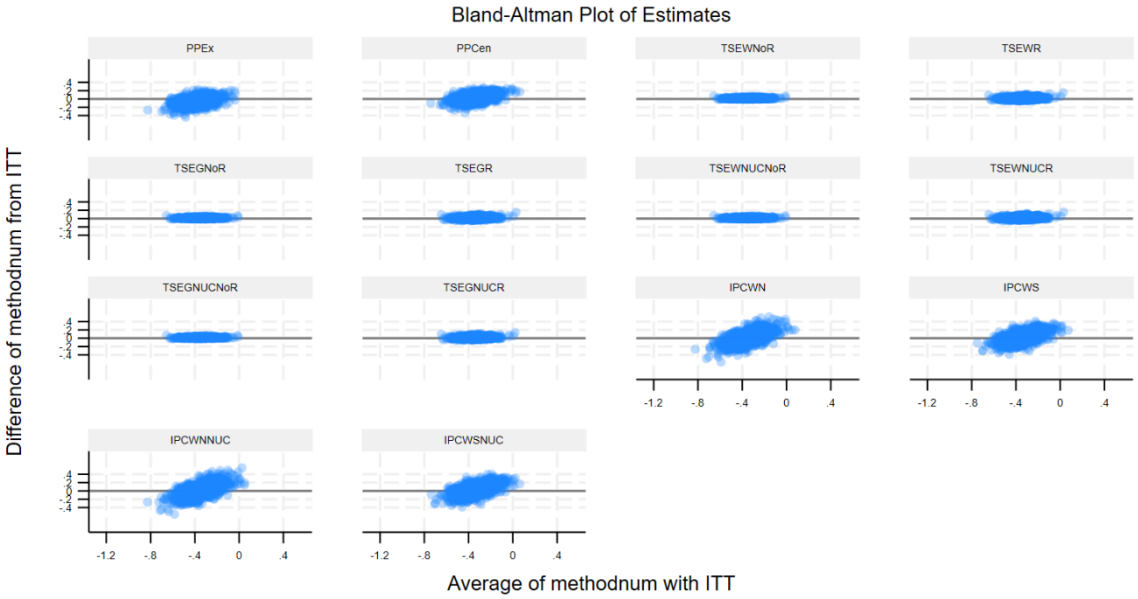
Graphs by Method
Estimates (coeff) for dgmnum=ThreeMonthlyFup
Data generated under the constant treatment effect assumption



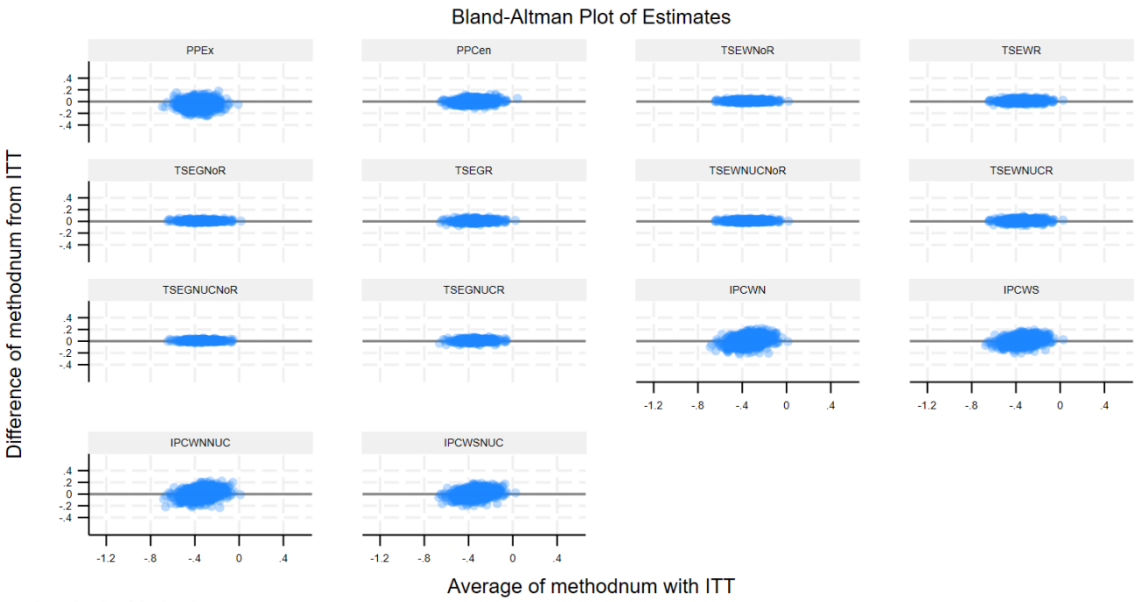
Graphs by Method
Estimates (coeff) for dgmnum=MajorityExpTrtB
Data generated under the constant treatment effect assumption



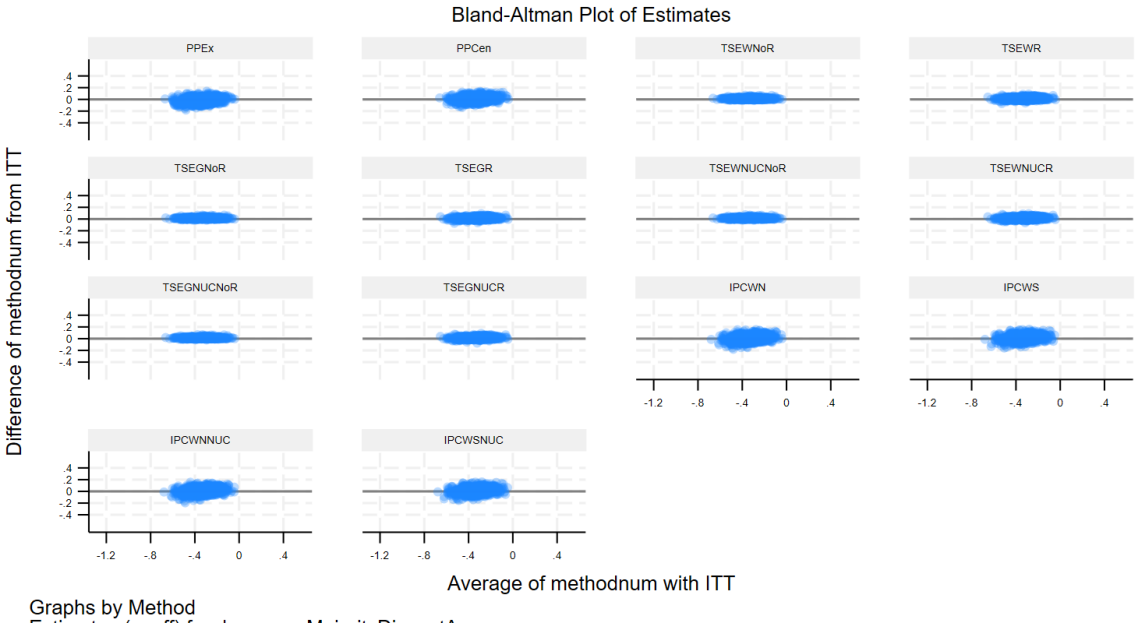
Graphs by Method
Estimates (coeff) for dgmnum=MajorityDiscontExp
Data generated under the constant treatment effect assumption



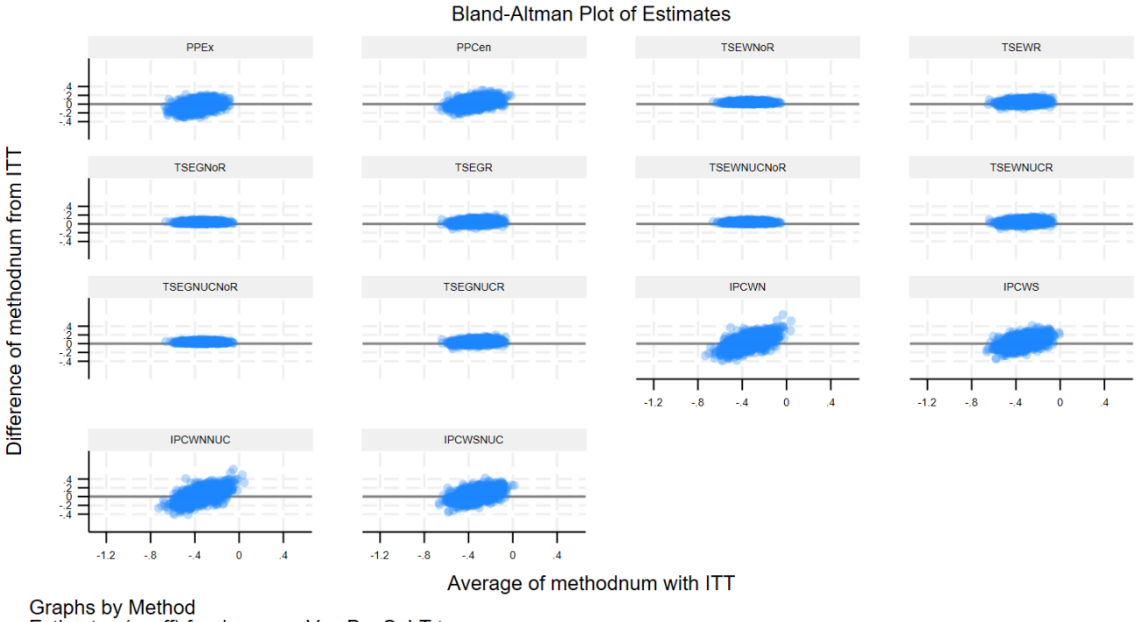
Graphs by Method
Estimates (coeff) for dgmnum=EarlyDiscontinuation
Data generated under the constant treatment effect assumption



Graphs by Method
Estimates (coeff) for dgmnum=LateDiscontinuation
Data generated under the constant treatment effect assumption

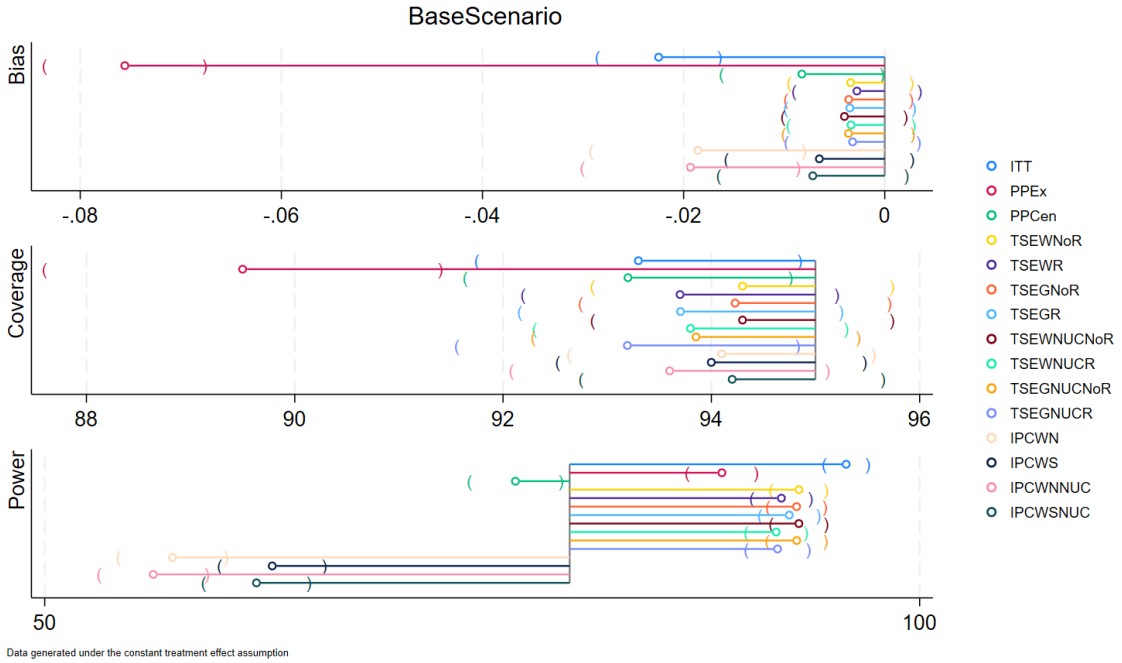


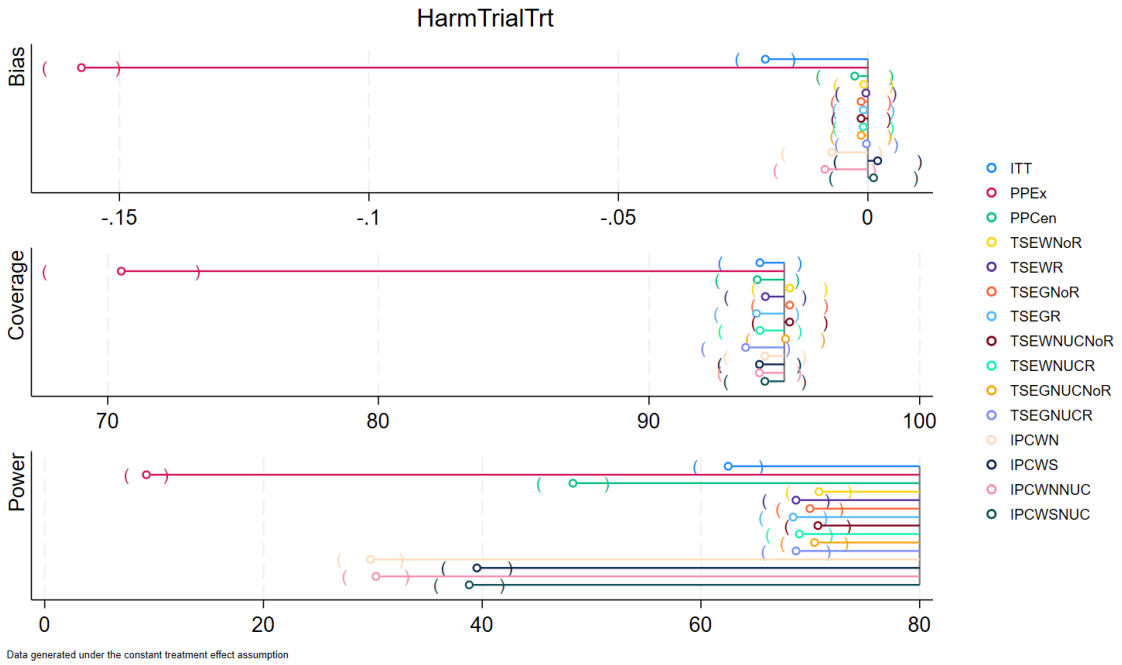
Graphs by Method
Estimates (coeff) for dgmnum=MajorityDiscontA
Data generated under the constant treatment effect assumption

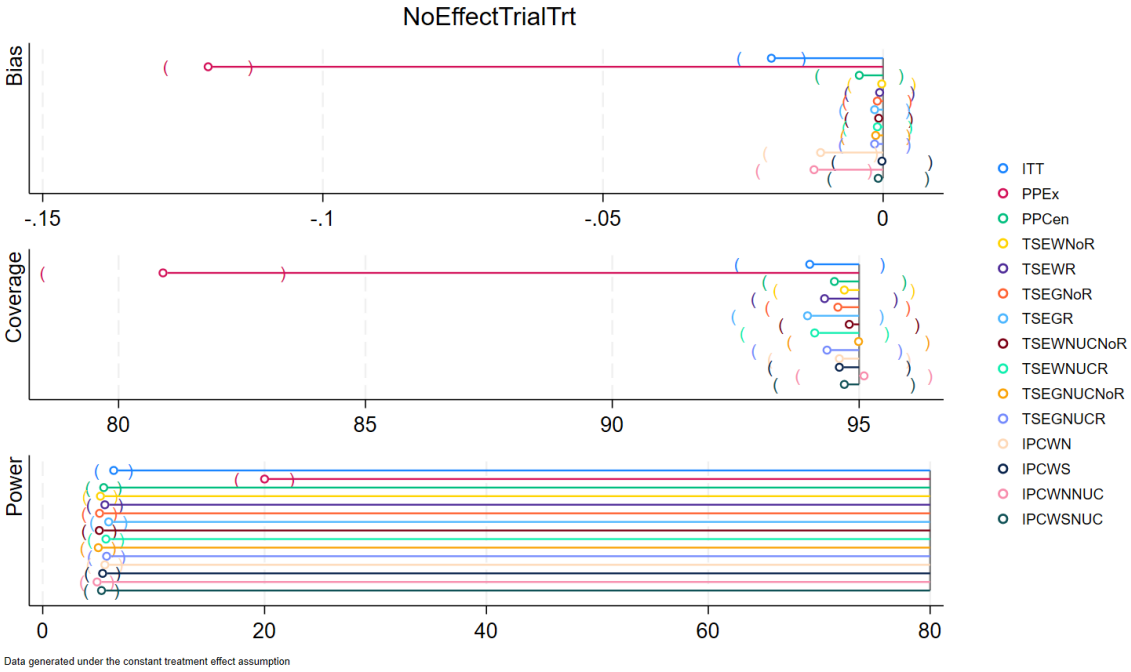


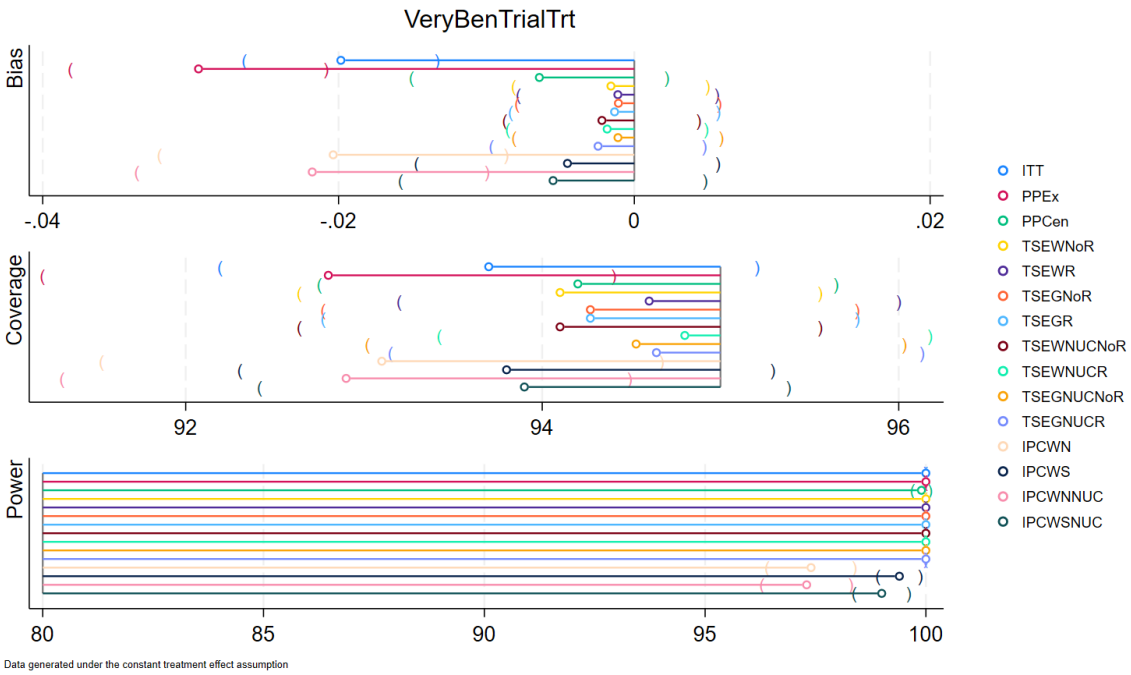
Graphs by Method
Estimates (coeff) for dgmnum=VeryBenSubTrt
Data generated under the constant treatment effect assumption

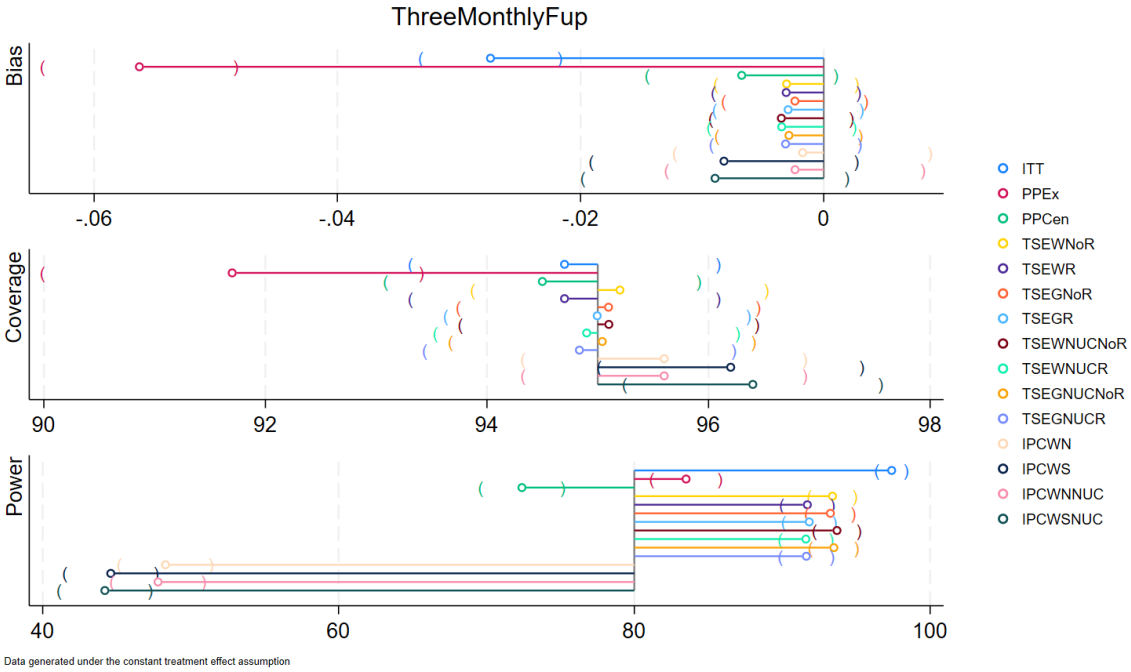
Appendix Figure D.20: Lollipop Plots of Performance Measures from Simulations - With the constant treatment effect assumption

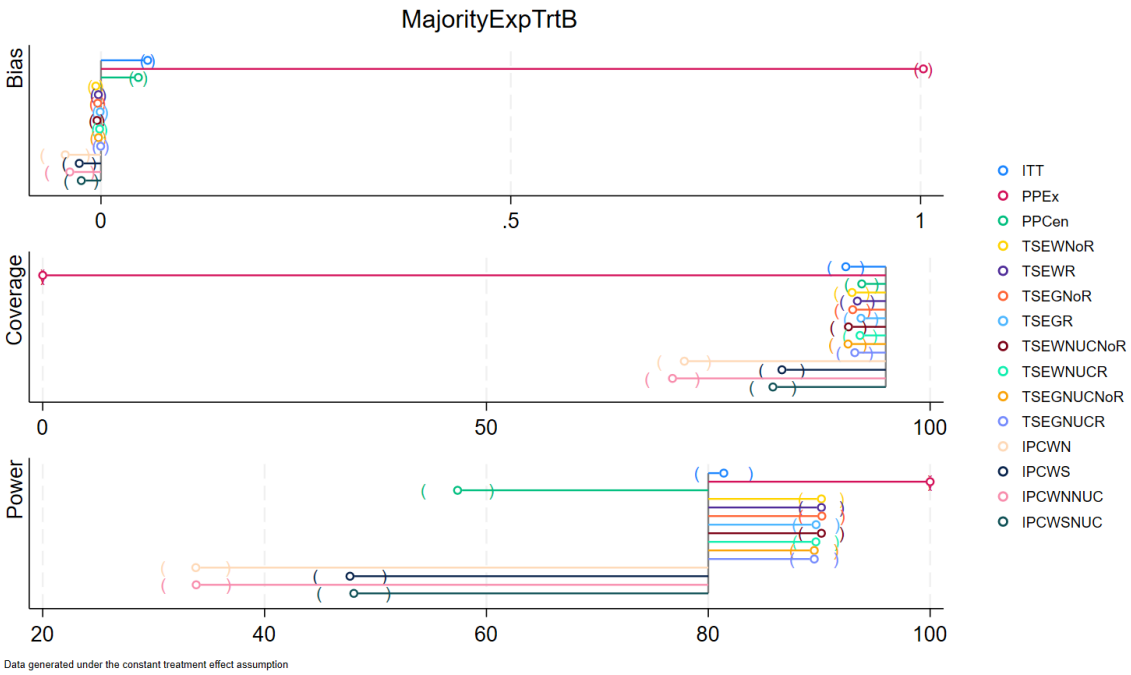


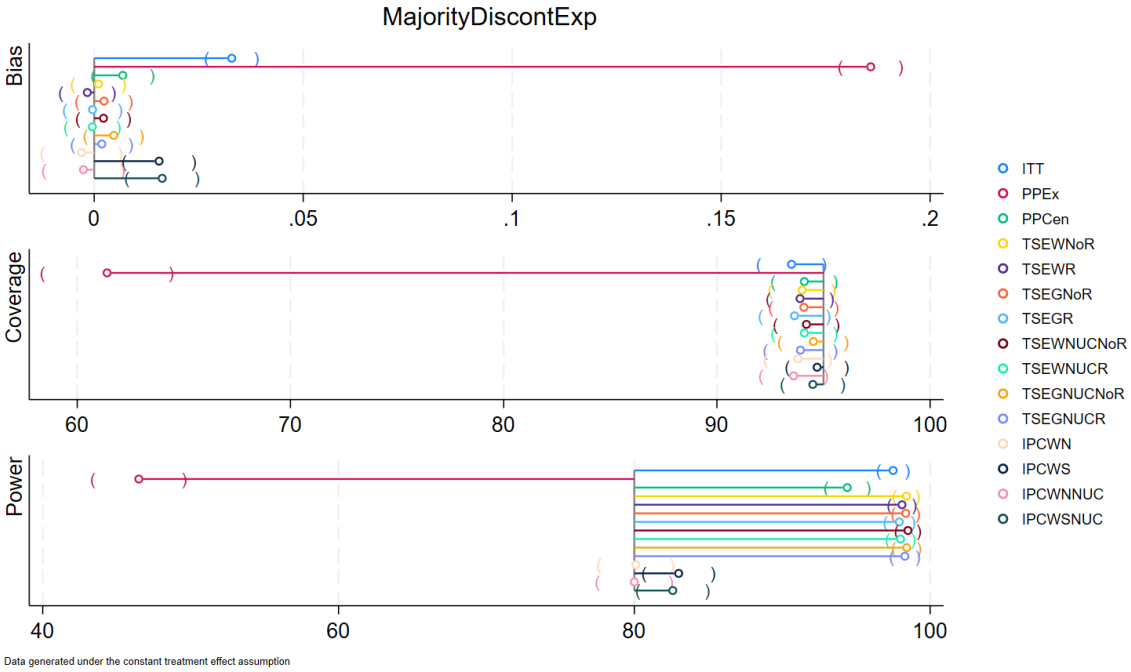


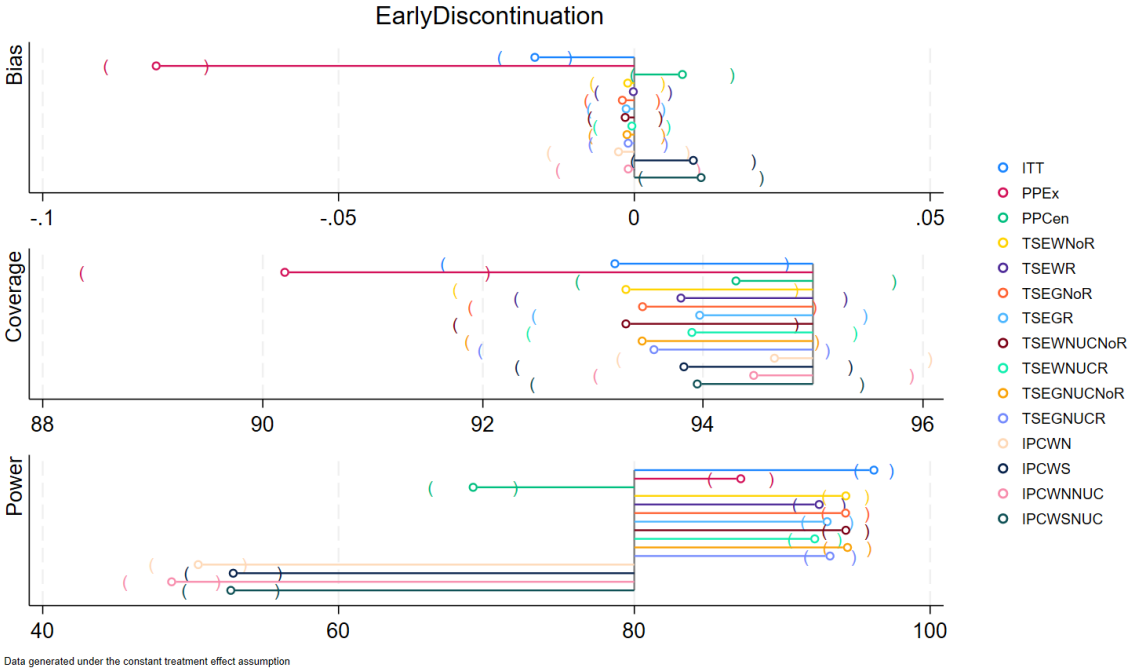

NoEffectTrialTrt
Bias
-.15
-.1
-.05
0
Coverage
80
85
90
95
Power
0
20
40
60
80
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

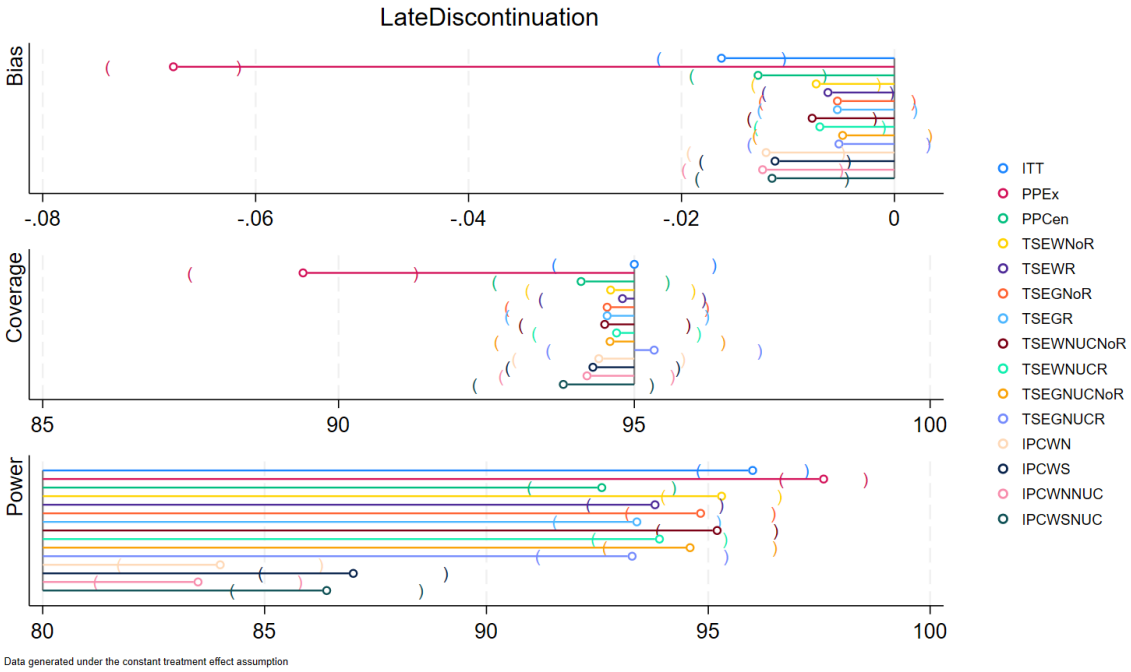

VeryBenTrialTrt
Bias
-.04
-.02
0
.02
Coverage
92
94
96
Power
80
85
90
95
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

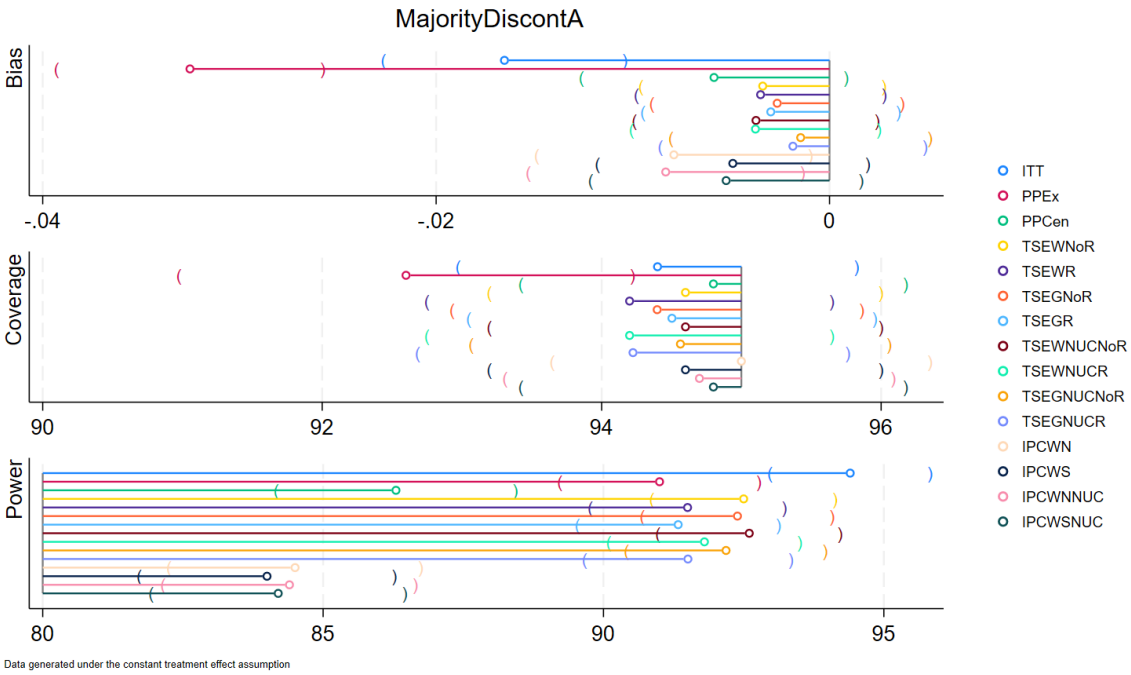

ThreeMonthlyFup
Bias
-.06
-.04
-.02
0
Coverage
90
92
94
96
98
Power
40
60
80
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

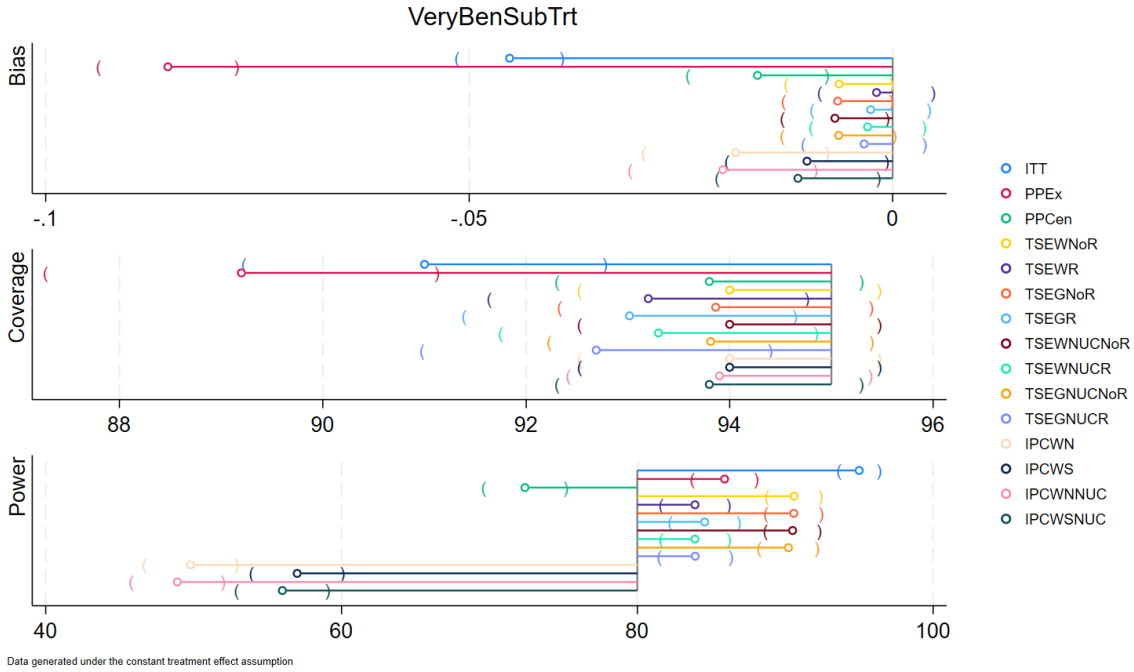

MajorityExpTrtB
Bias
0
.5
1
Coverage
0
50
100
Power
20
40
60
80
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC


MajorityDiscontExp
Bias
0
.05
.1
.15
.2
Coverage
60
70
80
90
100
Power
40
60
80
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

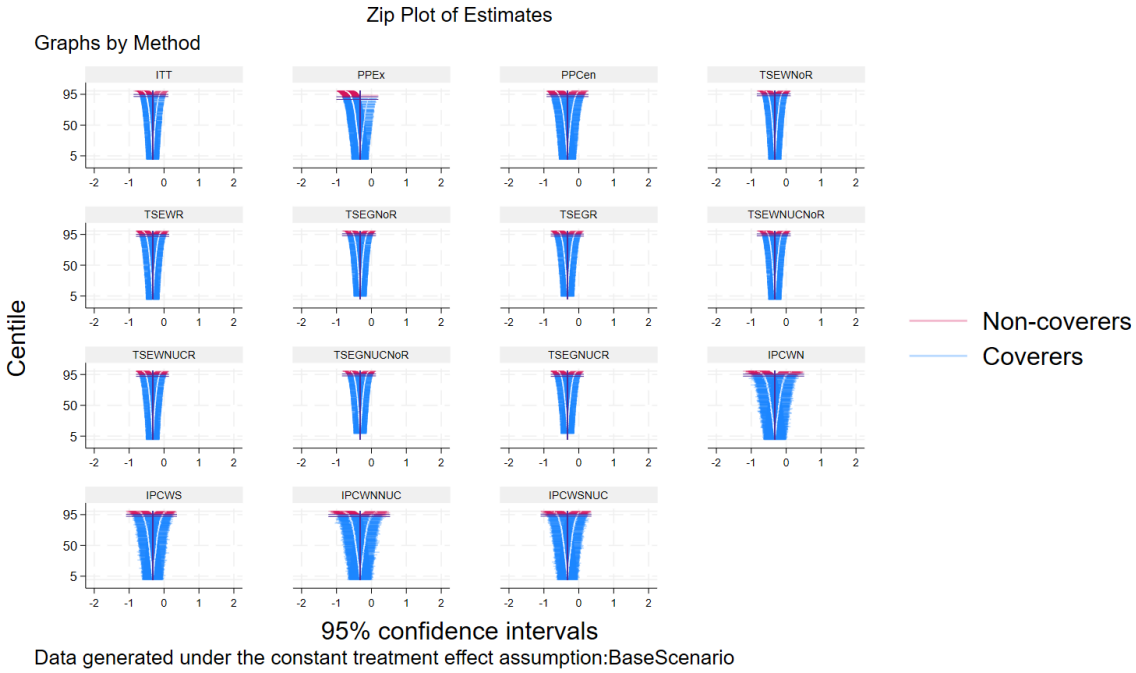

EarlyDiscontinuation
Bias
-.1
-.05
0
.05
Coverage
88
90
92
94
96
Power
40
60
80
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC


LateDiscontinuation
Bias
-.08
-.06
-.04
-.02
0
Coverage
85
90
95
100
Power
80
85
90
95
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

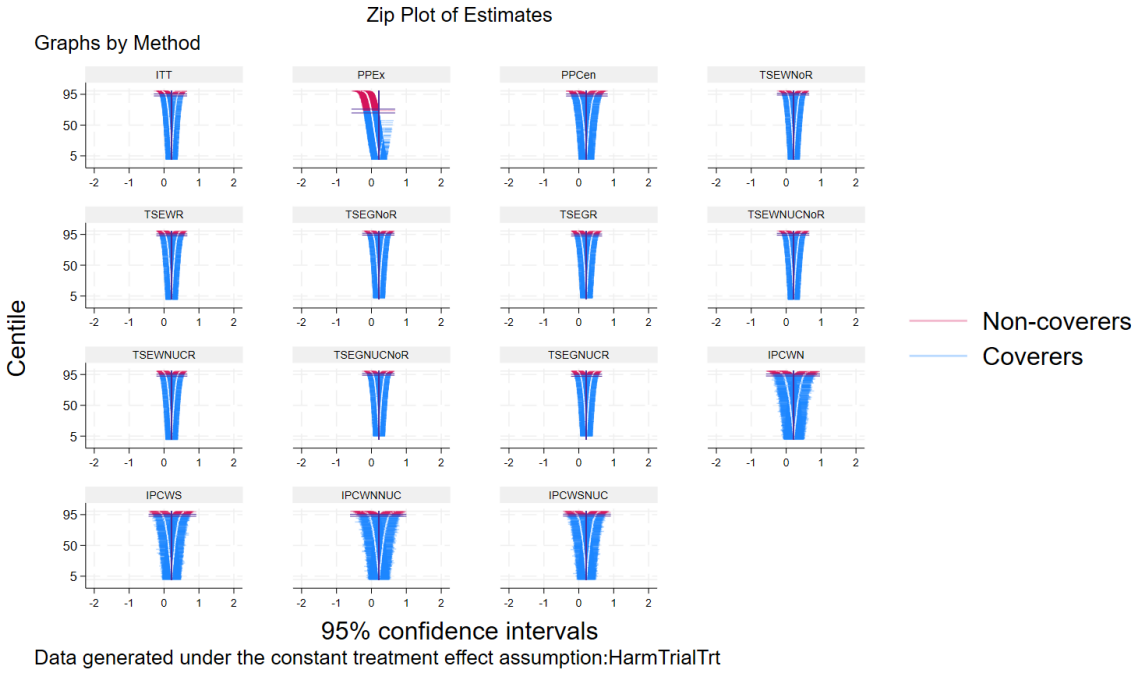

MajorityDiscontA
Bias
-.04
-.02
0
Coverage
90
92
94
96
Power
80
85
90
95
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC


VeryBenSubTrt
Bias
-.1
-.05
0
Coverage
88
90
92
94
96
Power
40
60
80
100
Data generated under the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

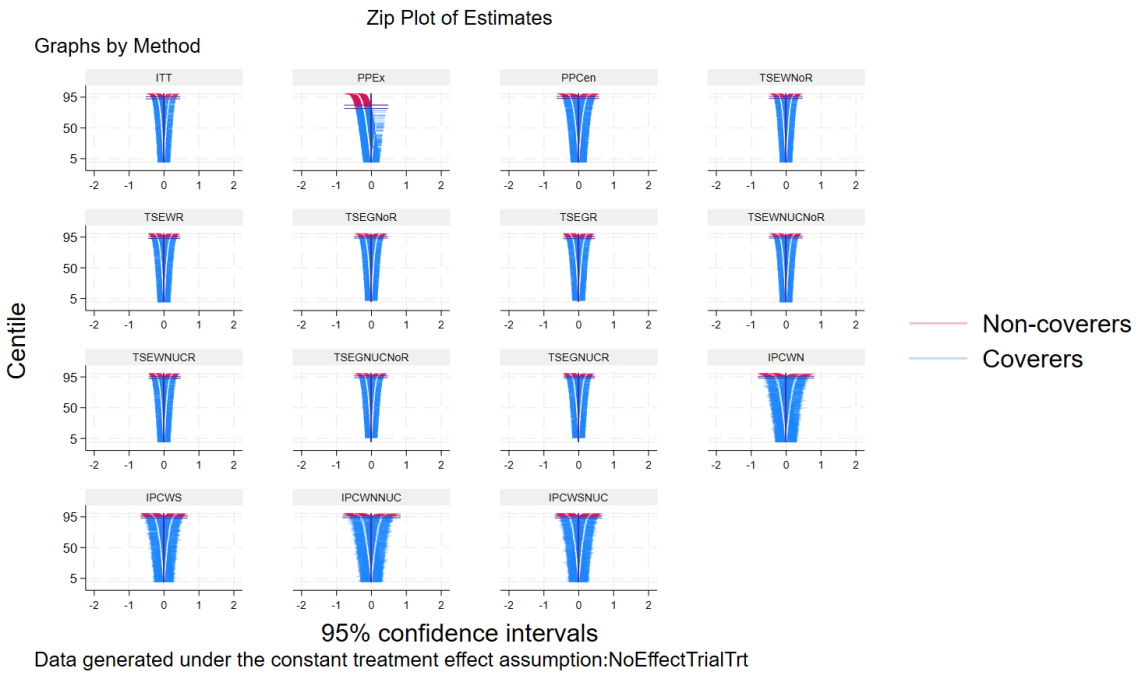
Appendix Figure D.21: Zip Plots Estimates Across the Simulations - With the constant treatment effect assumption



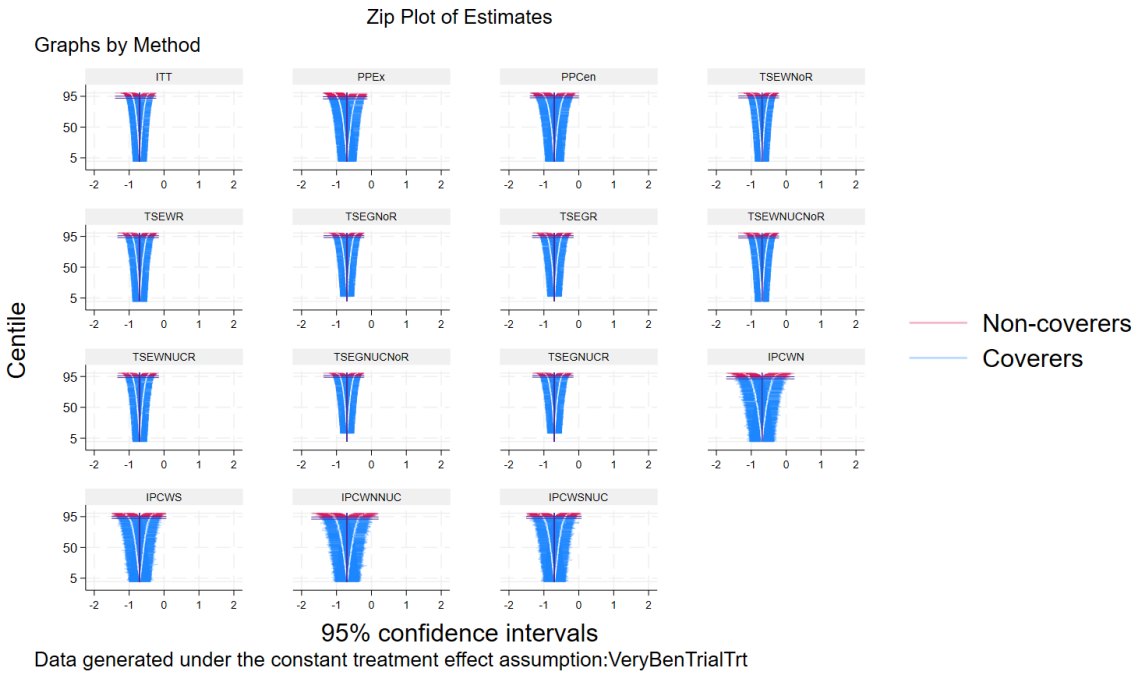
Data generated under the constant treatment effect assumption:BaseScenario



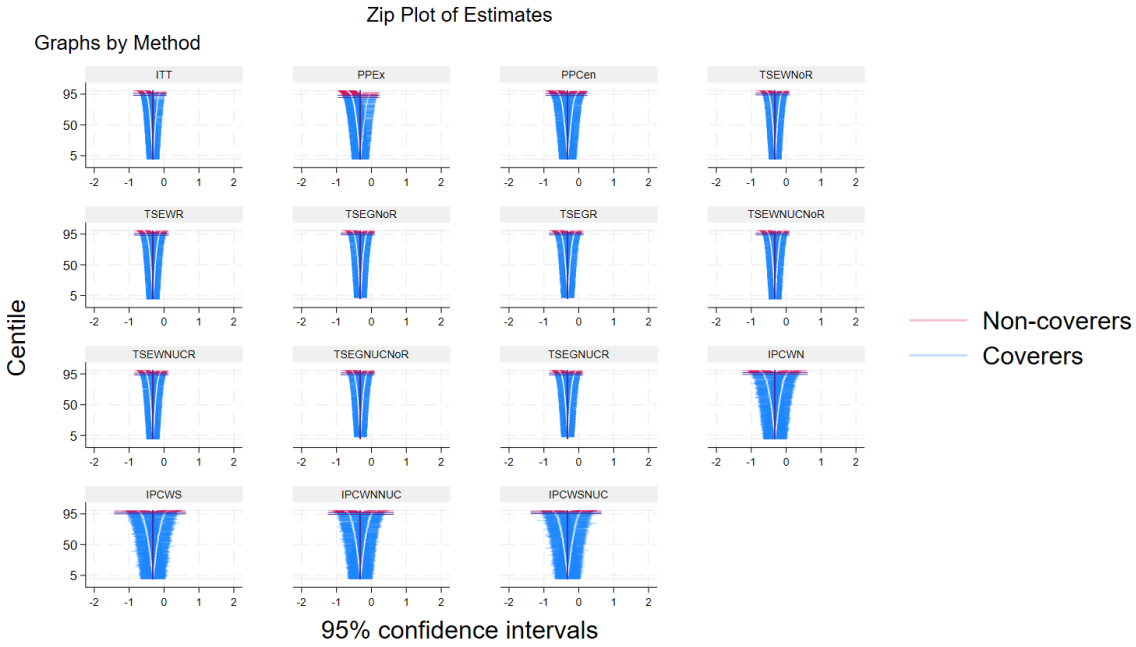
Data generated under the constant treatment effect assumption:HarmTrialTrt

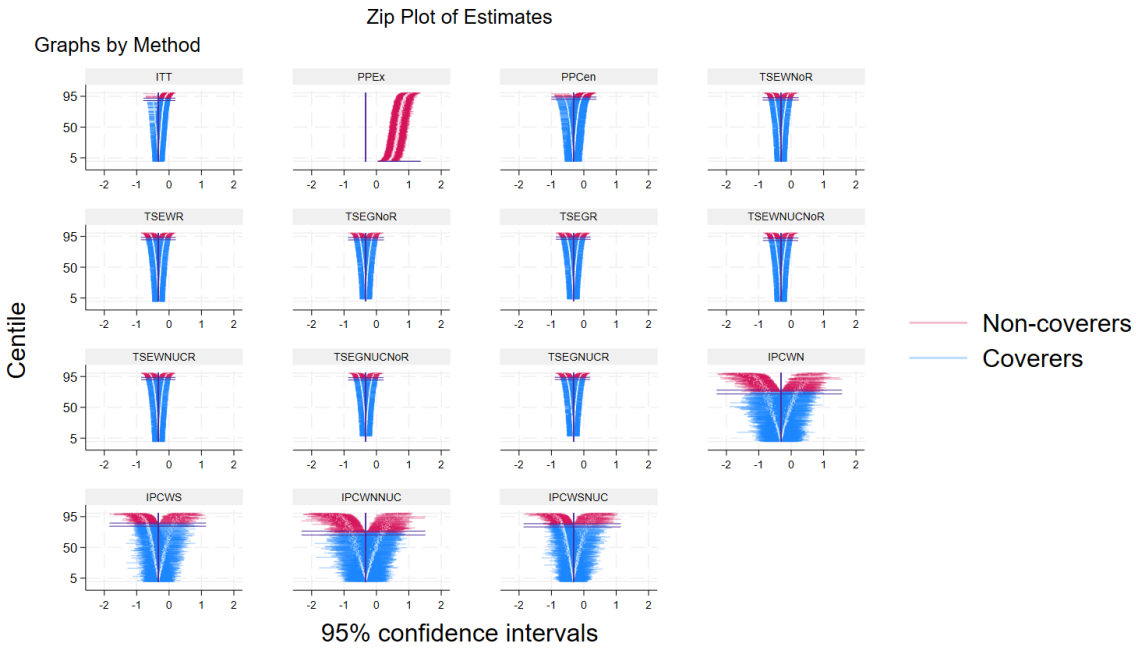


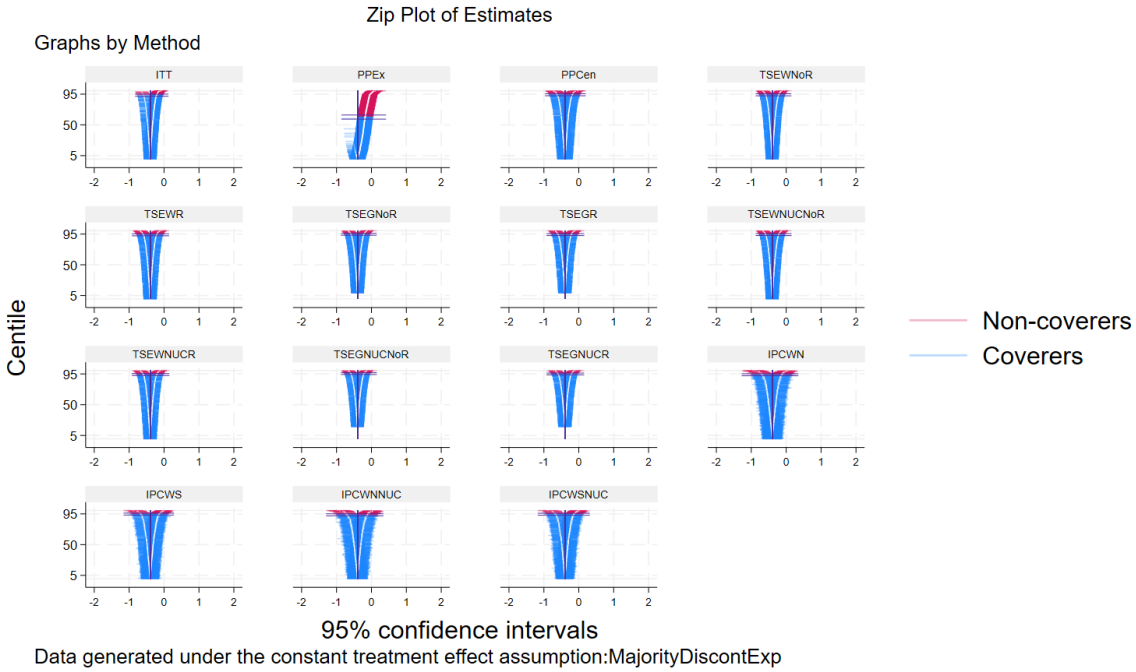
Data generated under the constant treatment effect assumption:NoEffectTrialTrt



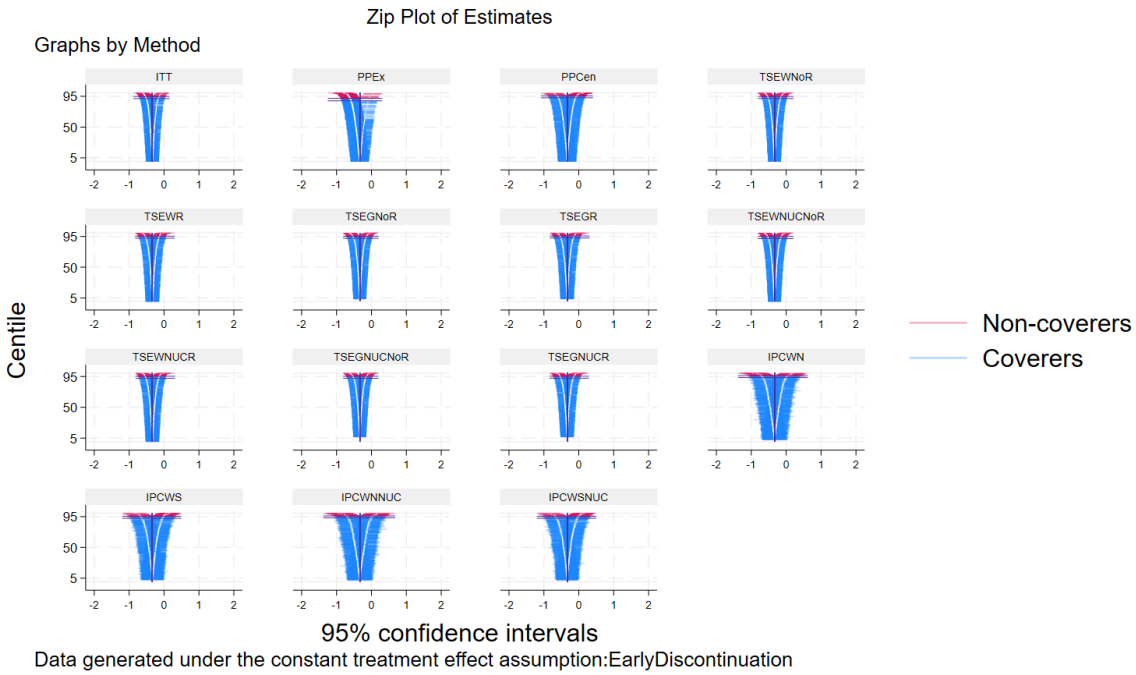
Data generated under the constant treatment effect assumption:VeryBenTrialTrt


Zip Plot of Estimates
Graphs by Method
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC
95
50
5
-2
-1
0
1
2
Centile
Non-coverers
Coverers
95% confidence intervals
Data generated under the constant treatment effect assumption:ThreeMonthlyFup

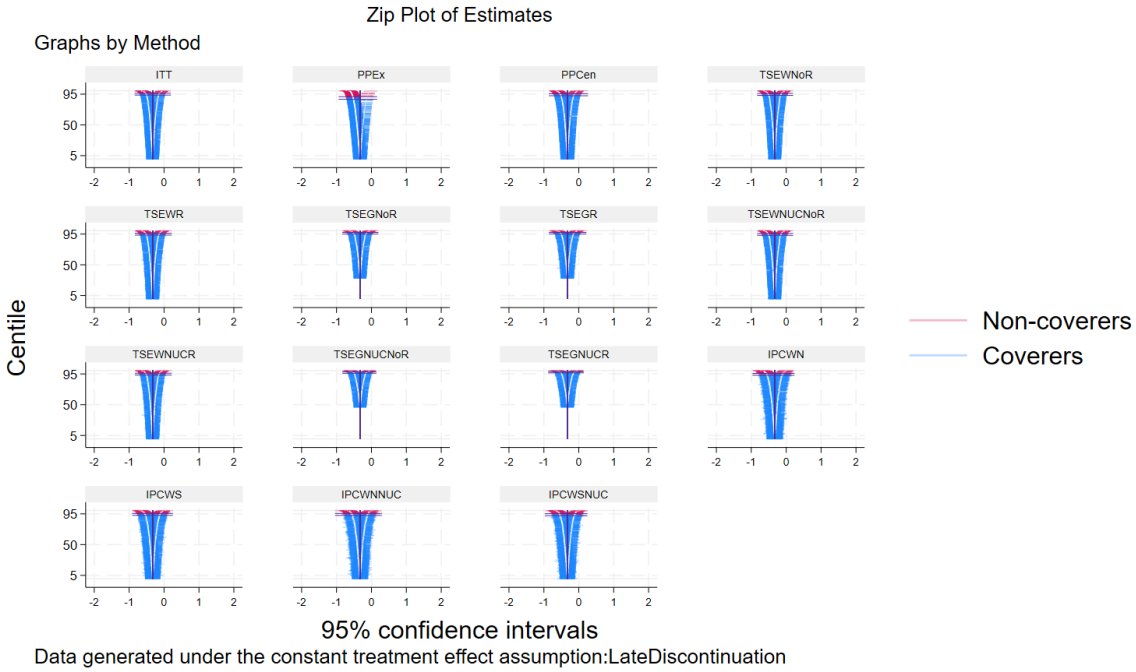

Zip Plot of Estimates
Graphs by Method
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC
95
50
5
-2
-1
0
1
2
Centile
Non-coverers
Coverers
95% confidence intervals
Data generated under the constant treatment effect assumption:MajorityExpTrtB



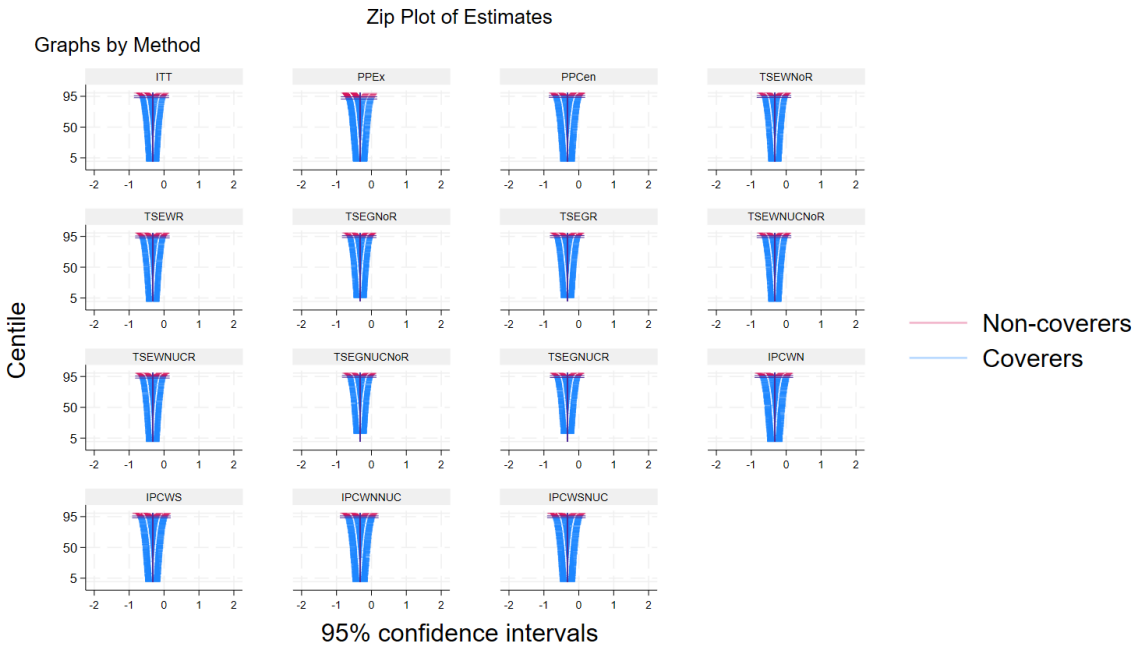
Data generated under the constant treatment effect assumption:MajorityDiscontExp



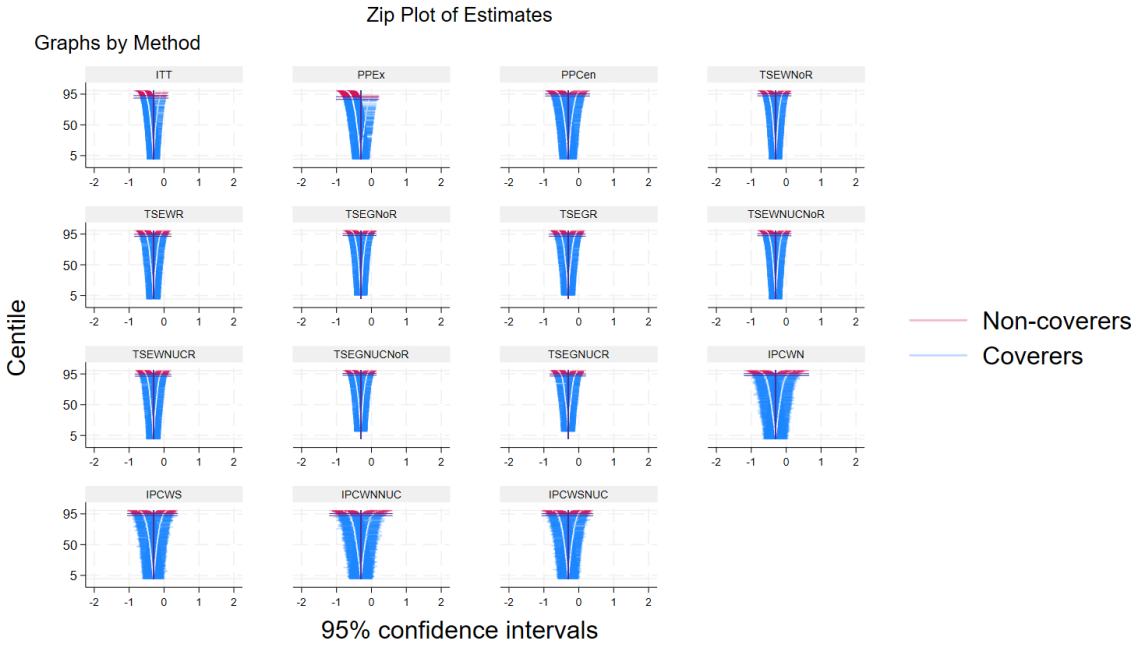
Data generated under the constant treatment effect assumption:EarlyDiscontinuation



Data generated under the constant treatment effect assumption:LateDiscontinuation



Data generated under the constant treatment effect assumption:MajorityDiscontA



Data generated under the constant treatment effect assumption:VeryBenSubTrt

D.4 Simulation Results – Relaxing the Constant Treatment Effect Assumption

This section provides the results of the simulation study where the constant treatment effect assumption was relaxed. For each of the scenarios, one thousand repetitions were generated, and the results outputted. Note that NoEffectTrialTrt was not repeated as the assumption had no impact on this scenario. In the scenarios which vary the impact of the treatment effect and/or methods where more censoring or exclusion of participants occurred, less events are observed or used in the analysis (Appendix Table D.9). This was the similar to when the constant treatment effect assumption was enforced.

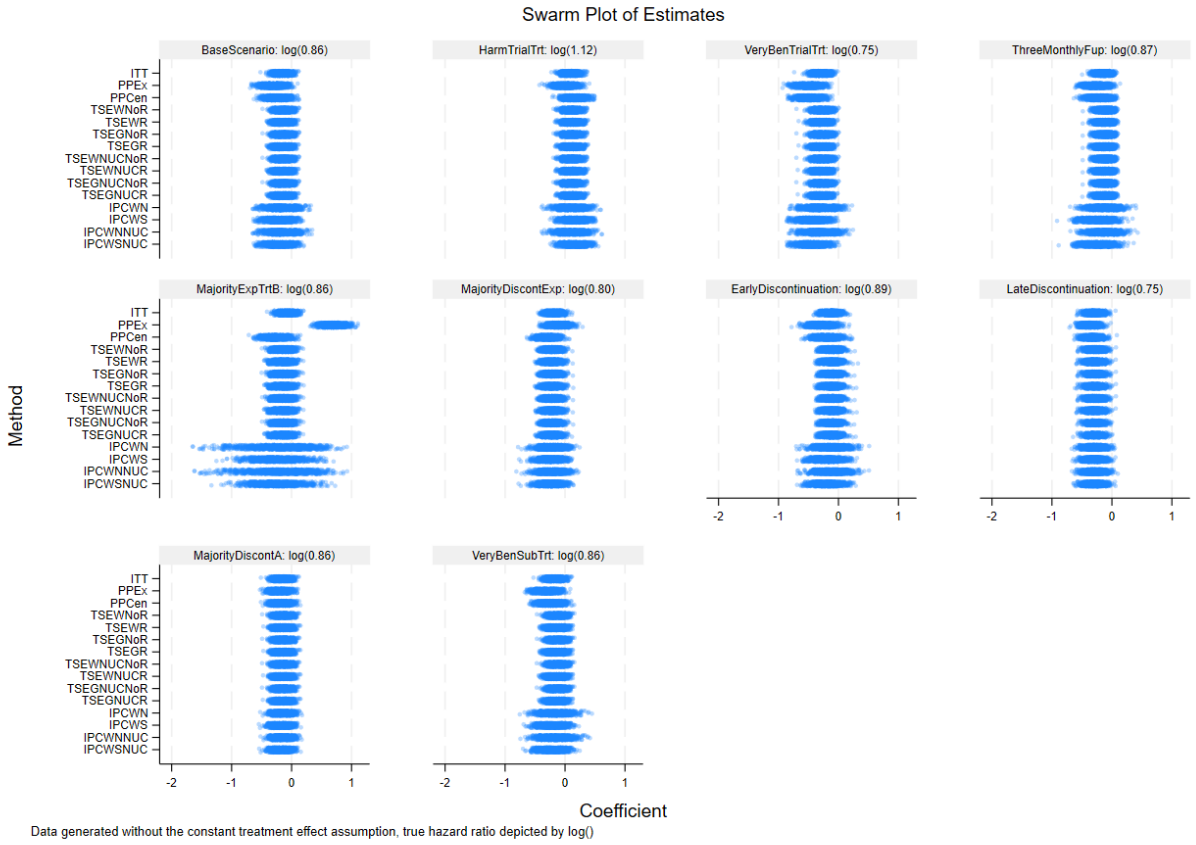
Appendix Table D.9: Mean (SD) of the Number of Events used in the Cox Model Averaged Across the Repetitions - Without the constant treatment effect assumption

	ITT	PPEx	PPCen	TSE WNoR	TSE WR	TSE GNoR	TSE GR	TSEW NUCNoR	TSE WNUCR	TSEG NUCNoR	TSE GNUCR	IPCWN	IPCWS	IPCWN NUC	IPCWS NUC
BaseScenario	490 (12)	280 (13)	280 (13)	490 (12)	461 (21)	490 (12)	462 (20)	490 (12)	462 (20)	490 (12)	463 (20)	463 (25)	258 (16)	464 (26)	258 (16)
HarmTrialTrt	514 (11)	321 (13)	321 (13)	514 (11)	487 (20)	514 (11)	488 (19)	514 (11)	488 (19)	514 (11)	489 (19)	489 (25)	302 (17)	490 (26)	302 (17)
VeryBenTrialTrt	477 (12)	259 (13)	259 (13)	477 (12)	447 (21)	477 (12)	448 (20)	477 (12)	448 (21)	477 (12)	449 (20)	449 (26)	237 (16)	450 (26)	237 (16)
ThreeMonthlyFup	510 (12)	274 (13)	274 (13)	510 (12)	481 (21)	510 (12)	482 (21)	510 (12)	482 (21)	510 (12)	482 (21)	483 (25)	143 (12)	484 (26)	143 (12)
MajorityExpTrtB	491 (12)	280 (13)	280 (13)	491 (12)	461 (22)	491 (12)	462 (22)	491 (12)	461 (22)	491 (12)	462 (22)	464 (117)	258 (47)	466 (124)	259 (50)
MajorityDiscontExp	493 (12)	347 (13)	347 (13)	493 (12)	472 (18)	493 (12)	472 (17)	493 (12)	472 (18)	494 (12)	473 (18)	473 (22)	320 (17)	474 (23)	320 (17)
EarlyDiscontinuation	494 (12)	244 (13)	244 (13)	494 (12)	462 (22)	494 (12)	463 (21)	494 (12)	463 (22)	494 (12)	464 (21)	463 (27)	214 (16)	464 (28)	214 (16)
LateDiscontinuation	481 (12)	420 (13)	420 (13)	481 (12)	464 (21)	481 (12)	464 (23)	481 (12)	463 (22)	481 (12)	462 (25)	471 (20)	430 (17)	472 (20)	431 (17)
MajorityDiscontA	475 (12)	387 (13)	387 (13)	475 (12)	462 (14)	475 (12)	463 (14)	475 (12)	462 (14)	475 (12)	463 (14)	464 (15)	358 (14)	464 (15)	359 (14)

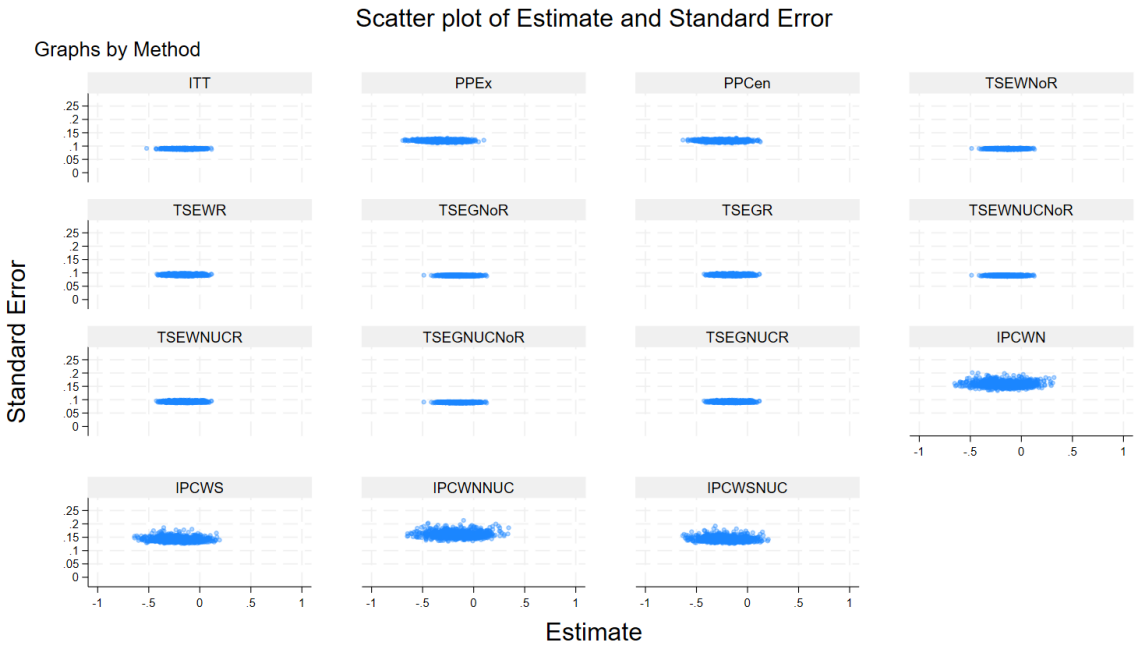
VeryBenSubTrt	476 (12)	266 (13)	266 (13)	476 (12)	415 (22)	476 (12)	416 (22)	476 (12)	415 (22)	476 (12)	417 (22)	415 (26)	239 (16)	416 (26)	240 (16)

The estimates from the 1000 repetitions without the constant treatment effect assumption are shown in Appendix Figure D.22. From this there are no obvious outliers to be concerned about. This is corroborated from the scatter plots of the estimates and their standard errors (Appendix Figure D.23). Within each scenario the estimate is similar across the different methods. However, in all scenarios the variation in the standard error in the IPCW methods is larger than all other methods. This may be due to the fact the IPCW methods rely on the derivation of weights from the data. Summaries of the weights are shown in Appendix Figure D.24. There are no obvious outliers across the repetitions within the scenarios. This is similar to when the constant treatment effect assumption was in place.

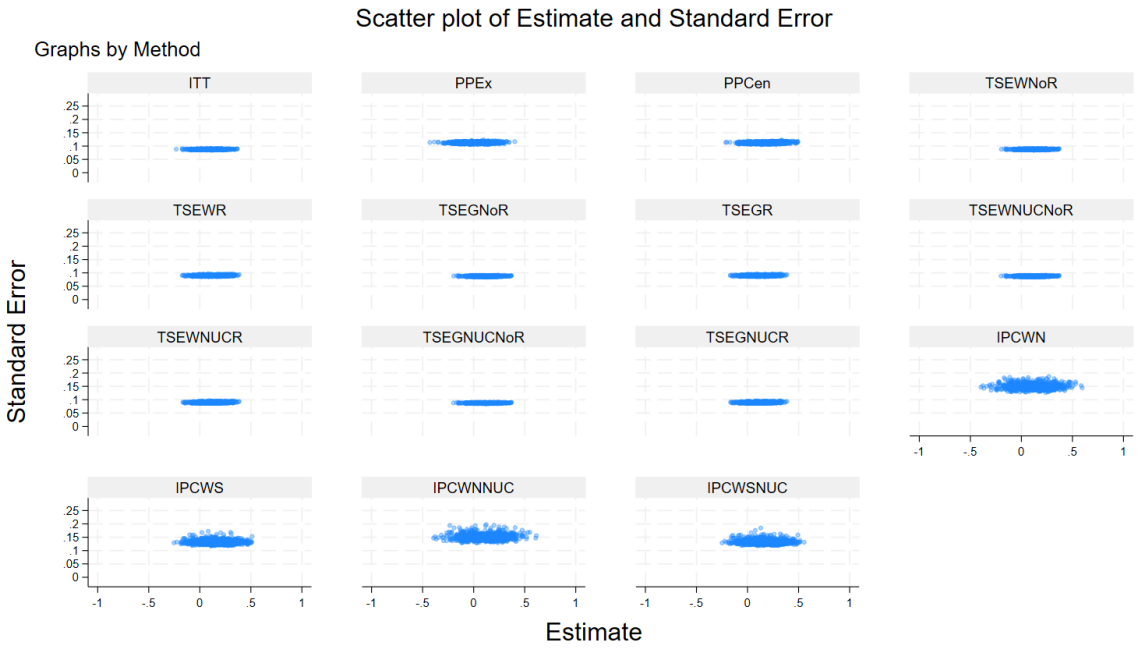
Appendix Figure D.22: Swarm Plot of Estimates Across the Scenarios - Without the constant treatment effect assumption



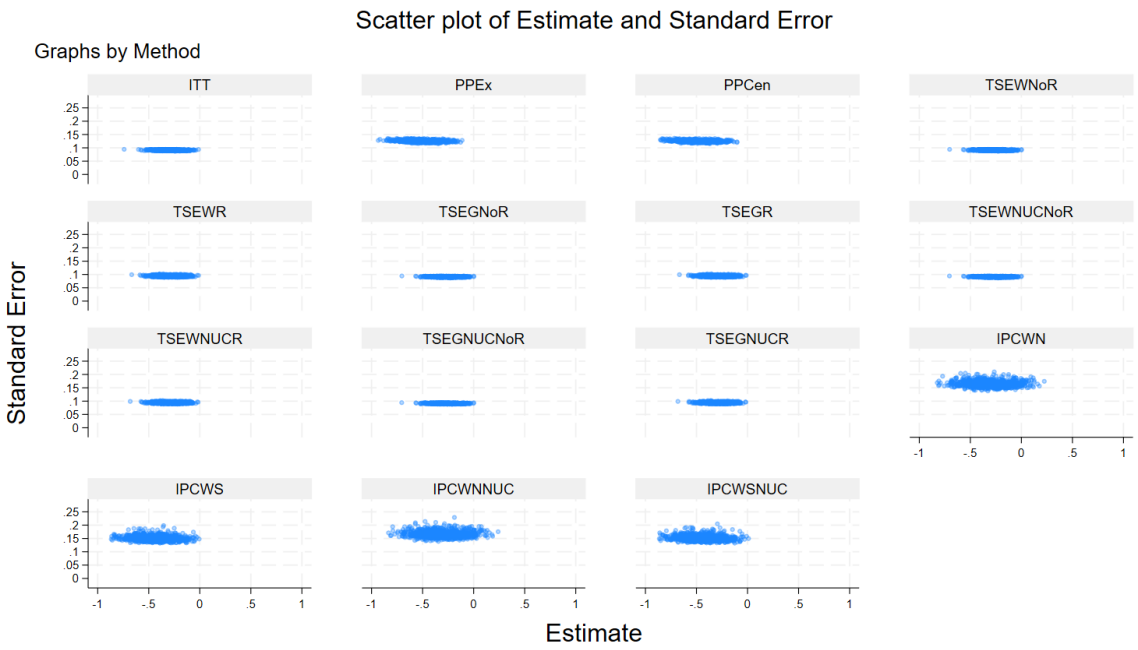
Appendix Figure D.23: Scatter Plots of the Estimates and their Standard Errors Across All Scenarios - Without the constant treatment effect assumption



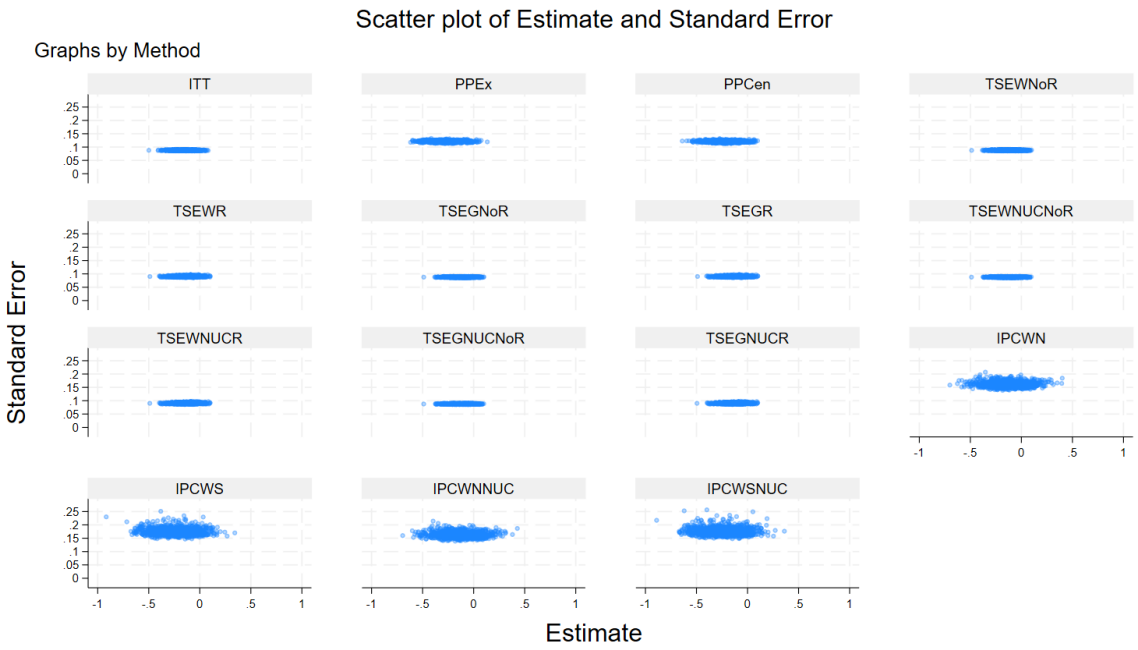
Data generated without the constant treatment effect assumption: BaseScenario



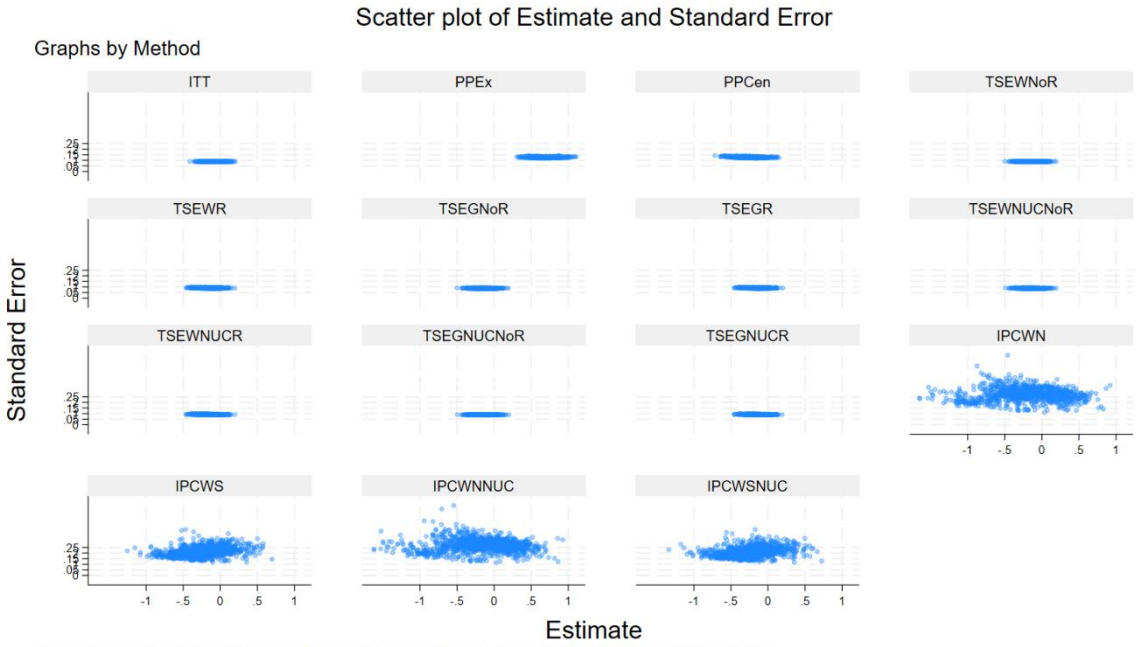
Data generated without the constant treatment effect assumption: HarmTrialTrt



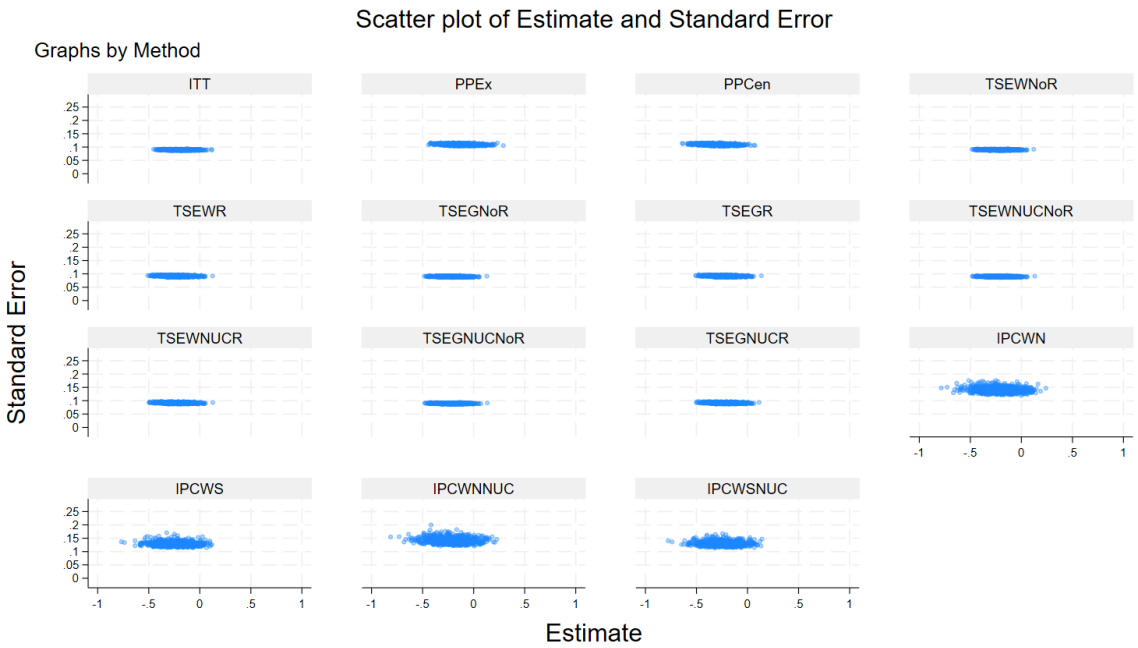
Data generated without the constant treatment effect assumption: VeryBenTrialTrt



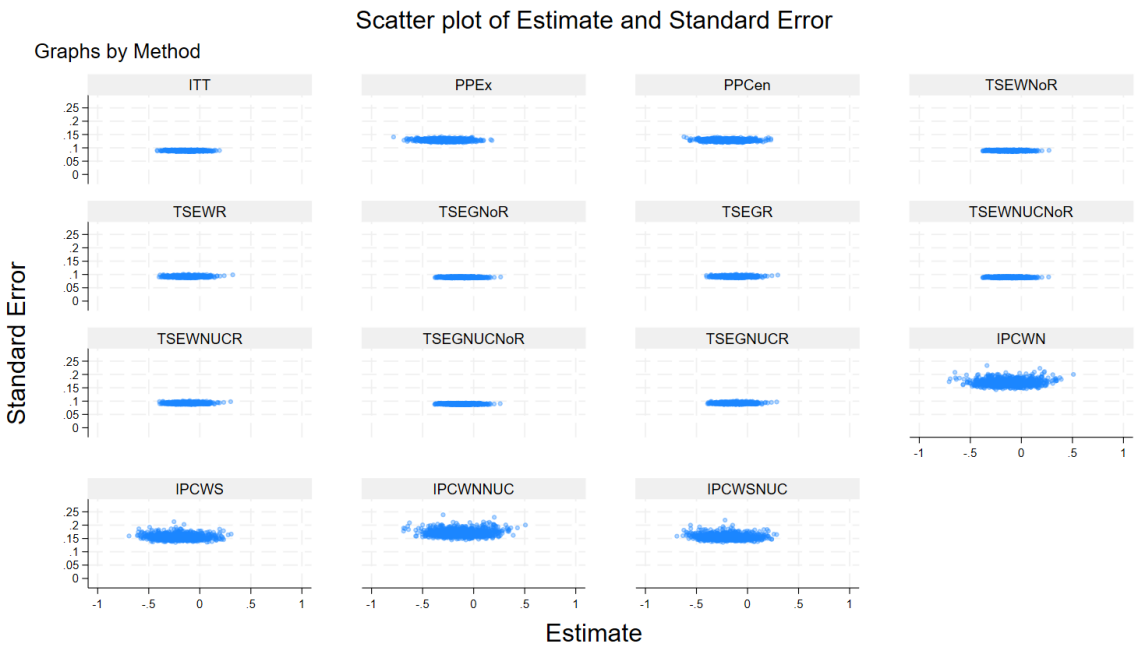
Data generated without the constant treatment effect assumption: ThreeMonthlyFup



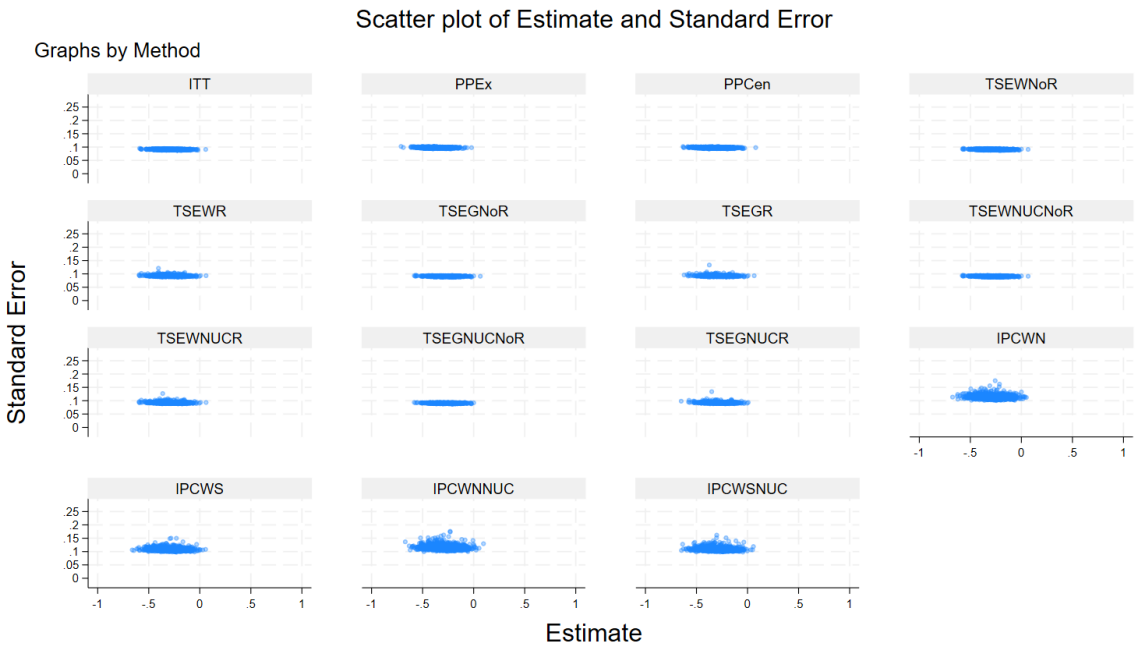
Data generated without the constant treatment effect assumption: MajorityExpTrtB



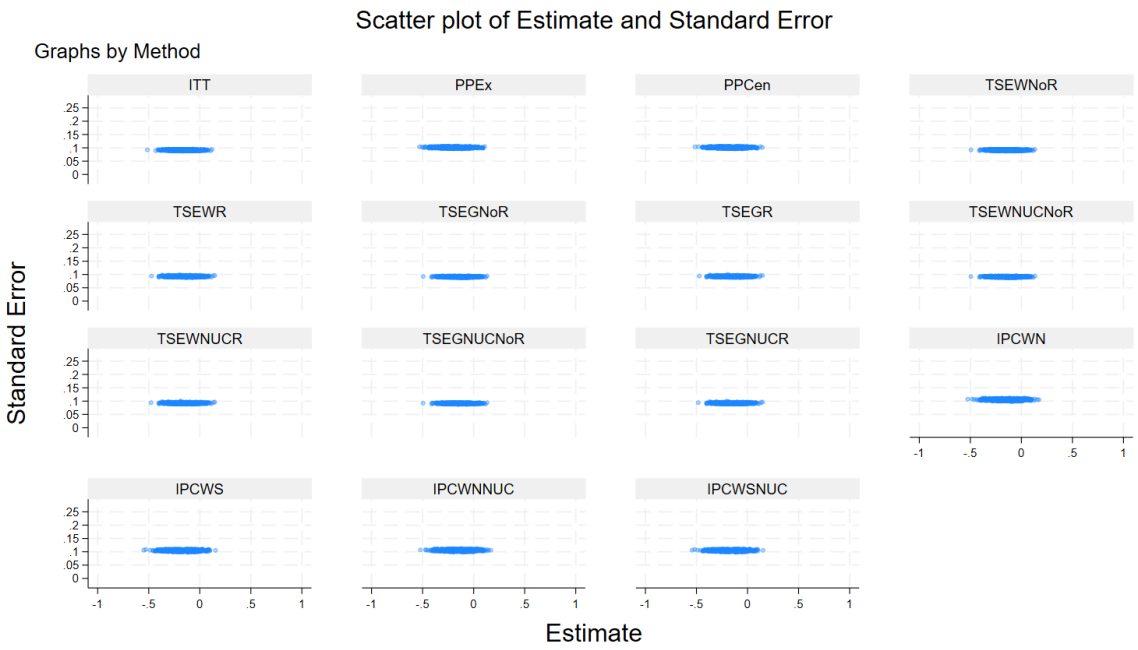
Data generated without the constant treatment effect assumption: MajorityDiscontExp



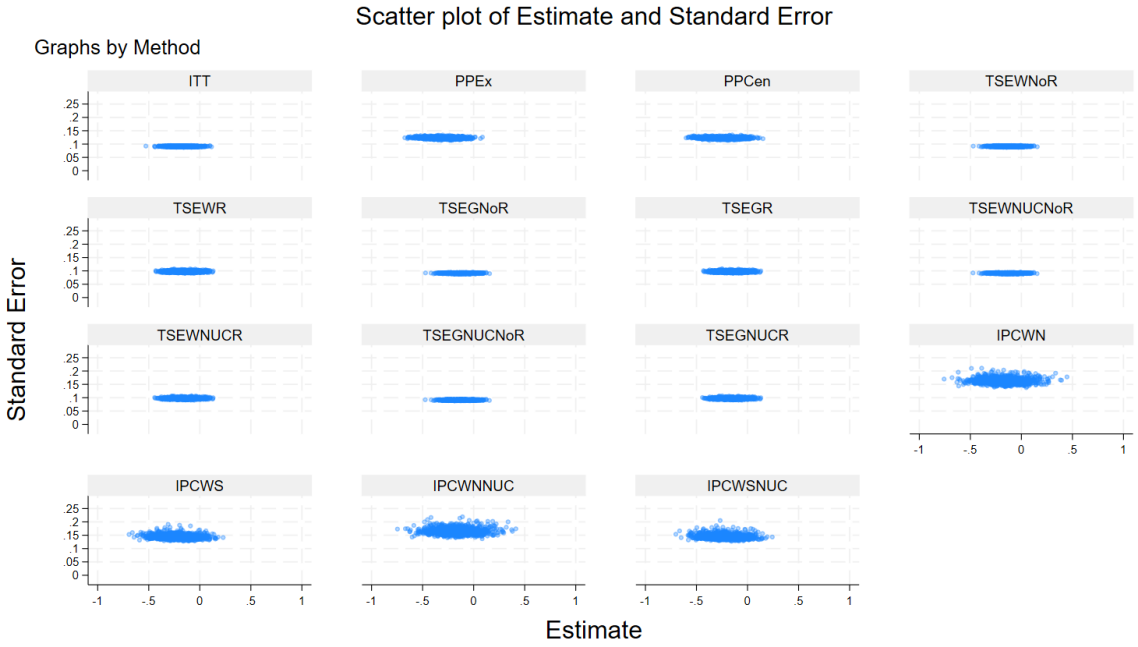
Data generated without the constant treatment effect assumption: EarlyDiscontinuation



Data generated without the constant treatment effect assumption: LateDiscontinuation

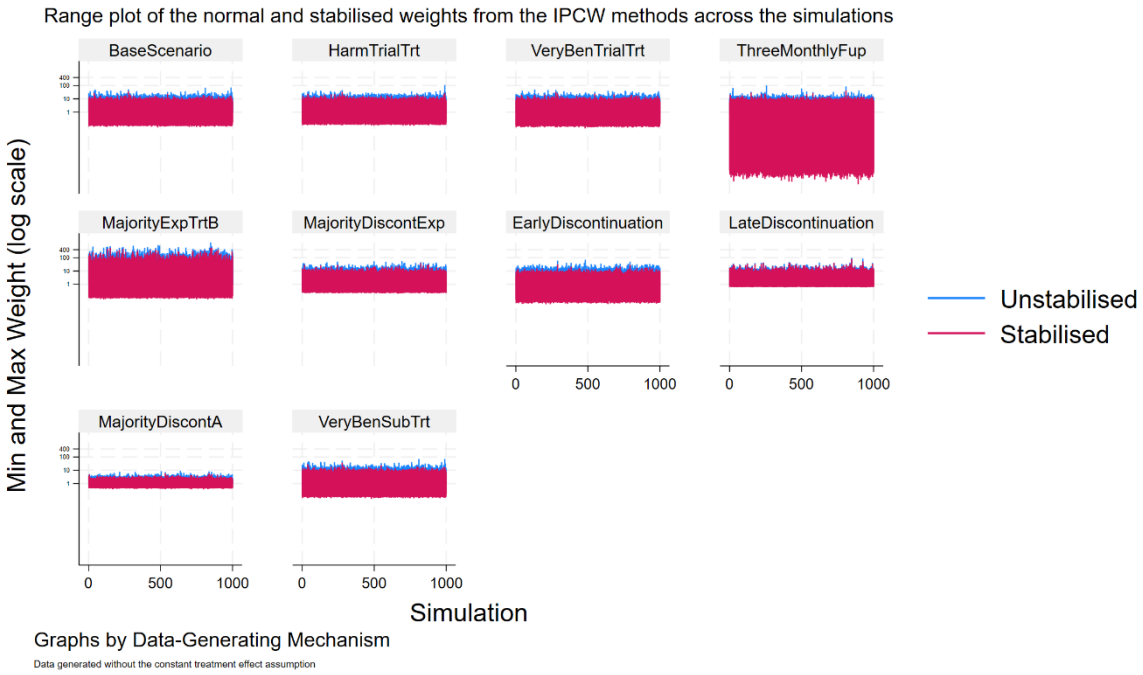


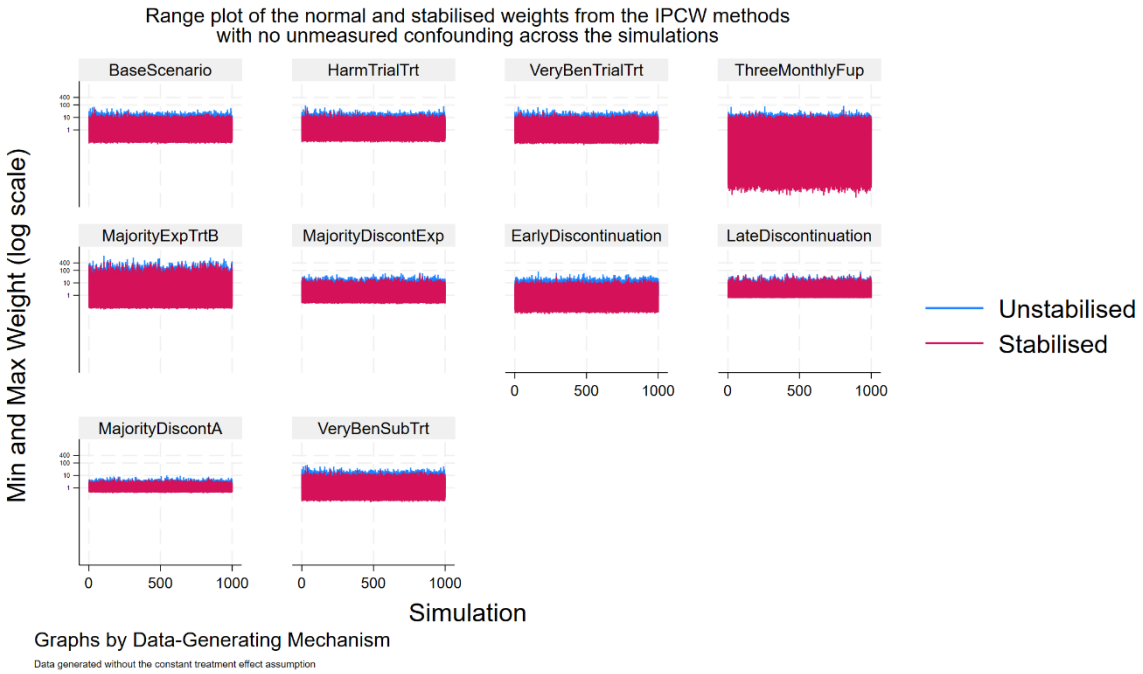
Data generated without the constant treatment effect assumption: MajorityDiscontA



Data generated without the constant treatment effect assumption: VeryBenSubTrt

Appendix Figure D.24: Range Plot of the Unstabilised and Stabilised weights from the IPCW Methods Across the Simulations and Scenarios - Without the constant treatment effect assumption





In terms of AFT model convergence, the TSE method using the generalised gamma model (TSEGN, TSEGR, TSEGNUCNoR, TSEWNUCR) failed to converge on a number of occurrences (Appendix Table D.10). This is more noticeable in LateDiscontinuation when there is less information available for AFT model building. However, if model selection had been implemented, the Weibull model would have been selected for the majority of repetition across all scenarios. In terms of the methods with and without unmeasured confounding. Notable changes in the proportion of repetitions not converging occur in MajorityDiscontExp and LateDiscontinuation. This may be due to the fact there is less data available in these scenarios for AFT model fitting but more variables in the NUC methods. This is similar to what was seen when the constant treatment effect assumption was in place (Appendix Table D.6). In addition, the mean subsequent anti-cancer therapy effect (A vs B) derived from the AFT models was similar to the reciprocal of the inputted value (Table 1: 1/0.8 = 1.25 or 1/0.6 = 1.67) (Appendix Table D.11).

Appendix Table D.10: AIC Goodness of Fit Comparisons for TSE AFT Models – Without the constant treatment effect assumption

	Unmeasured Confounding			No Unmeasured Confounding		
	Gamma	**Weibull**	**N/A**	**Gamma**	**Weibull**	**N/A**
BaseScenario	204 (20.40%)	771 (77.10%)	25 (2.50%)	196 (19.60%)	748 (74.80%)	56 (5.60%)
HarmTrialTrt	194 (19.40%)	773 (77.30%)	33 (3.30%)	176 (17.60%)	758 (75.80%)	66 (6.60%)
VeryBenTrialTrt	203 (20.30%)	774 (77.40%)	23 (2.30%)	193 (19.30%)	755 (75.50%)	52 (5.20%)
ThreeMonthlyFup	174 (17.40%)	819 (81.90%)	7 (0.70%)	175 (17.50%)	808 (80.80%)	17 (1.70%)
MajorityExpTrtB	179 (17.90%)	790 (79.00%)	31 (3.10%)	182 (18.20%)	763 (76.30%)	55 (5.50%)
MajorityDiscontExp	168 (16.80%)	772 (77.20%)	60 (6.00%)	164 (16.40%)	725 (72.50%)	111 (11.10%)
EarlyDiscontinuation	173 (17.30%)	814 (81.40%)	13 (1.30%)	180 (18.00%)	795 (79.50%)	25 (2.50%)
LateDiscontinuation	110 (11.00%)	610 (61.00%)	280 (28.00%)	88 (8.80%)	455 (45.50%)	457 (45.70%)
MajorityDiscontA	190 (19.00%)	775 (77.50%)	35 (3.50%)	181 (18.10%)	735 (73.50%)	84 (8.40%)
VeryBenSubTrt	175 (17.50%)	787 (78.70%)	38 (3.80%)	180 (18.00%)	748 (74.80%)	72 (7.20%)

N/A: Indicates that the generalised gamma model did not converge so no comparison can be made. Percentages have been calculated out of the 1000 repetitions for each method within in scenario.

Appendix Table D.11: Mean (SD) of Theta from the TSE Methods Calculated Across the Simulations

	TSEGNUCNoR	TSEGNoR	TSEWNUCNoR	TSEWNoR
BaseScenario	1.25 (0.17)	1.26 (0.17)	1.26 (0.18)	1.27 (0.18)
HarmTrialTrt	1.25 (0.18)	1.26 (0.18)	1.26 (0.19)	1.27 (0.19)
VeryBenTrialTrt	1.25 (0.17)	1.26 (0.17)	1.26 (0.17)	1.26 (0.17)
ThreeMonthlyFup	1.26 (0.17)	1.26 (0.17)	1.26 (0.17)	1.26 (0.17)
MajorityExpTrtB	1.24 (0.22)	1.25 (0.22)	1.25 (0.22)	1.26 (0.22)
MajorityDiscontExp	1.25 (0.20)	1.26 (0.20)	1.25 (0.20)	1.26 (0.20)
EarlyDiscontinuation	1.23 (0.15)	1.24 (0.15)	1.24 (0.16)	1.25 (0.16)
LateDiscontinuation	1.29 (0.36)	1.30 (0.35)	1.30 (0.37)	1.31 (0.36)
MajorityDiscontA	1.25 (0.18)	1.25 (0.18)	1.25 (0.18)	1.26 (0.18)
VeryBenSubTrt	1.64 (0.26)	1.66 (0.26)	1.66 (0.26)	1.67 (0.26)

In terms of proportional hazards in almost all scenarios and methods, randomisation allocation violated the proportional hazards assumption above the 5% nominal level. This was expected as the same trial treatment effect is not the in play throughout all of follow-up. LateDiscontinuation is the best scenario across the methods. This is not surprising as the trial treatment effect is in play supporting proportional hazards for the majority of this scenario. On the reverse VeryBenTrialTrt has the worse performance. This may be because there is a bigger additive effect lost between the two arms as participants discontinue.

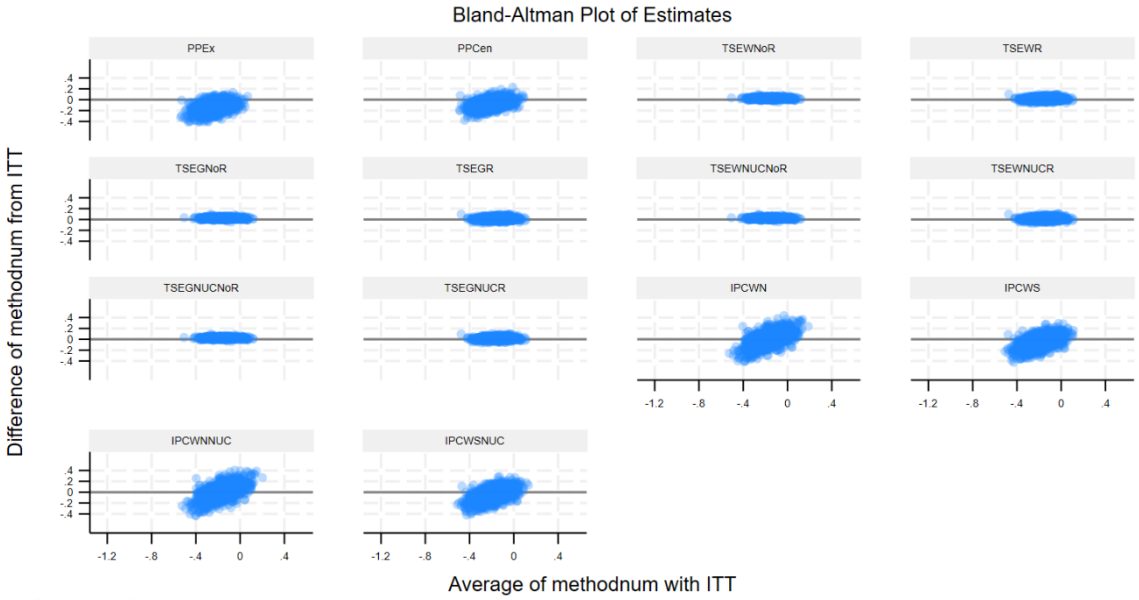
Appendix Table D.12: Percentage of Simulations Randomisation Allocation Violates Proportional Hazards Within Each Implemented Method

	ITT	PPCen	PPEx	TSEG NUCNoR	TSEG NUCR	TSEG NoR	TSEG R	TSEW NUCNoR	TSEW NUCR	TSE WNoR	TSE WR	IPCW N	IPCW NNUC	IPCW S	IPCW SNUC
BaseScenario	18%	11%	20%	19%	16%	20%	17%	21%	18%	21%	18%	56%	57%	51%	51%
HarmTrialTrt	18%	11%	7%	16%	16%	17%	16%	17%	16%	17%	16%	25%	26%	18%	19%
VeryBenTrialTrt	58%	33%	42%	58%	56%	61%	58%	62%	60%	62%	60%	80%	80%	77%	77%
ThreeMonthlyFup	15%	11%	25%	16%	17%	16%	17%	16%	17%	16%	18%	57%	59%	63%	62%
MajorityExpTrtB	31%	4%	8%	17%	18%	17%	18%	18%	19%	18%	19%	81%	80%	78%	76%
MajorityDiscontExp	12%	6%	14%	11%	11%	11%	11%	12%	13%	12%	13%	32%	32%	28%	28%
EarlyDiscontinuation	15%	12%	25%	16%	15%	16%	16%	16%	15%	16%	16%	52%	53%	50%	50%
LateDiscontinuation	10%	6%	5%	7%	6%	9%	8%	12%	10%	12%	10%	22%	24%	19%	20%
MajorityDiscontA	18%	17%	23%	17%	17%	18%	18%	19%	20%	19%	19%	31%	31%	32%	31%
VeryBenSubTrt	15%	11%	23%	18%	17%	18%	19%	20%	19%	20%	19%	57%	58%	50%	51%

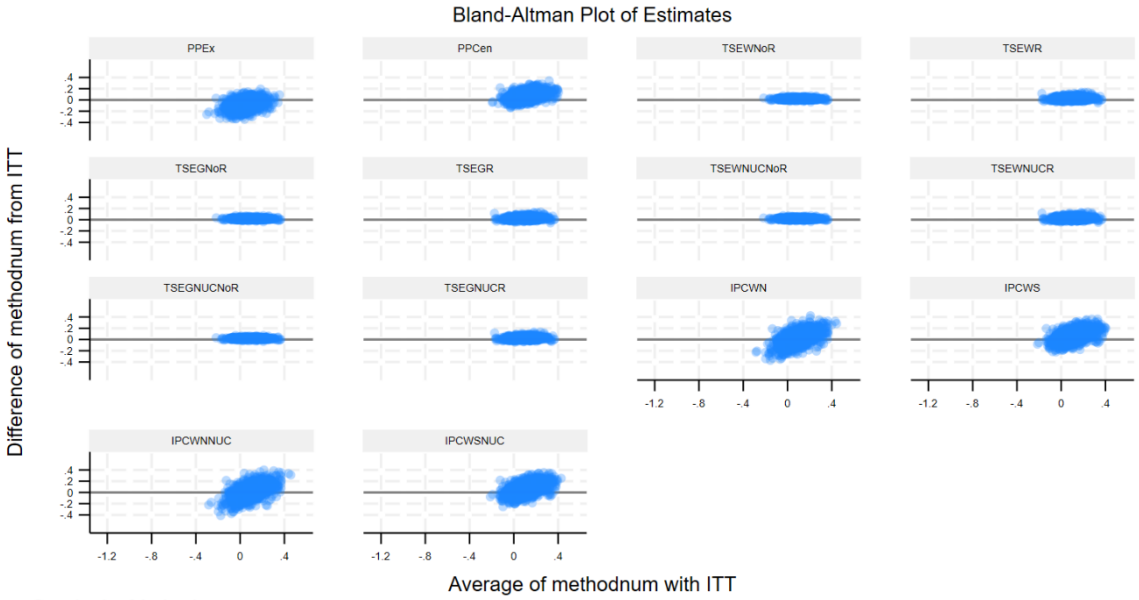
Overall, the TSE method estimates are the most similar to the ITT method estimates across all scenarios (Appendix Figure D.25). For the other methods, the similarities depend on the amount of censoring. This can be seen through how all methods are similar to ITT in MajorityDiscontA and LateDiscontinuation when there is the least impact of the intercurrent event. This is similar to what was observed when the constant treatment effect assumption was in place (Appendix Figure D.19).

The performance of the methods also depends on the impact of the intercurrent event (Appendix Figure D.26) As seen when the constant treatment effect assumption was in place (Figure 4.3 and Appendix Figure D.20) PPEx performs the worst across all the different scenarios. However, in general all methods perform worse in terms of bias and power. The TSE methods are the most consistent performers across the different approaches.

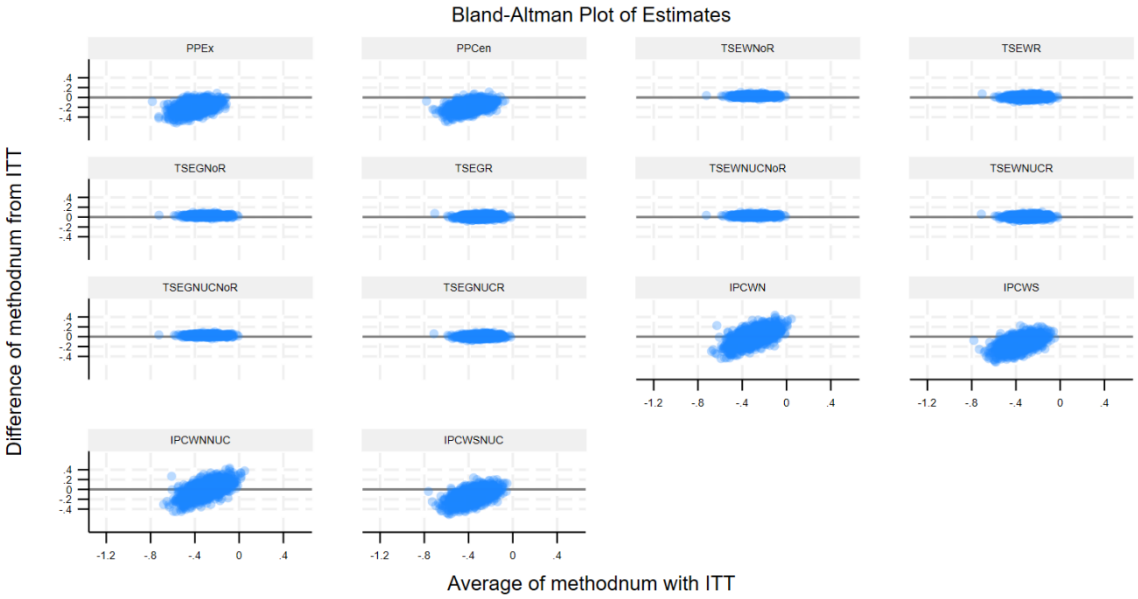
Appendix Figure D.25: Bland-Altman Plots of Estimates Across the Simulations - Without the constant treatment effect assumption



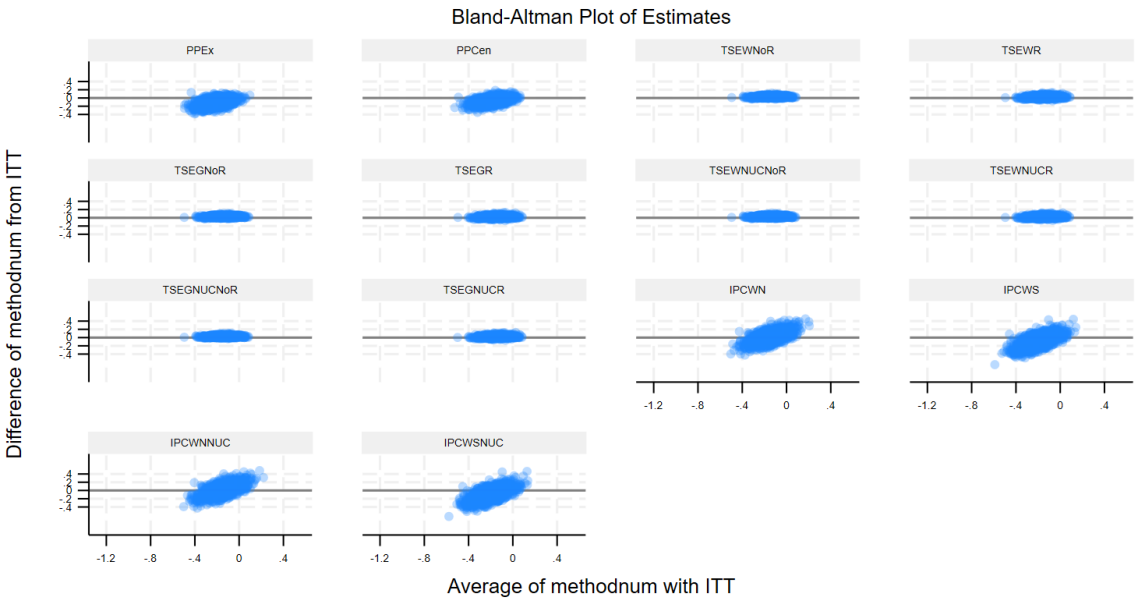
Graphs by Method
Estimates (coeff) for dgmnum=BaseScenario
Data generated without the constant treatment effect assumption



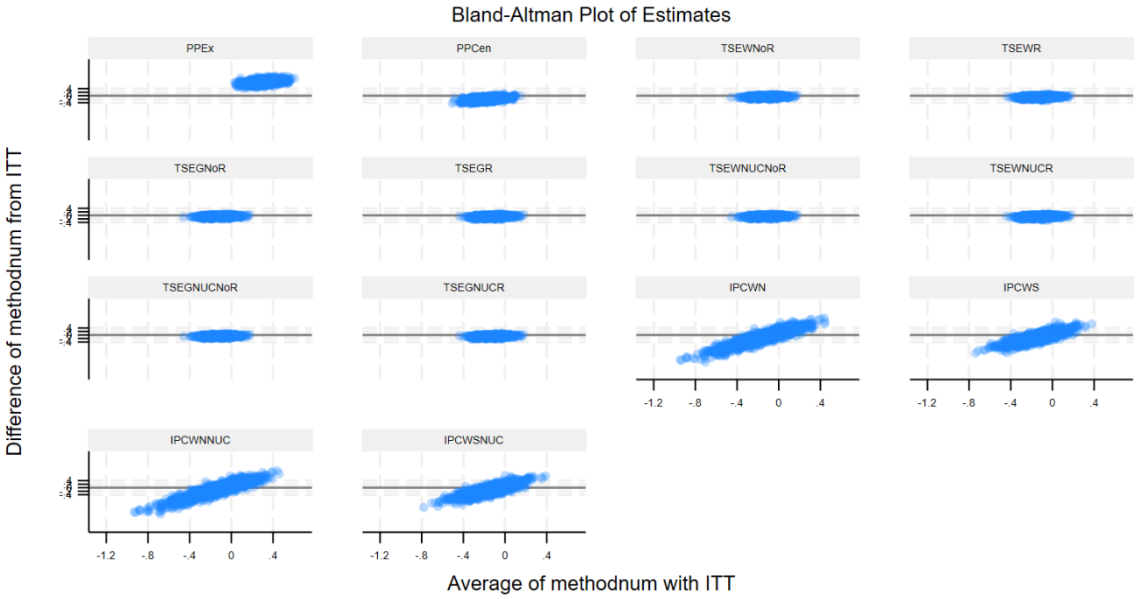
Graphs by Method
Estimates (coeff) for dgmnum=HarmTrialTrt
Data generated without the constant treatment effect assumption



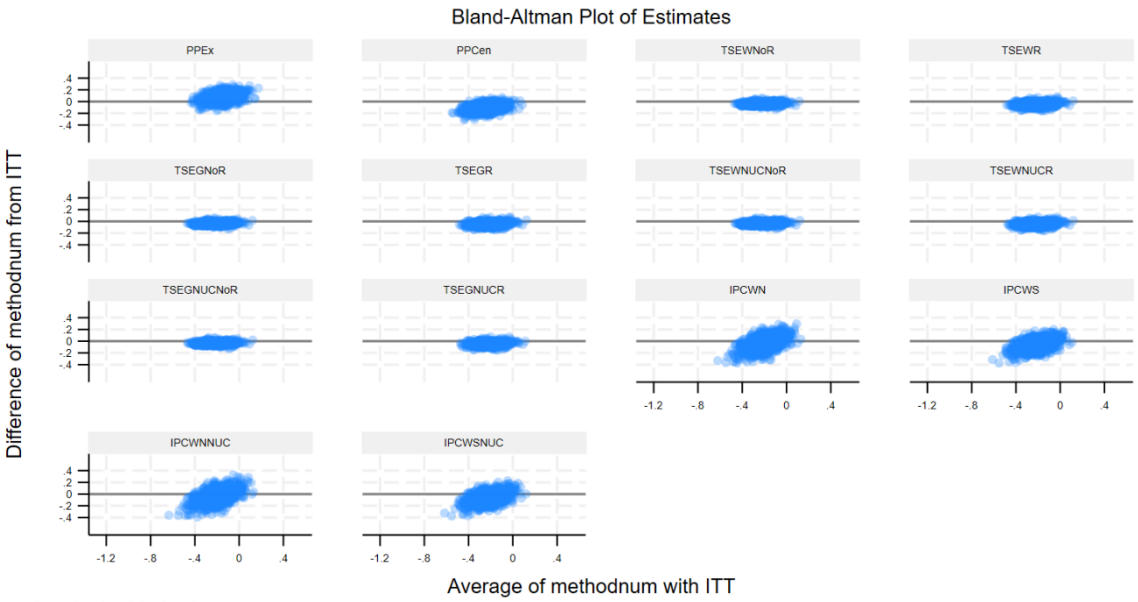
Graphs by Method
Estimates (coeff) for dgmnum=VeryBenTrialTrt
Data generated without the constant treatment effect assumption



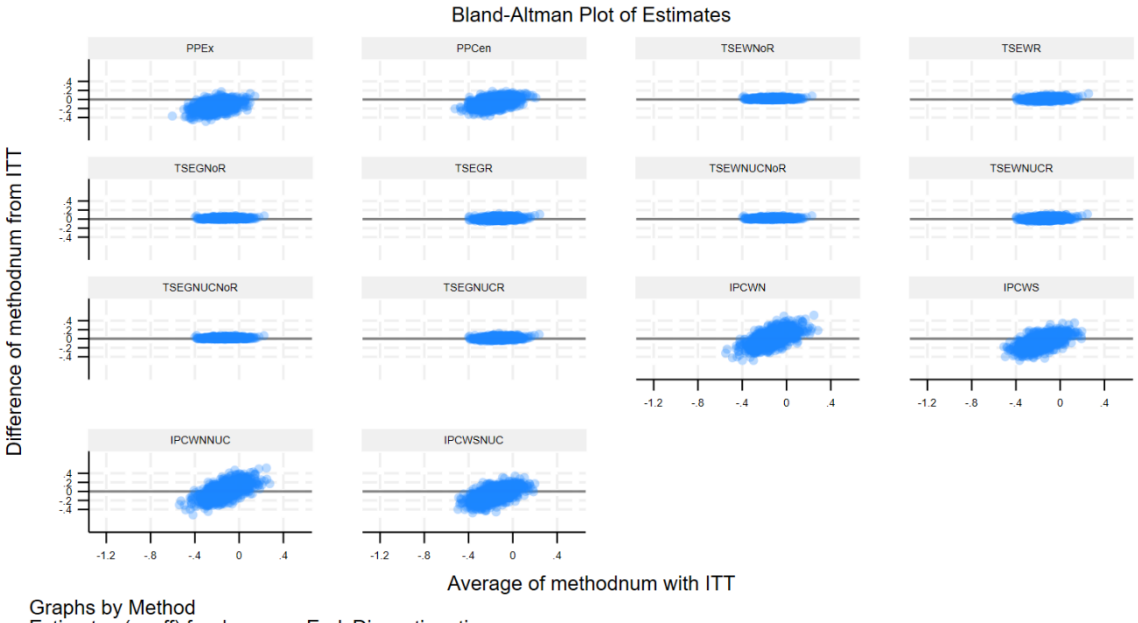
Graphs by Method
Estimates (coeff) for dgmnum=ThreeMonthlyFup
Data generated without the constant treatment effect assumption



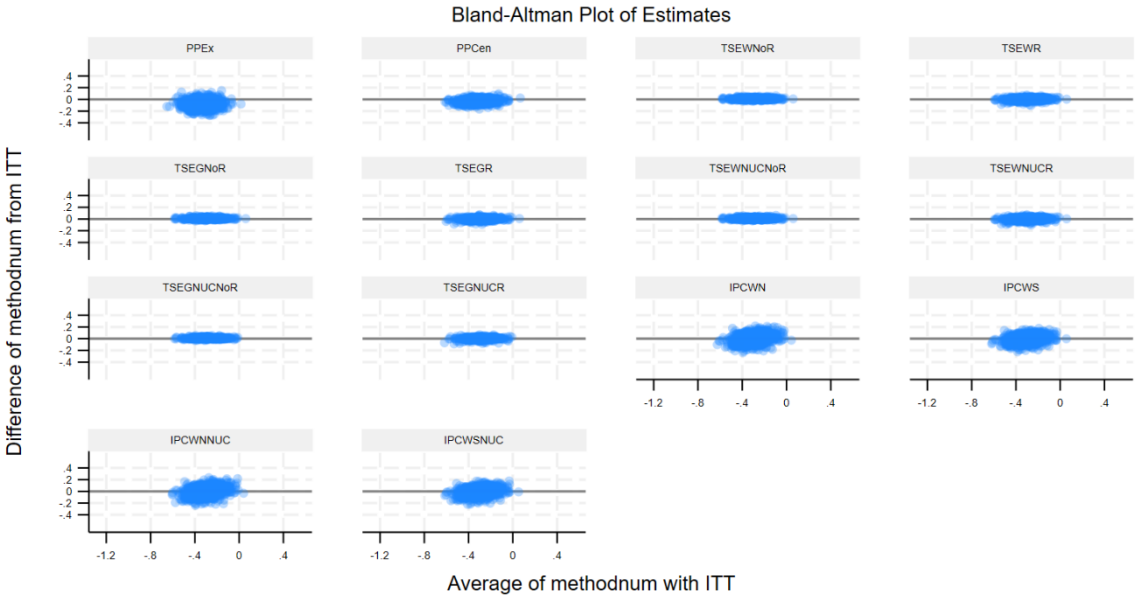
Graphs by Method
Estimates (coeff) for dgmnum=MajorityExpTrtB
Data generated without the constant treatment effect assumption



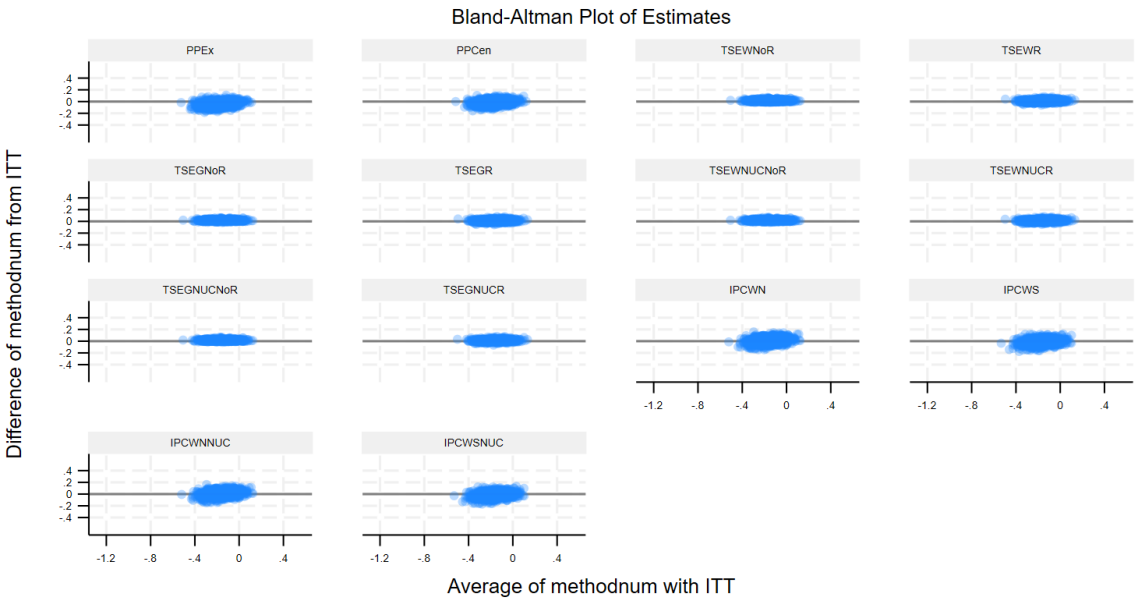
Graphs by Method
Estimates (coeff) for dgmnum=MajorityDiscontExp
Data generated without the constant treatment effect assumption



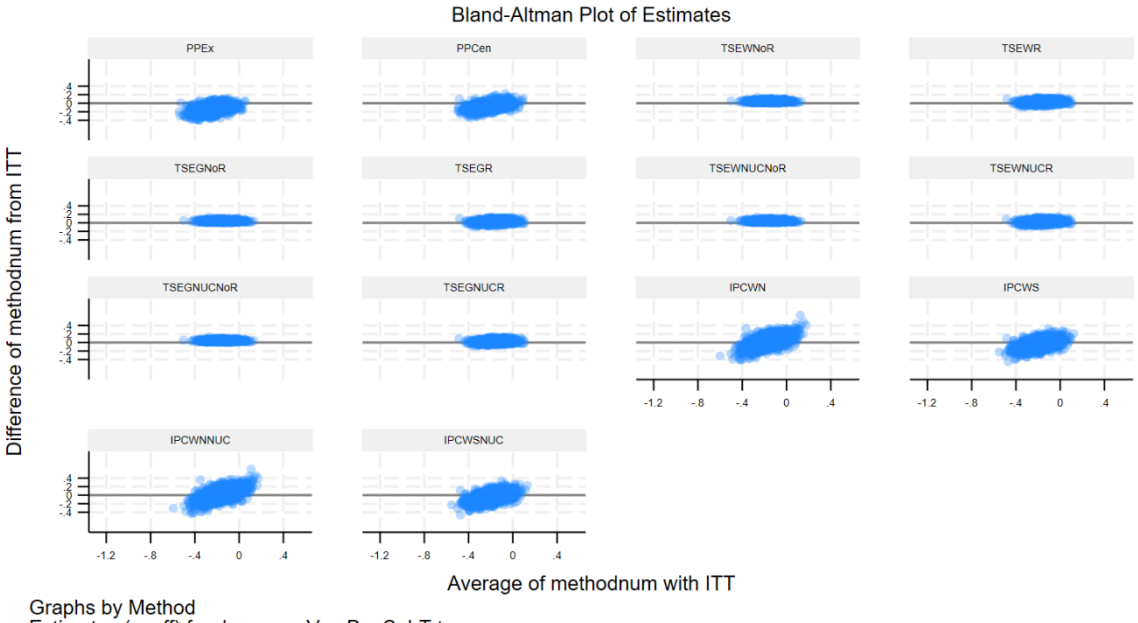
Graphs by Method
Estimates (coeff) for dgmnum=EarlyDiscontinuation
Data generated without the constant treatment effect assumption



Graphs by Method
Estimates (coeff) for dgmnum=LateDiscontinuation
Data generated without the constant treatment effect assumption

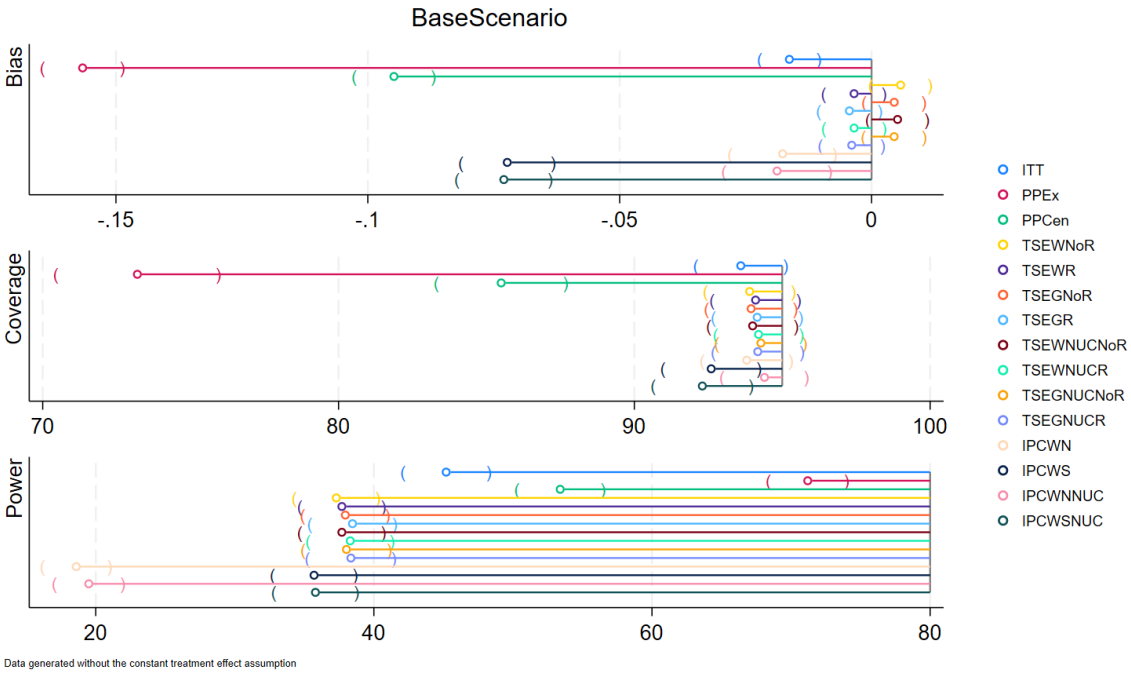


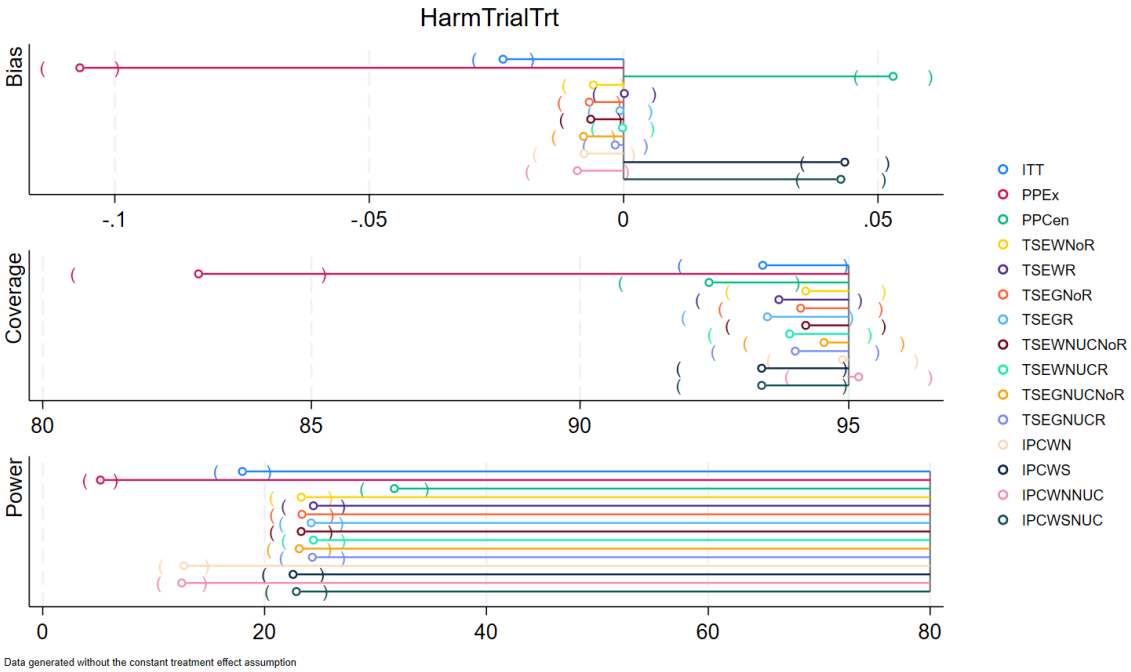
Graphs by Method
Estimates (coeff) for dgmnum=MajorityDiscontA
Data generated without the constant treatment effect assumption

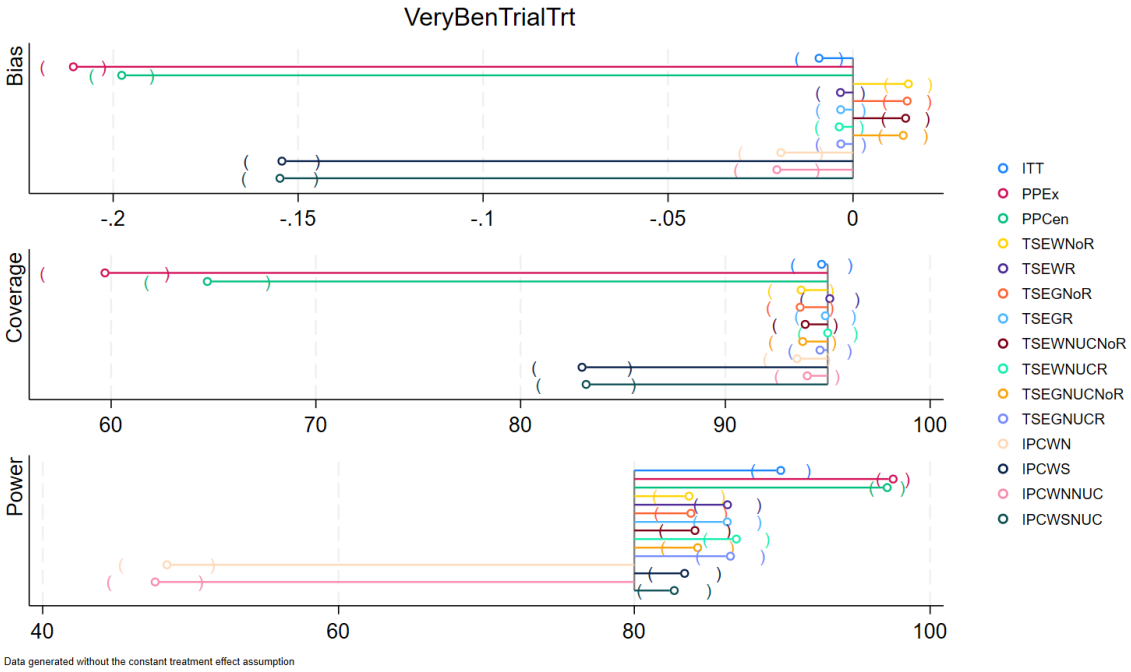


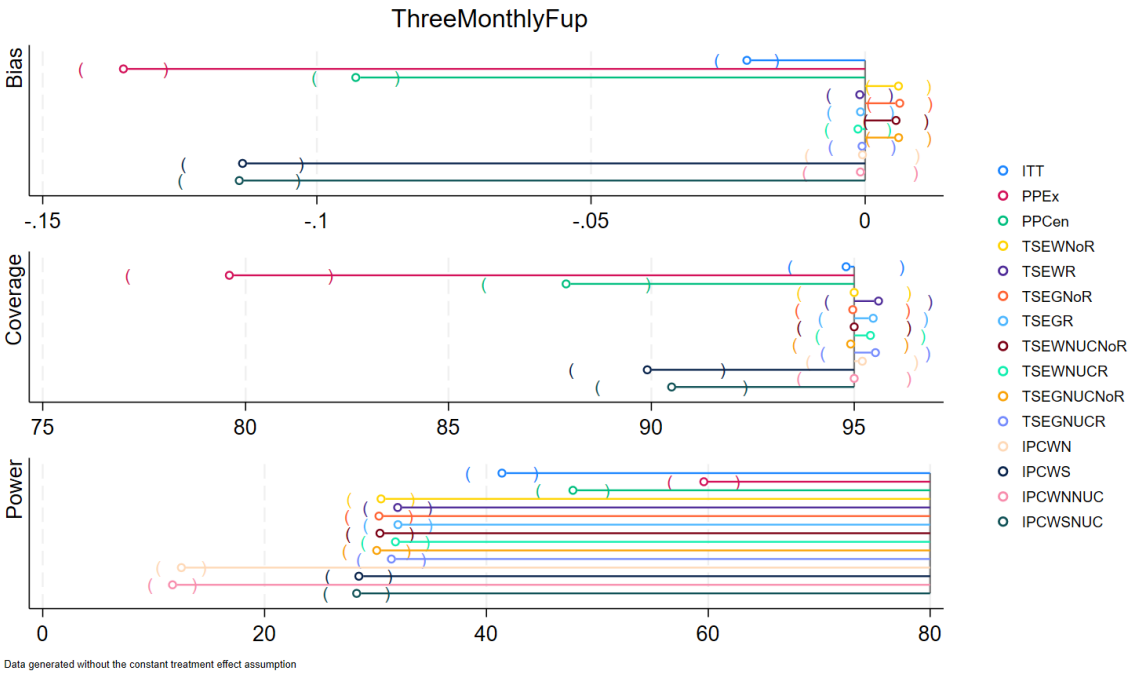
Graphs by Method
Estimates (coeff) for dgmnum=VeryBenSubTrt
Data generated without the constant treatment effect assumption

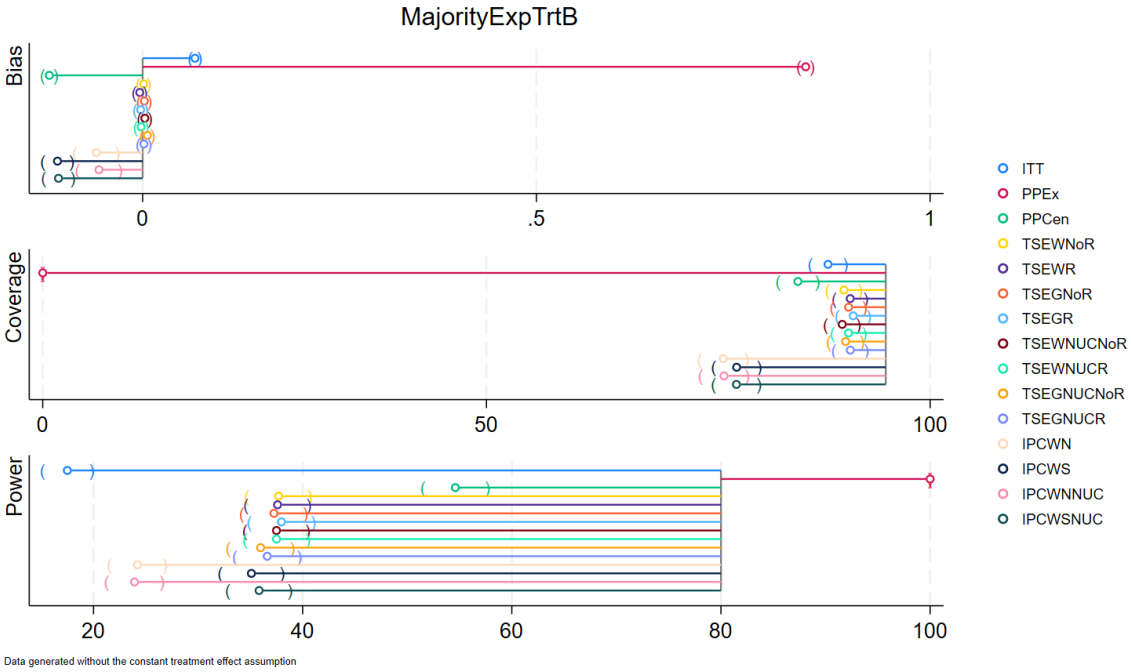
Appendix Figure D.26: Lollipop Plots of Performance Measures from Simulations - Without the constant treatment effect assumption

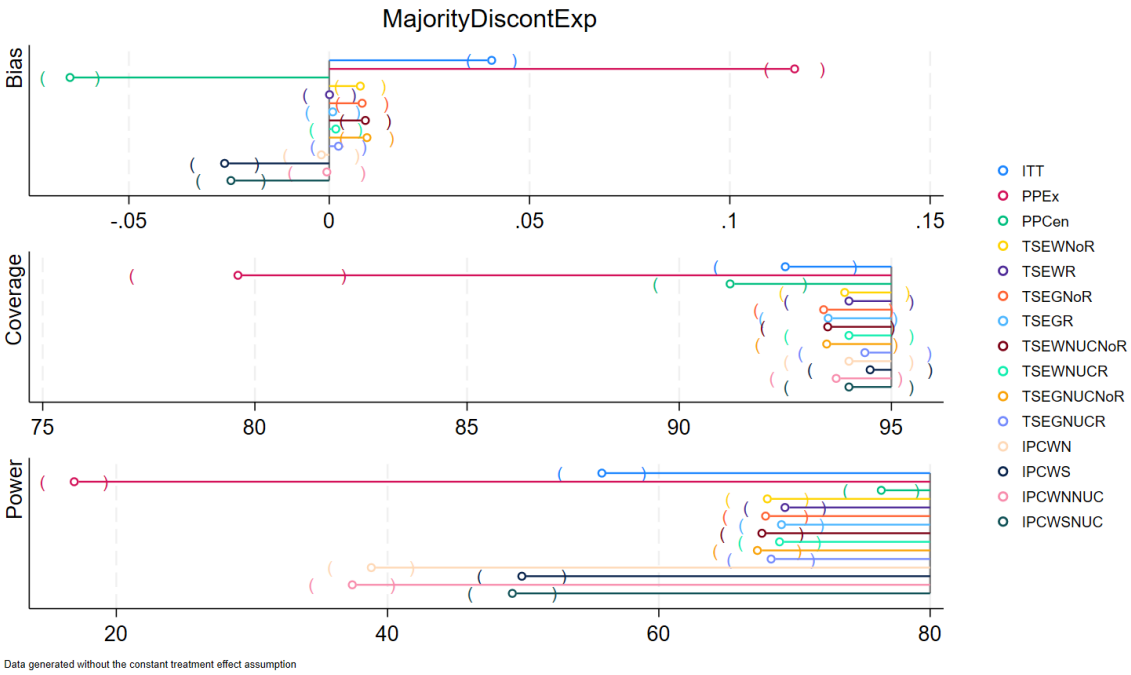


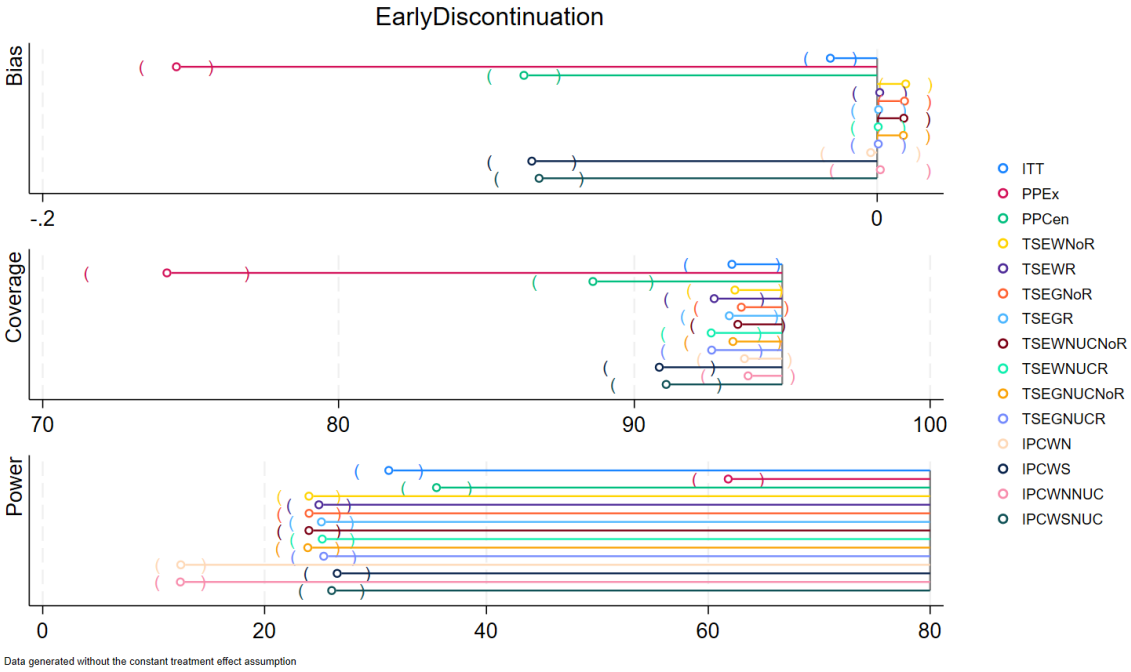


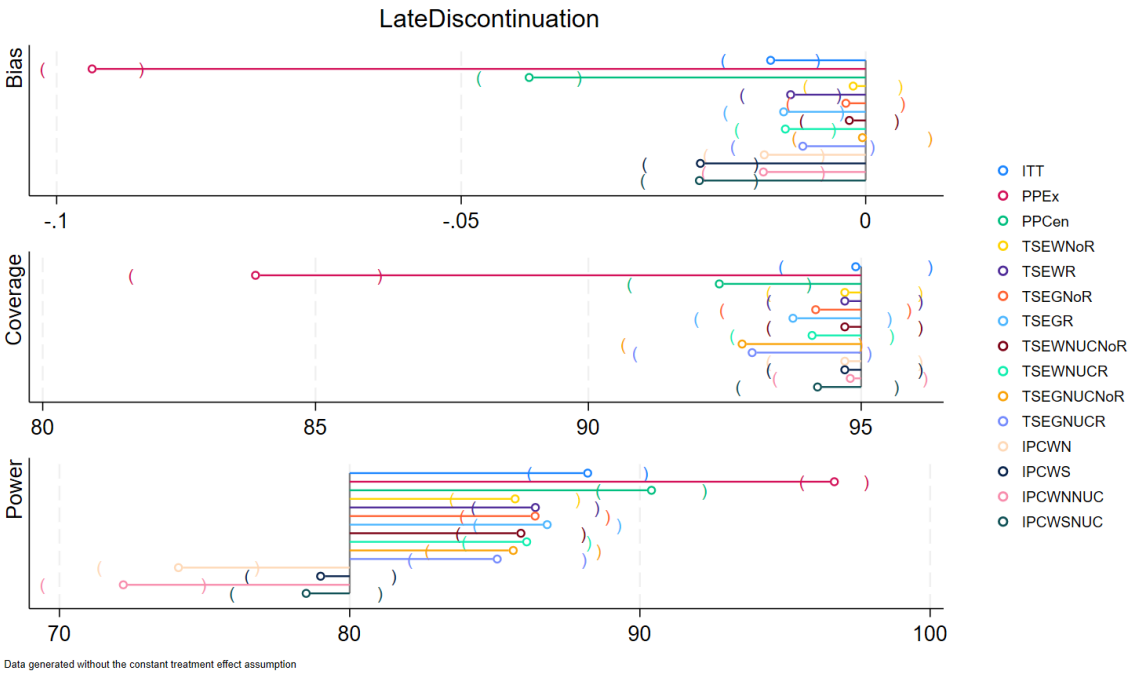

VeryBenTrialTrt
Bias
-.2
-.15
-.1
-.05
0
Coverage
60
70
80
90
100
Power
40
60
80
100
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

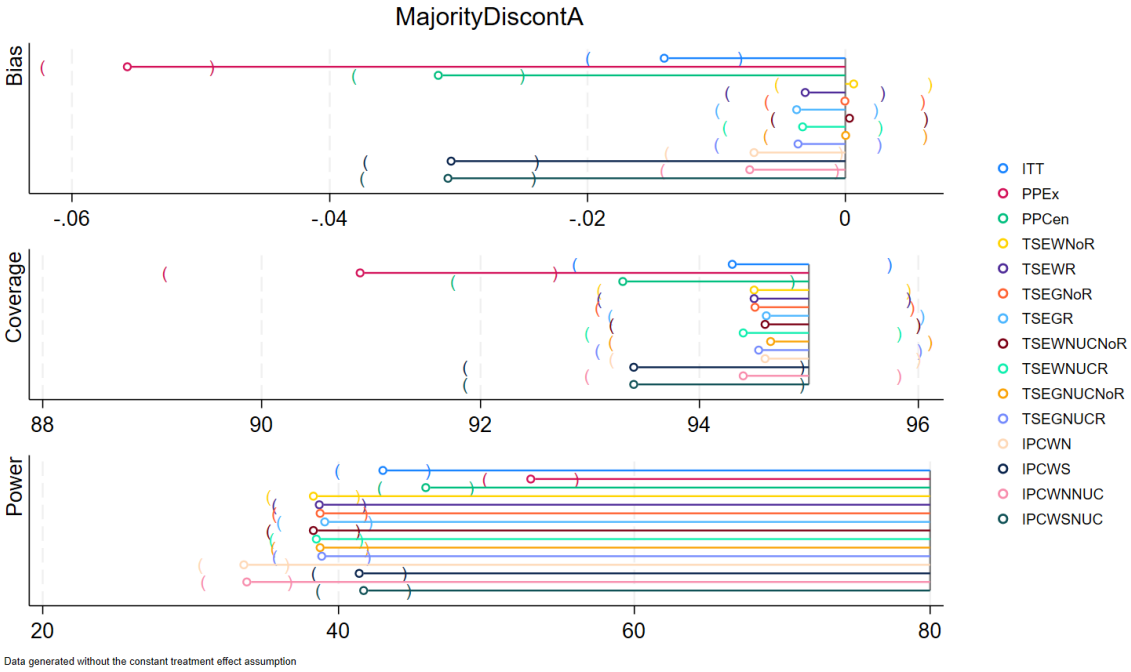

ThreeMonthlyFup
Bias
-.15
-.1
-.05
0
Coverage
75
80
85
90
95
Power
0
20
40
60
80
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

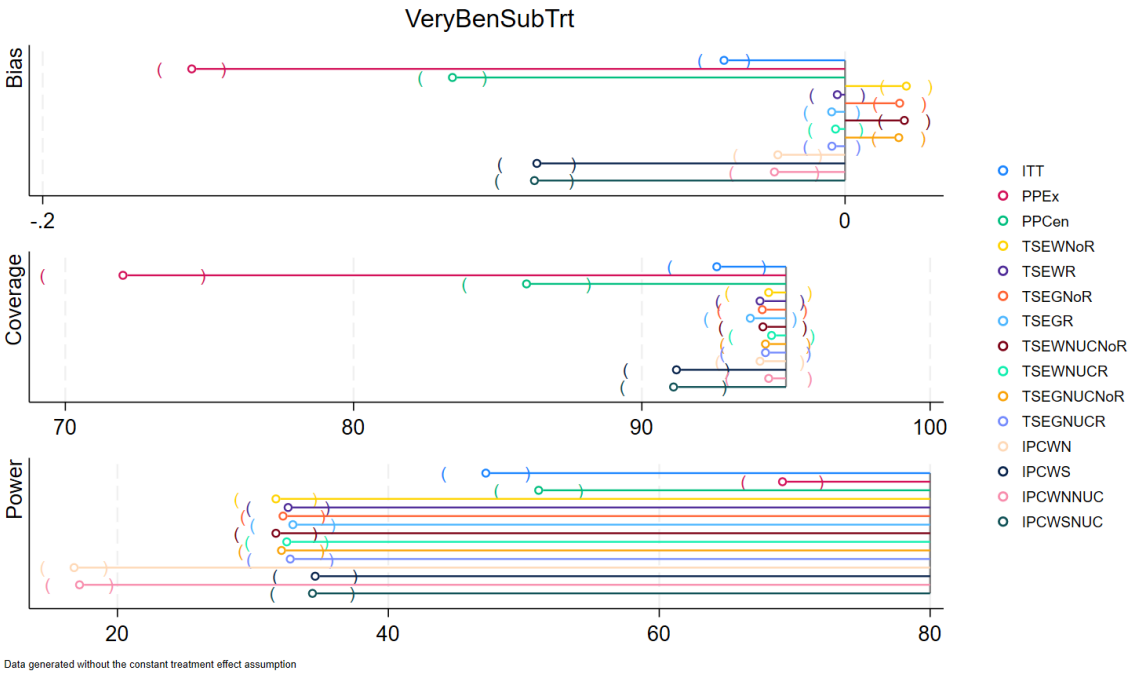

MajorityExpTrtB
Bias
0
.5
1
Coverage
0
50
100
Power
20
40
60
80
100
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC


MajorityDiscontExp
Bias
-.05
0
.05
.1
.15
Coverage
75
80
85
90
95
Power
20
40
60
80
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

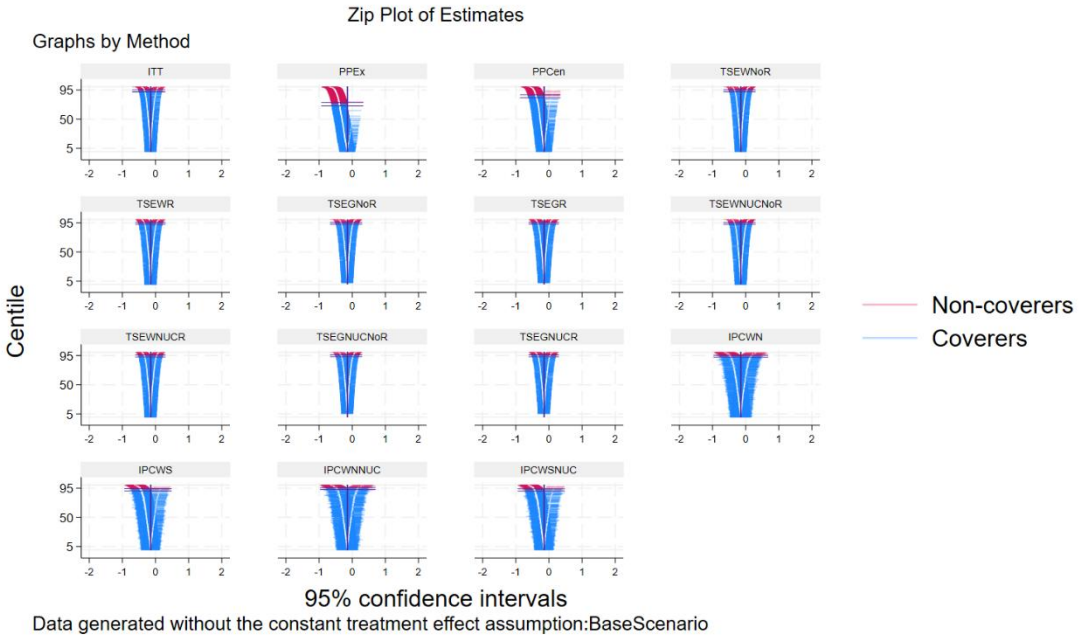

EarlyDiscontinuation
Bias
-.2
0
Coverage
70
80
90
100
Power
0
20
40
60
80
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

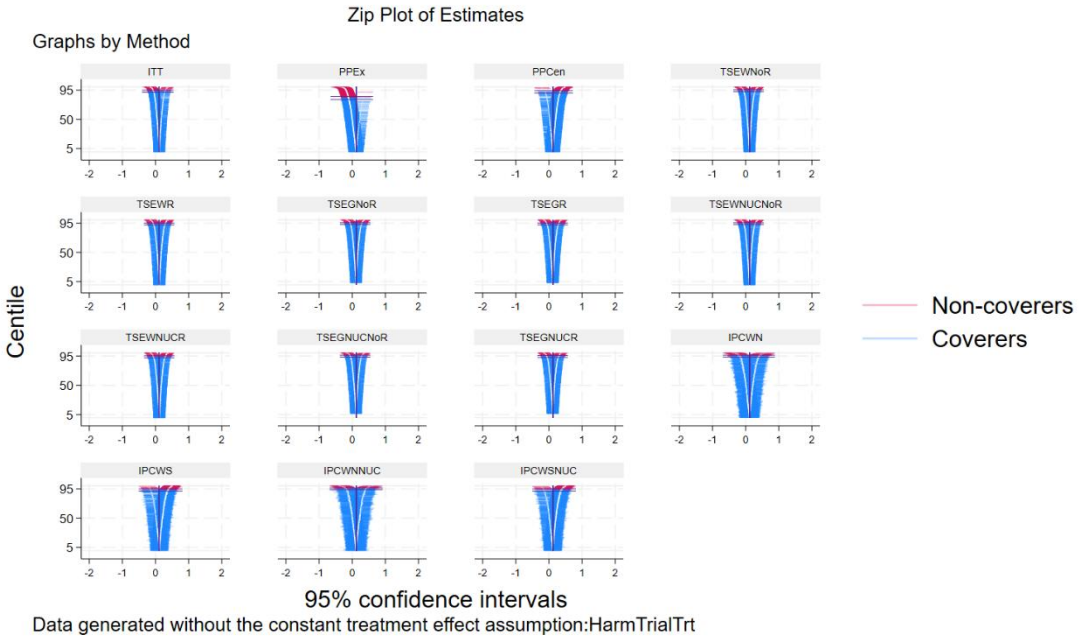

LateDiscontinuation
Bias
-.1
-.05
0
Coverage
80
85
90
95
Power
70
80
90
100
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

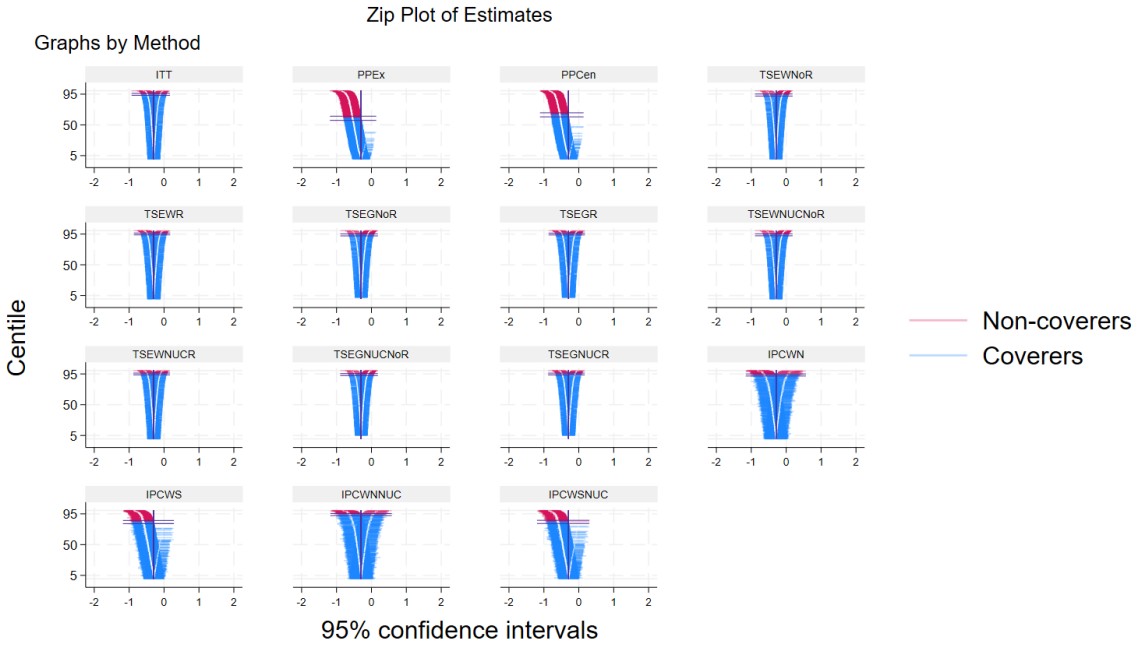

MajorityDiscontA
Bias
-.06
-.04
-.02
0
Coverage
88
90
92
94
96
Power
20
40
60
80
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC


VeryBenSubTrt
Bias
-.2
0
Coverage
70
80
90
100
Power
20
40
60
80
Data generated without the constant treatment effect assumption
ITT
PPEx
PPCen
TSEWNoR
TSEWR
TSEGNoR
TSEGR
TSEWNUCNoR
TSEWNUCR
TSEGNUCNoR
TSEGNUCR
IPCWN
IPCWS
IPCWNNUC
IPCWSNUC

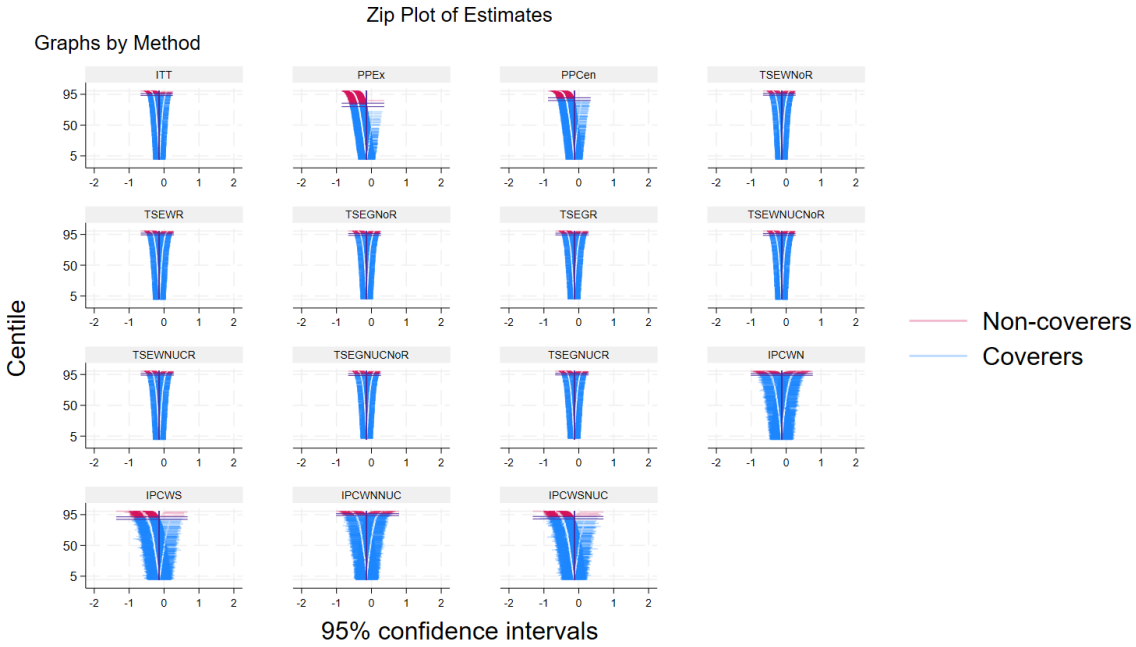
Appendix Figure D.27: Zip Plots Estimates Across the Simulations - Without the constant treatment effect assumption



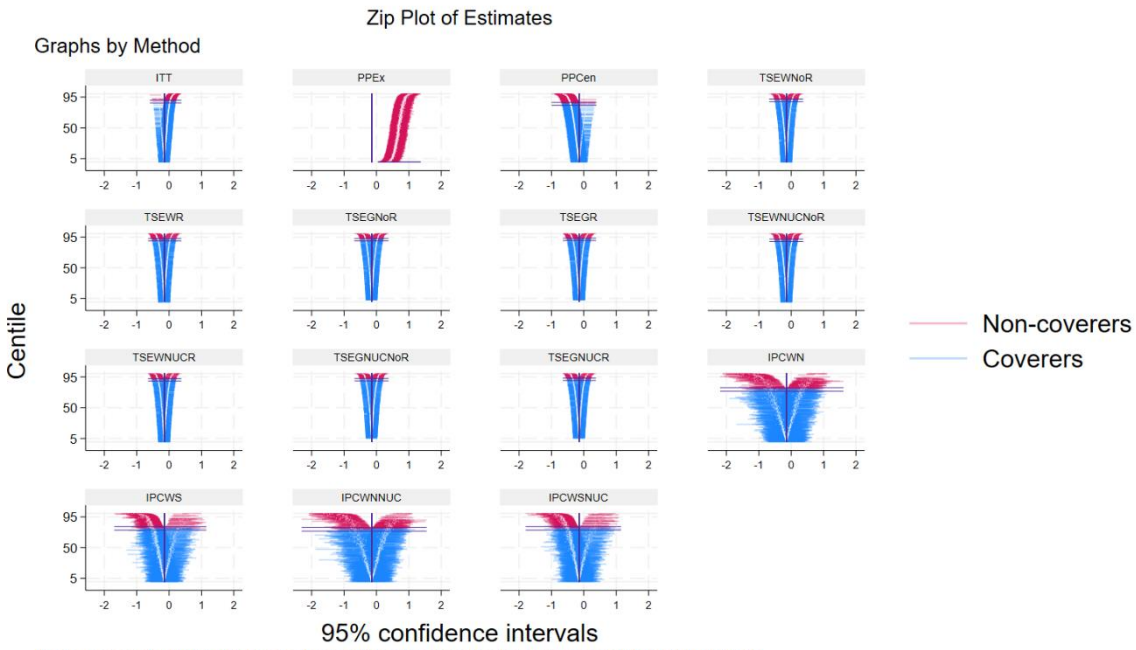




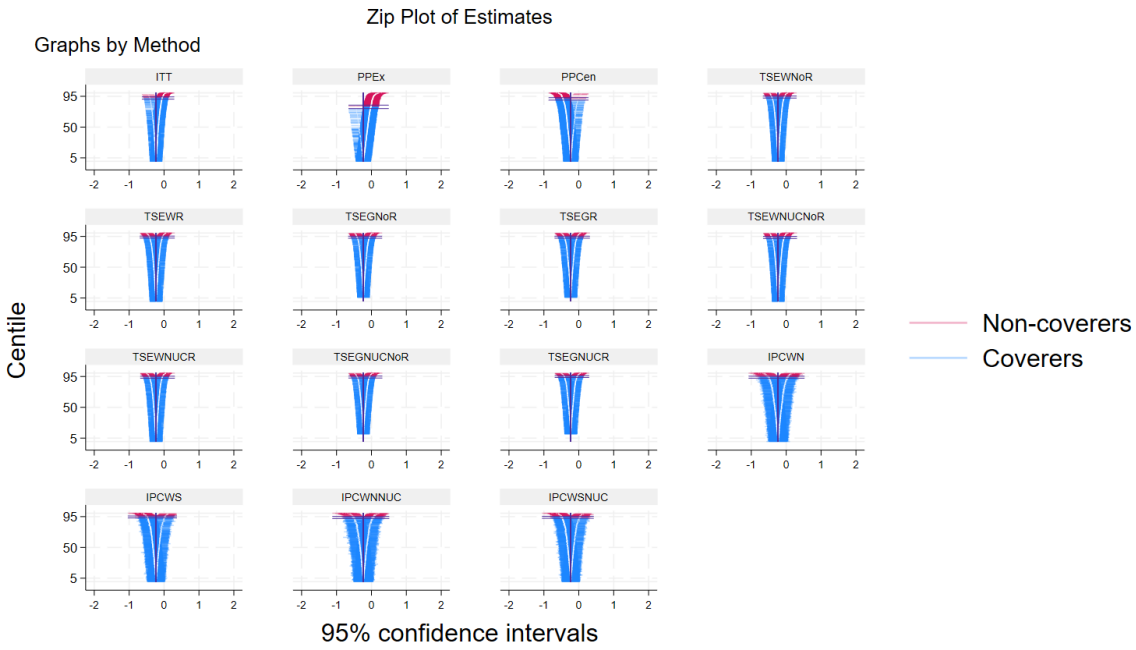
Data generated without the constant treatment effect assumption:VeryBenTrialTrt



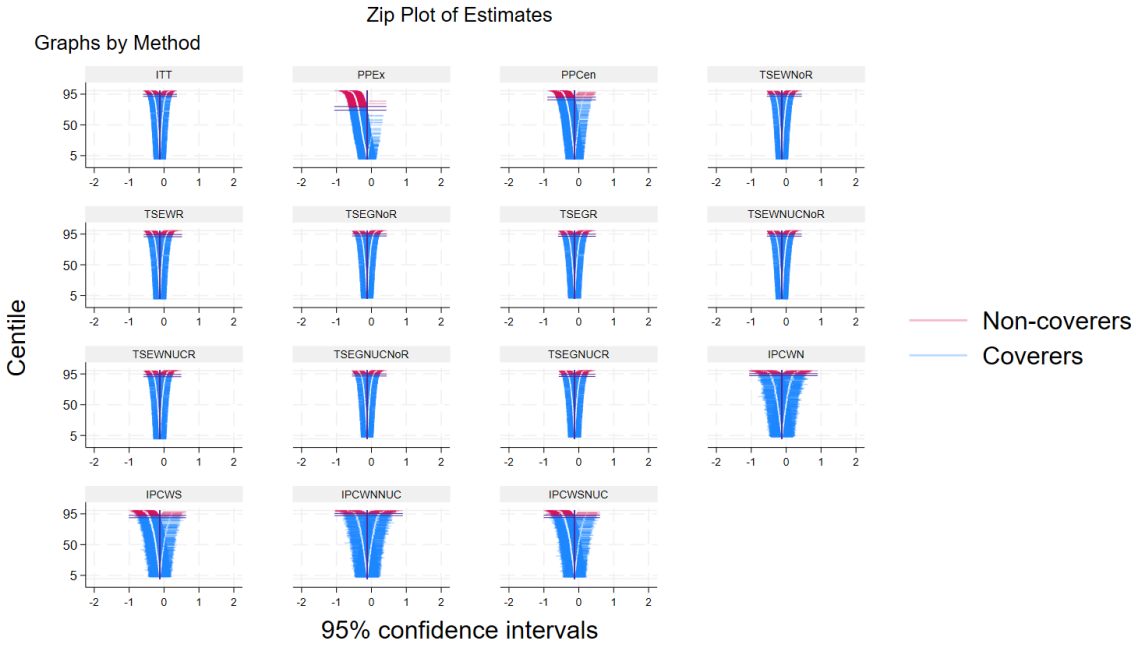
Data generated without the constant treatment effect assumption:ThreeMonthlyFup



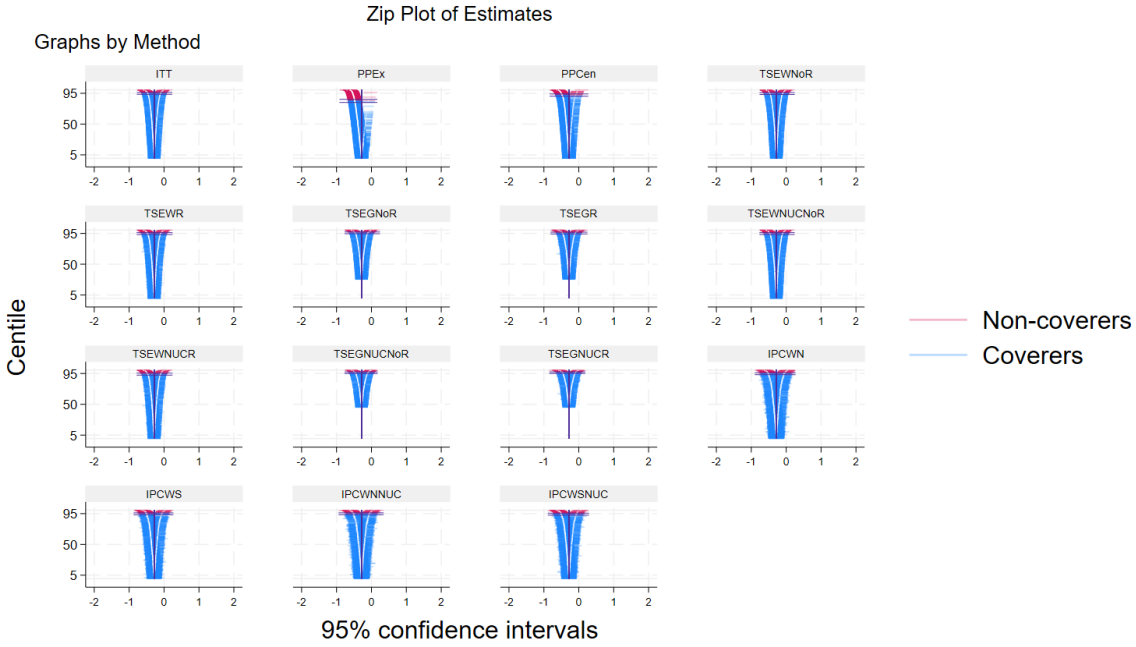
Data generated without the constant treatment effect assumption:MajorityExpTrtB



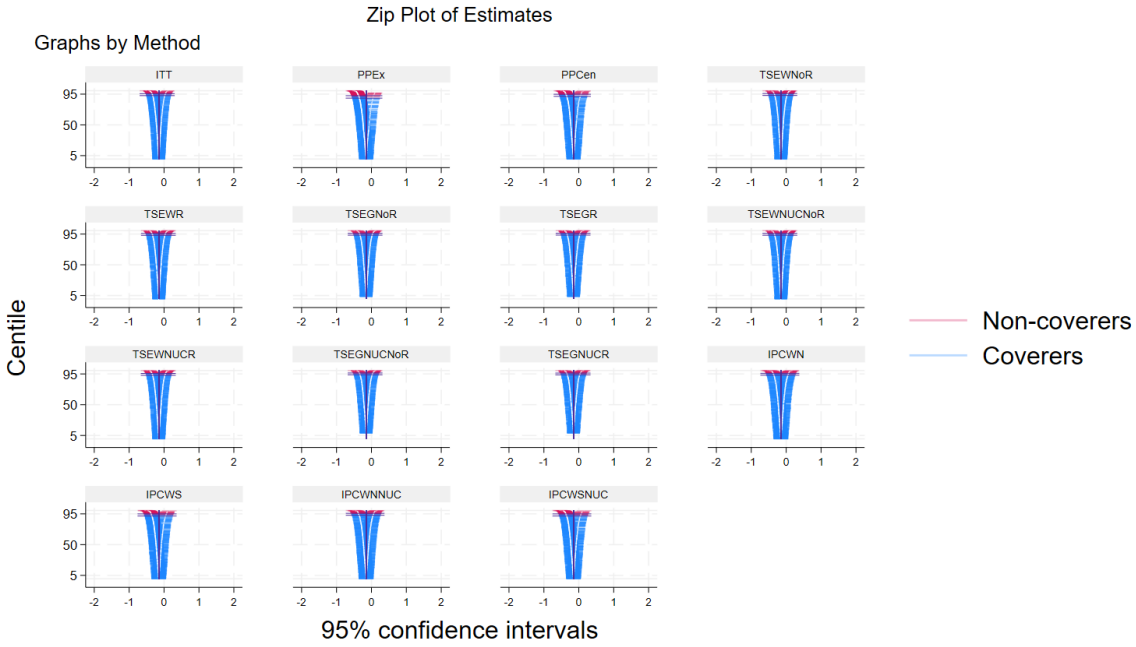
Data generated without the constant treatment effect assumption:MajorityDiscontExp



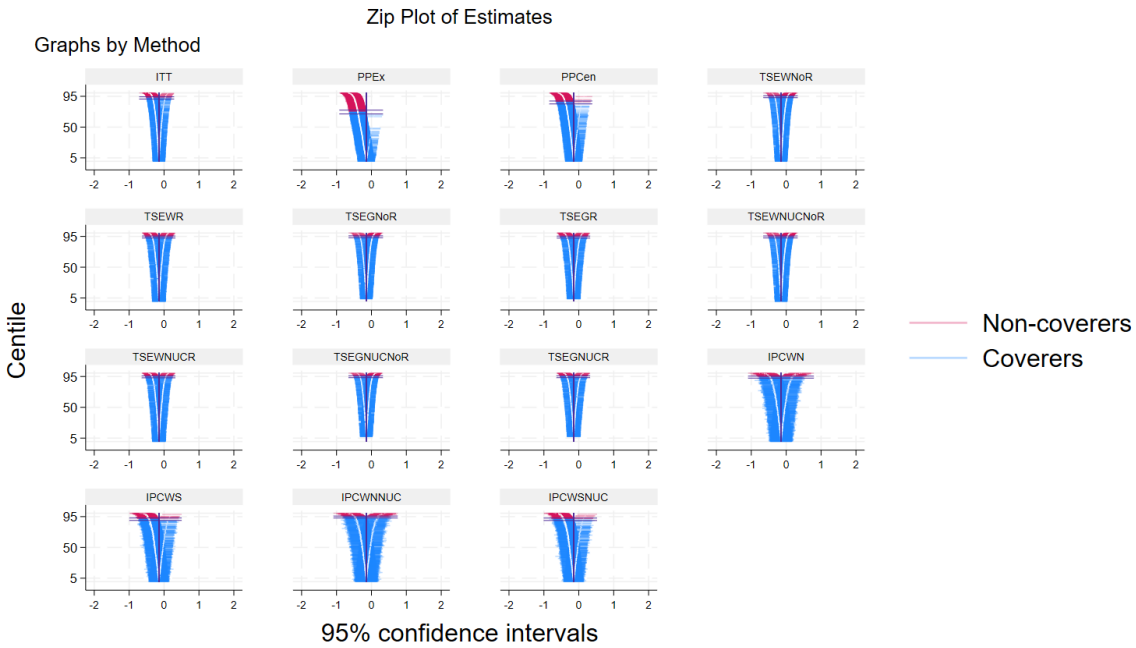
Data generated without the constant treatment effect assumption:EarlyDiscontinuation



Data generated without the constant treatment effect assumption:LateDiscontinuation



Data generated without the constant treatment effect assumption:MajorityDiscontA



Data generated without the constant treatment effect assumption:VeryBenSubTrt

D.5 Additional Case Study Results

This section provides additional results from the real data case study which are referenced in the main body of the manuscript.

Appendix Table D.13: STAR trial: Follow-up Treatment Overview

	CCS (n=461)	DFIS (n=458)	Total (n=919)
Still on trial treatment	8 (1.7%)	34 (7.4%)	42 (4.6%)
Died at the point of discontinuation	12 (2.6%)	31 (6.8%)	43 (4.7%)
Recorded as receiving systemic anti-cancer treatment	315 (68.3%)	249 (54.4%)	564 (61.4%)
Recorded as only receiving anti-cancer surgery	4 (0.9%)	2 (0.4%)	6 (0.7%)
Recorded as only receiving anti-cancer radiotherapy	9 (2.0%)	13 (2.8%)	22 (2.4%)
Recorded as only receiving palliative care	22 (4.8%)	35 (7.6%)	57 (6.2%)
Recorded as only receiving a combination of anti-cancer surgery, anti-cancer radiotherapy and palliative care	15 (3.3%)	7 (1.5%)	22 (2.4%)
Withdrew from data collection prior to information being recorded	7 (1.5%)	7 (1.5%)	14 (1.5%)
Not recorded as receiving any treatment post-trial discontinuation	41 (8.9%)	47 (10.3%)	88 (9.6%)
Died prior to follow-up visit	27 (5.9%)	30 (6.6%)	57 (6.2%)
Data collection ended prior to follow-up visit	1 (0.2%)	3 (0.7%)	4 (0.4%)

Appendix E Chapter 5 Appendix

Appendix Table E.1: Summarised Papers from Updated Literature Review

Title	1	2	3	4	5	6	7	8	9	10
Addition of Carboplatin to Sequential Taxane-Anthracycline Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer: A Phase III Randomized Controlled Trial. [1]	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Surgery Versus Ablation for Hepatocellular Carcinoma: A Randomized Controlled Trial (SURF-RCT Trial) and a Nonrandomized Prospective Observational Trial (SURF-Cohort Trial). [2]	No	Yes	Yes	Yes	No	Yes	No	No	N/A	N/A
Retifanlimab with carboplatin and paclitaxel for locally recurrent or metastatic squamous cell carcinoma of the anal canal (POD1UM-303/InterAACT-2): a global, phase 3 randomised controlled trial. [3]	Yes	Yes	Yes	Yes	No	Yes (Crossover only)	No	Yes	RPSFTM / IPCW for switching only	No
Claudin-18 isoform 2-specific CAR T-cell therapy (satri-cel) versus treatment of physician's choice for previously treated advanced gastric or gastro-oesophageal junction cancer (CT041-ST-01): a randomised, open-label, phase 2 trial. [4]	Yes	Yes	Yes	Yes	No	No	No	Yes	RPSFTM for switching only	No
177Lu-PSMA-617 versus a change of androgen receptor pathway inhibitor therapy for taxane-naive patients with progressive metastatic castration-resistant prostate cancer (PSMAfore): a phase 3, randomised, controlled trial. [5]	Yes	Yes	Yes	Yes	No	Yes	Yes	No	N/A	Yes
Phase 3 Trial of Stereotactic Body Radiotherapy in Localized Prostate Cancer. [6]	No	No	No	No	No	No	No	No	No	N/A

Title	1	2	3	4	5	6	7	8	9	10
Adding 6 months of androgen deprivation therapy to postoperative radiotherapy for prostate cancer: a comparison of short-course versus no androgen deprivation therapy in the RADICALS-HD randomised controlled trial. [7]	Yes	Yes	Yes (Non-Protocol ADT Only)	Yes (Non-Protocol ADT Only)	No	Yes (Non-Protocol ADT Only)	Yes	Yes	Time to non-protocol ADT	No
Liver transplantation plus chemotherapy versus chemotherapy alone in patients with permanently unresectable colorectal liver metastases (TransMet): results from a multicentre, open-label, prospective, randomised controlled trial. [8]	Yes	Yes	Yes (Chemo alone only)	Yes (Chemo alone only)	Yes (Chemo alone arm only)	No	Yes	Yes	PP Exclude	No
Isatuximab plus carfilzomib-dexamethasone versus carfilzomib-dexamethasone in patients with relapsed multiple myeloma (IKEMA): overall survival analysis of a phase 3, randomised, controlled trial. [9]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	RPSFTM, IPCW, Censoring	Yes
Plinabulin plus docetaxel versus docetaxel in patients with non-small-cell lung cancer after disease progression on platinum-based regimen (DUBLIN-3): a phase 3, international, multicentre, single-blind, parallel group, randomised controlled trial. [10]	No	Yes	No	No	No	Yes	Yes	No	N/A	No
Extent of Lymphadenectomy for Surgical Management of Right-Sided Colon Cancer: The Randomized Phase III RELARC Trial. [11]	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Temporary treatment cessation compared with continuation of tyrosine kinase inhibitors for adults with renal cancer: the STAR non-inferiority RCT. [12]	No	Yes	No	No	No	No	Yes	No	N/A	Yes

Title	1	2	3	4	5	6	7	8	9	10
Comparison of platinum combination chemotherapy plus pembrolizumab versus platinum combination chemotherapy plus nivolumab-ipilimumab for treatment-naive advanced non-small-cell lung cancer in Japan (JCOG2007): an open-label, multicentre, randomised, phase 3 trial. [13]	No	Yes	Yes	Yes	Yes	Yes	No	No	N/A	N/A
Site-specific therapy guided by a 90-gene expression assay versus empirical chemotherapy in patients with cancer of unknown primary (Fudan CUP-001): a randomised controlled trial. [14]	Yes	Yes	Yes	Yes	No	Yes	No	No	N/A	N/A
First-line bevacizumab plus chemotherapy in Chinese patients with stage III/IV epithelial ovarian cancer, fallopian tube cancer or primary peritoneal cancer: a phase III randomized controlled trial. [15]	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Prophylactic Radiation Therapy Versus Standard of Care for Patients With High-Risk Asymptomatic Bone Metastases: A Multicenter, Randomized Phase II Clinical Trial. [16]	No	No	No	No	No	No	No	No	N/A	N/A
PD-L1 expression guidance on sintilimab versus pembrolizumab with or without platinum-doublet chemotherapy in untreated patients with advanced non-small cell lung cancer (CTONG1901): A phase 2, randomized, controlled trial. [17]	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Enzalutamide versus Abiraterone Plus Prednisolone for Nonmetastatic Castration-Resistant Prostate Cancer: A Sub-Analysis from the ENABLE Study for PCa. [18]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	N/A	Yes
Mirvetuximab Soravtansine in FRalpha-Positive, Platinum-Resistant Ovarian Cancer. [19]	No	No	No	No	No	No	No	No	N/A	N/A
Pembrolizumab plus chemotherapy versus chemotherapy in untreated advanced pleural mesothelioma in Canada, Italy, and France: a phase 3, open-label, randomised controlled trial. [20]	Yes	Yes	Yes	Yes	No	Yes (Immunotherapy only)	Yes	No	N/A	No

Title	1	2	3	4	5	6	7	8	9	10
First-line nivolumab, paclitaxel, carboplatin, and bevacizumab for advanced non-squamous non-small cell lung cancer: Updated survival analysis of the ONO-4538-52/TASUKI-52 randomized controlled trial. [21]	No	No	No	No	No	No	No	No	N/A	N/A
Phase II randomised, double-blind study of mFOLFIRINOX plus ramucirumab versus mFOLFIRINOX plus placebo in advanced pancreatic cancer patients (HCRN GI14-198). [22]	No	Yes	No	No	No	No	Yes	No	N/A	Yes
Efficacy and Safety of First-Line Cytokines versus Sunitinib and Second-Line Axitinib for Patients with Metastatic Renal Cell Carcinoma (ESCAPE Study): A Phase III, Randomized, Sequential Open-Label Study. [23]	Yes	Yes	Yes	No	Yes (Second line only)	Yes (Second line only)	Yes	No	N/A	No
Impact of valganciclovir therapy on severe IRIS-Kaposi Sarcoma mortality: An open-label, parallel, randomized controlled trial. [24]	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Effect of Prior Local Therapy on Response to First-line Androgen Receptor Axis Targeted Therapy in Metastatic Castrate-resistant Prostate Cancer: A Secondary Analysis of the COU-AA-302 Trial. [25]	No	No	No	No	No	No	No	No	N/A	N/A
Molecular selection of therapy in metastatic colorectal cancer: the FOCUS4 molecularly stratified RCT. [26]	No	Yes	Yes (D only, if return to first line)	Yes (D only, if return to first-line)	No	Yes (D - if return to first line)	No	No	N/A	N/A
Clinical Efficacy of PD-1 Inhibitors Plus Split-Course Radiotherapy in the First-Line Treatment of Advanced Kidney Cancer: A Randomized Controlled Trial. [27]	No	No	No	No	No	No	No	No	N/A	N/A
Hybrid-control arm construction using historical trial data for an early-phase, randomized controlled trial in metastatic colorectal cancer. [28]	No	No	No	No	No	No	No	No	N/A	N/A
Enzalutamide Versus Abiraterone plus Prednisolone Before Chemotherapy for Castration-resistant Prostate Cancer: A Multicenter Randomized Controlled Trial. [29]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	N/A	No

Title	1	2	3	4	5	6	7	8	9	10
Radiotherapy to the prostate for men with metastatic prostate cancer in the UK and Switzerland: Long-term results from the STAMPEDE randomised controlled trial. [30]	Yes	Yes	No	No	No	No	Yes	Yes	Time to second-line treatment	No
Triplet (FOLFOXIRI) Versus Doublet (FOLFOX or FOLFIRI) Regimen as First Line Treatment in Metastatic Colorectal Carcinoma, a Prospective Phase II, Randomized Controlled Trial. [31]	Yes	Yes	Yes	Yes	Yes	Yes	No	No	N/A	N/A
High-dose radiotherapy and risk-adapted androgen deprivation in localised prostate cancer (DART 01/05): 10-year results of a phase 3 randomised, controlled trial. [32]	No	Yes	Yes	Yes	No	Yes	Yes	No	N/A	Yes
Response-adapted omission of radiotherapy and comparison of consolidation chemotherapy in children and adolescents with intermediate-stage and advanced-stage classical Hodgkin lymphoma (EuroNet-PHL-C1): a titration study with an open-label, embedded, multinational, non-inferiority, randomised controlled trial. [33]	No	No	No	No	No	No	No	No	N/A	N/A
Weekly Cisplatin Plus Radiation for Postoperative Head and Neck Cancer (JCOG1008): A Multicenter, Noninferiority, Phase II/III Randomized Controlled Trial. [34]	No	No	No	No	No	No	No	No	N/A	N/A
Deintensified Chemoradiotherapy for Pretreatment Epstein-Barr Virus DNA-Selected Low-Risk Locoregionally Advanced Nasopharyngeal Carcinoma: A Phase II Randomized Noninferiority Trial. [35]	No	Yes	No	No	No	No	No	No	N/A	N/A
Prospective Phase II Open-Label Randomized Controlled Trial to Compare Mandibular Preservation in Upfront Surgery With Neoadjuvant Chemotherapy Followed by Surgery in Operable Oral Cavity Cancer. [36]	No	No	No	No	No	No	No	No	No	N/A
Induction chemotherapy with albumin-bound paclitaxel plus lobaplatin followed by concurrent radiochemotherapy for locally advanced esophageal cancer. [37]	No	No	No	No	No	No	No	No	N/A	N/A

Column guide: 1 = Obvious Two-Stage Application? 2 = Are Subsequent Lines Discussed? 3 = Summarised % in results or appendix 4 =Summarised number OR % in results or appendix? 5=Included a breakdown or summary of the number of subsequent lines (Two, Three, Four) in manuscript or appendix? 6=Included a breakdown of subsequent treatments (Treatment 1, Treatment 2, Treatment 3) in manuscript or appendix? 7=Mentioned in Discussion 8=Did some form of analysis 9=What analysis? 10= Do they think they are an issue for their analysis?

E.1 Appendix E References

1. Gupta S, Nair N, Hawaldar RW, et al. Addition of Carboplatin to Sequential Taxane-Anthracycline Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer: A Phase III Randomized Controlled Trial. *J Clin Oncol* 2025: JCO2501023.
2. Kawaguchi Y, Hasegawa K, Kashiwabara K, et al. Surgery Versus Ablation for Hepatocellular Carcinoma: A Randomized Controlled Trial (SURF-RCT Trial) and a Nonrandomized Prospective Observational Trial (SURF-Cohort Trial). *J Clin Oncol* 2025; 43: 2628-2638. Randomized Controlled Trial Observational Study Multicenter Study Comparative Study.
3. Rao S, Samalin-Scalzi E, Evesque L, et al. Retifanlimab with carboplatin and paclitaxel for locally recurrent or metastatic squamous cell carcinoma of the anal canal (POD1UM-303/InterAACT-2): a global, phase 3 randomised controlled trial. *Lancet* 2025; 405: 2144-2152. Randomized Controlled Trial Clinical Trial, Phase III Multicenter Study.
4. Qi C, Liu C, Peng Z, et al. Claudin-18 isoform 2-specific CAR T-cell therapy (satri-cel) versus treatment of physician's choice for previously treated advanced gastric or gastro-oesophageal junction cancer (CT041-ST-01): a randomised, open-label, phase 2 trial. *Lancet* 2025; 405: 2049-2060. Clinical Trial, Phase II Randomized Controlled Trial Multicenter Study.
5. Morris MJ, Castellano D, Herrmann K, et al. ¹⁷⁷Lu-PSMA-617 versus a change of androgen receptor pathway inhibitor therapy for taxane-naive patients with progressive metastatic castration-resistant prostate cancer (PSMAfore): a phase 3, randomised, controlled trial. *Lancet* 2024; 404: 1227-1239. Randomized Controlled Trial Clinical Trial, Phase III Multicenter Study Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(24)01653-2.
6. van As N, Griffin C, Tree A, et al. Phase 3 Trial of Stereotactic Body Radiotherapy in Localized Prostate Cancer. *N Engl J Med* 2024; 391: 1413-1425. Clinical Trial, Phase III Equivalence Trial Multicenter Study. DOI: https://dx.doi.org/10.1056/NEJMoa2403365.
7. Parker CC, Clarke NW, Cook AD, et al. Adding 6 months of androgen deprivation therapy to postoperative radiotherapy for prostate cancer: a comparison of short-course versus no androgen deprivation therapy in the RADICALS-HD randomised controlled trial. *Lancet* 2024; 403: 2405-2415. Randomized Controlled Trial Multicenter Study Comparative Study Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S0140-6736(24)00548-8.
8. Adam R, Piedvache C, Chiche L, et al. Liver transplantation plus chemotherapy versus chemotherapy alone in patients with permanently unresectable colorectal liver metastases (TransMet): results from a multicentre, open-label, prospective, randomised controlled trial. *Lancet* 2024; 404: 1107-1118. Multicenter Study Randomized Controlled Trial. DOI: https://dx.doi.org/10.1016/S0140-6736(24)01595-2.
9. Yong K, Martin T, Dimopoulos MA, et al. Isatuximab plus carfilzomib-dexamethasone versus carfilzomib-dexamethasone in patients with relapsed multiple myeloma (IKEMA): overall survival analysis of a phase 3, randomised, controlled trial. *The Lancet Haematology* 2024; 11: e741-e750. Randomized Controlled Trial Clinical Trial, Phase III Multicenter Study. DOI: https://dx.doi.org/10.1016/S2352-3026(24)00148-0.
10. Han B, Feinstein T, Shi Y, et al. Plinabulin plus docetaxel versus docetaxel in patients with non-small-cell lung cancer after disease progression on platinum-based regimen (DUBLIN-3): a phase 3, international, multicentre, single-blind, parallel group, randomised controlled trial. *Lancet Respir Med* 2024; 12: 775-786. Randomized Controlled Trial Multicenter Study Clinical Trial, Phase III. DOI: https://dx.doi.org/10.1016/S2213-2600(24)00178-4.

11. Lu J, Xing J, Zang L, et al. Extent of Lymphadenectomy for Surgical Management of Right-Sided Colon Cancer: The Randomized Phase III RELARC Trial. *J Clin Oncol* 2024; 42: 3957-3966. Randomized Controlled Trial Multicenter Study Clinical Trial, Phase III. DOI: https://dx.doi.org/10.1200/JCO.24.00393.
12. Collinson F, Royle K-L, Swain J, et al. Temporary treatment cessation compared with continuation of tyrosine kinase inhibitors for adults with renal cancer: the STAR non-inferiority RCT. 2024; 28: 45. DOI: 10.3310/JWTR4127.
13. Shiraishi Y, Nomura S, Sugawara S, et al. Comparison of platinum combination chemotherapy plus pembrolizumab versus platinum combination chemotherapy plus nivolumab-ipilimumab for treatment-naive advanced non-small-cell lung cancer in Japan (JCOG2007): an open-label, multicentre, randomised, phase 3 trial. *Lancet Respir Med* 2024; 12: 877-887. Randomized Controlled Trial Clinical Trial, Phase III Multicenter Study Comparative Study. DOI: https://dx.doi.org/10.1016/S2213-2600(24)00185-1.
14. Liu X, Zhang X, Jiang S, et al. Site-specific therapy guided by a 90-gene expression assay versus empirical chemotherapy in patients with cancer of unknown primary (Fudan CUP-001): a randomised controlled trial. *Lancet Oncol* 2024; 25: 1092-1102. Randomized Controlled Trial Comparative Study DOI: https://dx.doi.org/10.1016/S1470-2045(24)00313-9.
15. Wu X, Liu J, An R, et al. First-line bevacizumab plus chemotherapy in Chinese patients with stage III/IV epithelial ovarian cancer, fallopian tube cancer or primary peritoneal cancer: a phase III randomized controlled trial. *J* 2024; 35: e99. Clinical Trial, Phase III Randomized Controlled Trial. DOI: https://dx.doi.org/10.3802/jgo.2024.35.e99.
16. Gillespie EF, Yang JC, Mathis NJ, et al. Prophylactic Radiation Therapy Versus Standard of Care for Patients With High-Risk Asymptomatic Bone Metastases: A Multicenter, Randomized Phase II Clinical Trial. *J Clin Oncol* 2024; 42: 38-46. Randomized Controlled Trial Multicenter Study Clinical Trial, Phase II. DOI: https://dx.doi.org/10.1200/JCO.23.00753.
17. Maggie Liu SY, Huang J, Deng JY, et al. PD-L1 expression guidance on sintilimab versus pembrolizumab with or without platinum-doublet chemotherapy in untreated patients with advanced non-small cell lung cancer (CTONG1901): A phase 2, randomized, controlled trial. *Science Bulletin* 2024; 69: 535-543. Randomized Controlled Trial Clinical Trial, Phase II. DOI: https://dx.doi.org/10.1016/j.scib.2023.12.046.
18. Mita K, Izumi K, Goriki A, et al. Enzalutamide versus Abiraterone Plus Prednisolone for Nonmetastatic Castration-Resistant Prostate Cancer: A Sub-Analysis from the ENABLE Study for PCa. *Cancers (Basel)* 2024; 16: 24. DOI: https://dx.doi.org/10.3390/cancers16030508.
19. Moore KN, Angelergues A, Konecny GE, et al. Mirvetuximab Soravtansine in FRalpha-Positive, Platinum-Resistant Ovarian Cancer. *N Engl J Med* 2023; 389: 2162-2174. Clinical Trial, Phase III Comparative Study Randomized Controlled Trial. DOI: https://dx.doi.org/10.1056/NEJMoa2309169.
20. Chu Q, Perrone F, Greillier L, et al. Pembrolizumab plus chemotherapy versus chemotherapy in untreated advanced pleural mesothelioma in Canada, Italy, and France: a phase 3, open-label, randomised controlled trial. *Lancet* 2023; 402: 2295-2306. Randomized Controlled Trial Clinical Trial, Phase III Research Support, Non-U.S. Gov't DOI: https://dx.doi.org/10.1016/S0140-6736(23)01613-6.
21. Kim HR, Sugawara S, Lee JS, et al. First-line nivolumab, paclitaxel, carboplatin, and bevacizumab for advanced non-squamous non-small cell lung cancer: Updated survival analysis of the ONO-4538-52/TASUKI-52 randomized controlled trial. *Cancer Med* 2023; 12: 17061-17067. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1002/cam4.6348.

22. Shaib WL, Manali R, Liu Y, et al. Phase II randomised, double-blind study of mFOLFIRINOX plus ramucirumab versus mFOLFIRINOX plus placebo in advanced pancreatic cancer patients (HCRN GI14-198). *Eur J Cancer* 2023; 189: 112847. Randomized Controlled Trial Multicenter Study Clinical Trial, Phase II Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.ejca.2023.02.030.
23. Kadono Y, Konaka H, Nohara T, et al. Efficacy and Safety of First-Line Cytokines versus Sunitinib and Second-Line Axitinib for Patients with Metastatic Renal Cell Carcinoma (ESCAPE Study): A Phase III, Randomized, Sequential Open-Label Study. *Cancers (Basel)* 2023; 15: 13. DOI: https://dx.doi.org/10.3390/cancers15102745.
24. Volkow P, Chavez Galan L, Ramon-Luing L, et al. Impact of valganciclovir therapy on severe IRIS-Kaposi Sarcoma mortality: An open-label, parallel, randomized controlled trial. *PLoS ONE* 2023; 18: e0280209. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pone.0280209.
25. Roy S, Sun Y, Morgan SC, et al. Effect of Prior Local Therapy on Response to First-line Androgen Receptor Axis Targeted Therapy in Metastatic Castrate-resistant Prostate Cancer: A Secondary Analysis of the COU-AA-302 Trial. *Eur Urol* 2023; 83: 571-579. Randomized Controlled Trial Clinical Trial, Phase III Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/j.eururo.2023.02.017.
26. Brown LC, Fisher D, Adams R, et al. *National Institute for Health and Care Research* 2022; 12: 12. Review. DOI: https://dx.doi.org/10.3310/HTNB6908.
27. Li W, Cao Z, Chang P, et al. Clinical Efficacy of PD-1 Inhibitors Plus Split-Course Radiotherapy in the First-Line Treatment of Advanced Kidney Cancer: A Randomized Controlled Trial. *J* 2022; 2022: 8100323. DOI: https://dx.doi.org/10.1155/2022/8100323.
28. Li C, Ferro A, Mhatre SK, et al. Hybrid-control arm construction using historical trial data for an early-phase, randomized controlled trial in metastatic colorectal cancer. *Commun* 2022; 2: 90. DOI: https://dx.doi.org/10.1038/s43856-022-00155-y.
29. Izumi K, Shima T, Mita K, et al. Enzalutamide Versus Abiraterone plus Prednisolone Before Chemotherapy for Castration-resistant Prostate Cancer: A Multicenter Randomized Controlled Trial. *Eur Urol Open Sci* 2022; 41: 16-23. DOI: https://dx.doi.org/10.1016/j.euros.2022.04.016.
30. Parker CC, James ND, Brawley CD, et al. Radiotherapy to the prostate for men with metastatic prostate cancer in the UK and Switzerland: Long-term results from the STAMPEDE randomised controlled trial. *PLoS Med* 2022; 19: e1003998. Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1371/journal.pmed.1003998.
31. Khalil KA, Musallam HS, Hassan MA and Mahmoud IA. Triplet (FOLFOXIRI) Versus Doublet (FOLFOX or FOLFIRI) Regimen as First Line Treatment in Metastatic Colorectal Carcinoma, a Prospective Phase II, Randomized Controlled Trial. *Asian Pac J Cancer Prev* 2022; 23: 3421-3429. Randomized Controlled Trial Clinical Trial, Phase II. DOI: https://dx.doi.org/10.31557/APJCP.2022.23.10.3421.
32. Zapatero A, Guerrero A, Maldonado X, et al. High-dose radiotherapy and risk-adapted androgen deprivation in localised prostate cancer (DART 01/05): 10-year results of a phase 3 randomised, controlled trial. *Lancet Oncol* 2022; 23: 671-681. Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(22)00190-5.
33. Mauz-Korholz C, Landman-Parker J, Balwierz W, et al. Response-adapted omission of radiotherapy and comparison of consolidation chemotherapy in children and adolescents with intermediate-stage and advanced-stage classical Hodgkin lymphoma (EuroNet-PHL-C1): a titration study with an open-label, embedded, multinational, non-inferiority, randomised controlled trial. *Lancet Oncol* 2022; 23: 125-137. Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1016/S1470-2045(21)00470-8.

34. Kiyota N, Tahara M, Mizusawa J, et al. Weekly Cisplatin Plus Radiation for Postoperative Head and Neck Cancer (JCOG1008): A Multicenter, Noninferiority, Phase II/III Randomized Controlled Trial. *J Clin Oncol* 2022; 40: 1980-1990. Clinical Trial, Phase II Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.21.01293.
35. Li XY, Luo DH, Guo L, et al. Deintensified Chemoradiotherapy for Pretreatment Epstein-Barr Virus DNA-Selected Low-Risk Locoregionally Advanced Nasopharyngeal Carcinoma: A Phase II Randomized Noninferiority Trial. *J Clin Oncol* 2022; 40: 1163-1173. Clinical Trial, Phase II Equivalence Trial Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.21.01467.
36. Chaukar D, Prabash K, Rane P, et al. Prospective Phase II Open-Label Randomized Controlled Trial to Compare Mandibular Preservation in Upfront Surgery With Neoadjuvant Chemotherapy Followed by Surgery in Operable Oral Cavity Cancer. *J Clin Oncol* 2022; 40: 272-281. Clinical Trial, Phase II Comparative Study Randomized Controlled Trial Research Support, Non-U.S. Gov't. DOI: https://dx.doi.org/10.1200/JCO.21.00179.
37. Yan MH, Liu F, Qu BL, et al. Induction chemotherapy with albumin-bound paclitaxel plus lobaplatin followed by concurrent radiochemotherapy for locally advanced esophageal cancer. *World J Gastrointest Oncol* 2021; 13: 1781-1790. Clinical Trial. DOI: https://dx.doi.org/10.4251/wjgo.v13.i11.1781.