Developing methods to improve usefulness of economic Decision Analytical Models: case study in COPD telehealth monitoring

Kinga Sokolowska - Lowrie MSc.

Thesis submitted for a Degree of Doctor of Philosophy

School of Health and Related Research

Faculty of Medicine

University of Sheffield

January 2018

Abstract

Background: In the light of a scarcity of health resources and the growing needs of the population, there is considerable interest in the potential of telehealth technology to assist patients in self-management of chronic conditions, such as Chronic Obstructive Pulmonary Disease (COPD), heart failure (HF), and diabetes. However, despite ongoing support for this technology from the UK government, the uptake of the technology has been slower than anticipated, with the research suggesting the lack of evidence for the cost-effectiveness of this technology is one of the major barriers. Economic modelling is one of techniques that could facilitate deeper understanding of the long-term consequences and financial outcomes of telehealth interventions.

Objective: This thesis documents the process of doctoral research into the methods to enhance the use of decision analytical models in the NHS, using a case study of COPD telehealth. This research is predicated on understanding and challenging assumptions around the methods by which decision models are developed, used, disseminated, and evaluated. The study proposes the 'end-user mode' of model dissemination as an alternative to currently used practices.

Methods: During the model development process, the conceptual modelling was undertaken using the existing conceptual framework. The framework was altered to suit the needs of the research, with 29 qualitative interviews conducted to elicit stakeholders' requirements. A usability evaluation of the model was conducted with end-users in a series of 16 tests, with both qualitative and video data analysed. Finally, when the model was released in Open Access, the stakeholder satisfaction was evaluated, using the end-user satisfaction questionnaire to conduct seven further qualitative interviews.

Findings:

A number of specific requirements for the model were elicited during qualitative interviews and fed back to modellers during model development process. The usability evaluation resulted in several problems being identified and eradicated in consecutive phases of development; and the study led to the development of a decision tool that was well-received by NHS stakeholders. The user satisfaction evaluation revealed high satisfaction with the model.

Conclusions: The findings suggest that the 'end-user mode' approach is viable in the development and the dissemination of a decision model for telehealth. Importantly, several potential areas for future research were identified, including the need to develop methods to improve the uptake and the use of modelling in the NHS, and the development of the concept and instruments for end-user satisfaction in modelling and simulation domain.

Acknowledgements

'I can do all things through Christ who strengthens me' (Philippians 4:13 New International Version). I thank God for providing me with the means and grace to complete this work.

I am indebted to many people who have given the support and guidance throughout this PhD.

I cannot thank enough my supervisors: Professor Simon Dixon, Professor Gail Mountain and Dr Praveen Thokala, whose guidance and knowledge made this PhD possible. Thank you, Simon, for your patience, generosity of time, and sharing your knowledge and expertise with me. Thank you, Gail, for guiding me through nuances of telehealth and qualitative research, and for your patience with my English. Thank you, Praveen, for your help with conceptual modelling.

Thank you All, for giving me the emotional support and on-going encouragement.

I also wish to thank researchers involved in the MALT project, Prof Alan Brennan, for commenting on drafts chapters of this thesis and Dr Lizzie Coates for managing and organising support for the project. Also, Dr Daniel Wolstenholme, for advising on methods for usability evaluation, and Dr Georgina Jones for her support on questionnaire design, and reviewing the qualitative chapter.

I am also grateful for people who made this PhD possible. Dr Claire Gibson and Prof Gail Mountain, whose initiative made this PhD research possible; Prof Suzanne Mason, Dr Katherine Jays-Ward and CLAHRC-SY team for awarding me the PhD studentship.

I would like to thank all my research participants, I am very grateful for their time and commitment to this project.

On the personal note, I would like to thank my husband, Neil Lowrie, for his love, support and understanding throughout the process and my Father-in-law, John Lowrie for proofreading the thesis. I am grateful for my PhD colleagues in ScHARR, in particular, Sabine, Sarah, Milad, Noura, Shraddha, Dua and Thamer for the time we spent together, support and humour.

Finally, I would like to dedicate this thesis to my Father, who fell in a coma in 2014. He was always inspiring me to pursue my dreams and would love to see this work completed.

Dissemination

The elements of this work have been presented on the following conferences:

1. 56th Operational Research Society Annual Conference, London September 2014, (oral presentation)
2. Health Economics Study Group, Glasgow June 2014 (poster presentation)
3. I Futures Conference, Sheffield July 2014 (oral presentation)
4. European OR Conference, Glasgow July 2015 (oral presentation)

Table of contents

List of Figures

List of Tables

List of Pictures

List of Abbreviations

BOR Behavioural Operations Research

CBA Cost Benefit Analysis

CCG   Clinical Commissioning Groups

CEA Cost-effectiveness analysis

CHF Chronic Heart Failure

COPD Chronic Obstructive Pulmonary Disease

CLAHRC-SY Collaboration for Leadership in Applied Health Research and Care for South Yorkshire

CPR Combined Predictive Model

FEV_1 Forced expiratory volume in 1 second

HES Hospital Episode Statistics

HTA Health Technology Assessment

ICT Information and Communications Technology

ICER Incremental Cost Effectiveness Ratio

IS Information Systems

MALT Mainstreaming Assistive Living Technologies

NICE National Institute for Clinical Excellence

ONS Office for National Statistics

OR Operations Research

OBC Outline business case

PCT Primary Care Trusts

PARR Patients at Risk of Re-hospitalisation

PBRA Person-Based Resource Allocation

PSA Probabilistic Sensitivity Analysis

RCT Randomized Controlled Trial

QALY Quality Adjusted Life Years

QIPP Quality, Innovation, Productivity and Prevention programme

QOF Quality and Outcomes Framework

1. Introduction

This thesis is concerned with the question of how to facilitate the use of decision analytic models in telehealth. In order to answer this question several issues had to be resolved, for example, of how to develop an economic decision analytic model for telehealth in the first place, how decision makers would like to use these models, and how to measure model's usefulness and user-satisfaction. Sections 1.1-1.4 give an overview of telehealth and explain barriers to its mainstreaming in the light of recent policy and the literature. Sections 1.5-1.7 outline major types of economic evaluations and weaknesses of these frameworks when they are used for evaluations of telehealth. Following this, Section 1.8 explains the rationale of the study and its aims and objectives. The chapter concludes with a discussion of the potential contribution of this work to the broad field of Operations Research and Behavioural Operations Research.

1.1 What is telehealth?

The use of Information and Communications Technology (ICT) in healthcare has expanded during the last decade (Haux, 2006). Information technology is now being used in a growing number of health disciplines, for example, to exchange medical data between either two medical professionals or a medical professional and a patient who are not in the same location.

The prefix 'tele'- is derived from Greek and means 'distant'. Telemedicine, telehealth, telecare and e-health encompass a wide variety of technologies delivered using ICT but the interpretation of what these terms mean varies widely. In order to choose the correct terminology for the study, I followed publications listed in Cochrane reviews, and also two major journals in the field: *Journal of Telemedicine and Telecare*, and *Journal of Telemedicine and e-health*. Chronologically the first to appear was telemedicine, defined as:

The delivery of healthcare services, where distance is a critical factor, by all healthcare professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of healthcare providers, all in the interests of advancing health of individuals and their communities. (WHO, 1998, p. 10)

Telemedicine encompasses technology that is used in clinical settings, to exchange information between healthcare professionals who are in different locations. This includes transmission of medical images and scans (teleradiology, telepathology, teledermatology), video consultations (telecardiology, telepsychiatry) and many others. Telemedicine is especially valuable when distance plays a role, for example, in the provision of healthcare services to rural or excluded populations such as prisons (Inglis, 2011).

Telemonitoring emerged as an expansion of telemedicine into domestic settings. It enables the delivery of health services as well as patient education at home, but current systems vary in the degree to which they merely monitor and educate patients. The term 'telehealth' was coined to highlight the health-related aspect of the technology, as, unlike telemedicine, it is used more for health prevention than curative purposes. However, the two are very often used interchangeably.

In a standard telehealth system, a patient is equipped with a device for collecting data on disease symptoms and peripherals for measuring vital signs such as spirometers, weighing scales, or glucometers (Adeogun *et al*., 2011a). Medical data from these devices is remotely transferred to monitoring centres for the attention of medical professionals. Any change of patients' parameters outside its normal pattern can be flagged up for the immediate attention of health professionals, who either call or visit the patient at home. Some types of telehealth systems provide explicit

educational services in addition to monitoring. Their aim is to improve self- management by educating patients on symptoms of the disease and stimulating them to adhere to medical instructions. Usually, patients receive electronic messages, quizzes and questionnaires to assess their health status, medication compliance and lifestyle choices. Deviations from activity perceived as normal can be spotted and acted upon quickly. The type of telehealth used in this thesis is discussed in Chapter 2.

Certain types of technology are used in social care, to support the elderly or disabled people to live independently. Examples of these include fall sensors, devices for home security, but also systems providing instructions for self-care, and psychological or social support. Usually, there are referred to as '*telecar*e', '*telehealthcare*', or 'a*ssisted living technologies'*.

To complicate things further, around the year 2000 researchers started to use the term e-health as encompassing all types of technology mentioned above and to label the services delivered using the Internet. Some debated that the term should not be used in academia, as it was first used more like business and marketing term, others that it should enter the research domain and be expanded (Della Mea, 2001; Eysenbach, 2001). In Cochrane reviews, the term is usually used for technologies such as electronic health records, health information systems, and other types of clinical electronic communication systems.

The interest of this study is in telehealth and its sub-function telemonitoring. Other aspects of technology might be mentioned in this thesis, especially in reviews or accounts of literature, but only because they are part of the same domain. Figure 1 below illustrates my understanding of terms used to label technology in this field.

Figure 1: My interpretation of major telehealth related terms
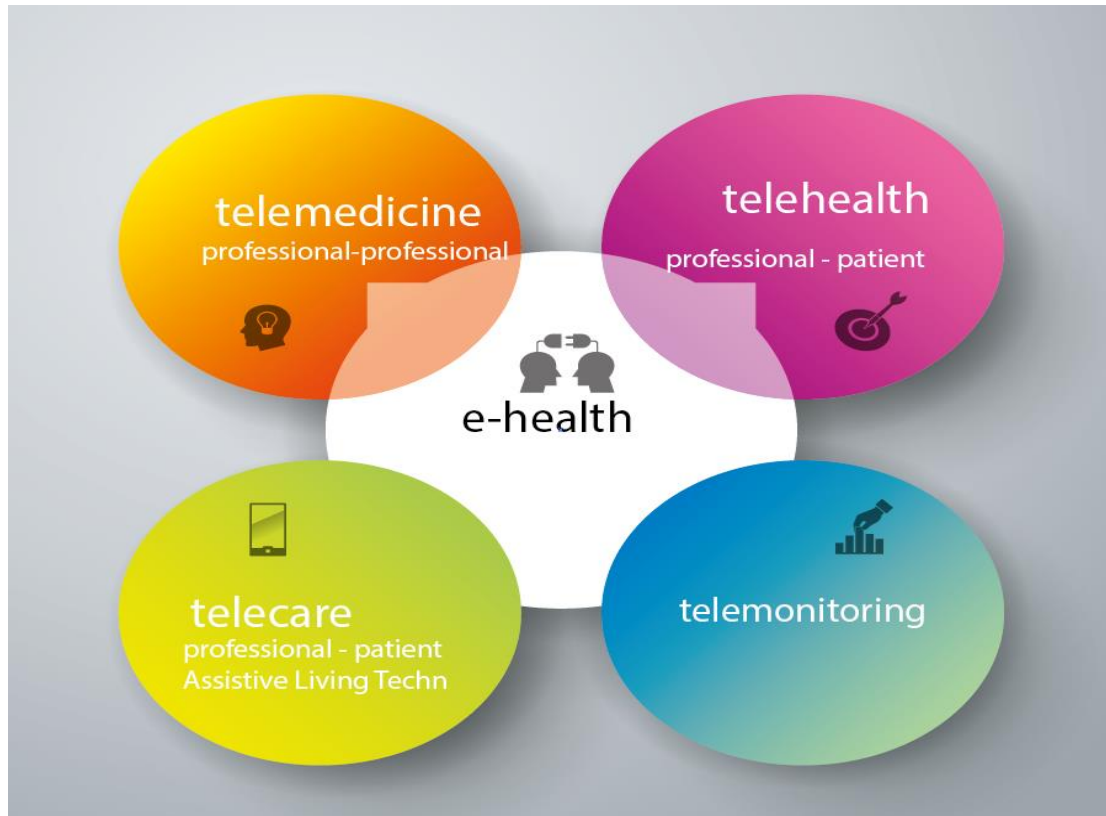
1.2 Benefits of telehealth in COPD

The interest of this study is in telehealth for Chronic Obstructive Pulmonary Disease, an incurable and irreversible long term condition (LTC) which is the fourth largest cause of morbidity and mortality in the world (Pauwels, 2014). The condition results in the reduction in airflow capacity that leads to serious pulmonary problems in advanced stages. Normal functioning of people with COPD is disturbed and their quality of life decreases as the disease worsens over time. Treatment strategies mainly aim at improving functionality and quality of life by preventing acute exacerbations, which are associated with significant risk of mortality (Mannino & Buist, 2007). Currently used clinical pathways for COPD are discussed in Chapter 2.

COPD represents a large and increasing financial burden on both the patients and healthcare provider. It is the second most common cause of emergency hospital admissions in the UK (Audit & Inspection, 2006), with a direct cost of £500million a year (Booker & Bent, 2004). A growth in COPD prevalence in the UK is expected in coming years (Simpson, 2010). It is thought by many that only by changing the way healthcare services are provided, focusing on prevention and supporting self-management, self-care and delivering care at home; can this increasing demand for healthcare be met (Brettle & Smith, 2013; Clark & Goodwin, 2010). Telehealth is perceived as a way to achieve this change.

There is ample evidence that telehealth can be a safe and an efficient mean of providing care to COPD patients (Bartoli *et al*., 2009; McKinstry, 2009; Trappenburg, 2008). When using telemonitoring, patients are under the constant supervision of medical staff, and any worsening of the symptoms of the disease could be quickly spotted and acted upon. This way telehealth significantly reduces the number and severity of exacerbations and even prevents their occurrence (Calvo *et al*., 2014). This leads to significant reductions in the number of A&E visits and hospitalisations (Bartoli *et al*., 2009; Brettle & Smith, 2013; Jódar-Sánchez, 2013). By delivering patient education and supporting lifestyle change, it improves self-management and adherence to medication (Dansky, 2008). Telehealth also improves access to healthcare services especially in remote or rural settings (Inglis, 2011).

Patients are largely satisfied with the use of telehealth (Finkelstein *et al*., 2004; Kruse *et al*., 2017; Polisena *et al*., 2009). Being under the constant supervision of medical staff gives them reassurance and reduces anxiety about symptoms of the disease, as an emergency can be quickly spotted. It also means that they do not have to travel to be checked by their GP as often. Mc Lean (2012) have conducted a meta-analysis of telehealth randomized controlled trials (RCTs) in COPD and reported that the majority of patients were satisfied with the technology, as long as

they could have face-to-face contact with a clinician on request. Finding a right balance between relying on technology yet maintaining some face-to-face contact, was found important, as some patients felt isolated or simply not able to adjust to the new mode of service delivery. It is understandable that within the patient population there will be individuals, who due to their age, cognitive disability and unfamiliarity with technology will not be able to use telemonitoring, especially patients with cognitive impairments or dementia.

Healthcare staff can benefit from telehealth in several ways. Technology can reduce unnecessary patients' visits, saving their time and resources. It helps to manage the caseload more effectively, as up-to-date data is available on patients' condition every day. However, this can only be achieved if personnel accept technology and overcome initial anxieties to use it. Nurses have to learn how to use telehealth, and how to interpret the results. Anecdotal evidence shows that it can be challenging for staff to comply with the script and not to visit a patient immediately after an alarm is detected.

There are ongoing difficulties in demonstrating evidence of benefits from telehealth to the wider healthcare system. One study showed the reduction of resource use for primary and community care, especially the reduction of GP and nurse home visits (Wakefield, 2009), but the number of good quality studies remains sparse.

1.3 Barriers to telehealth implementation

Although the technology is a promising solution for the challenges the healthcare system is facing, currently in the UK, more than 75% of telehealth initiatives fail during the operational phase (Clark & Goodwin, 2010), and many do not survive the research phase. Large-scale implementations are also rare. The unique example in the world of a successful large-scale telehealth system is Veterans Health Administration Care Coordination/ Telehealth Programme

in the United States, which between 2003 and 2007 supported over 30 000 veteran patients with chronic conditions (Cruickshank, 2012; Darkins *et al*., 2008). There were various facilitators of VA success; however, it seems that the programme managed to overcome barriers with which UK programmes are still struggling: the integration of telehealth into existing hospital and community care services, and ensuring sustained funding for this type of services (Cruickshank, 2012; Darkins *et al*., 2008).

The difficulties involved in mainstreaming telehealth can be attributed to inter-related behavioural, technical, financial and organisational issues (Broens *et al*., 2007; May *et al*., 2003; Tanriverdi & Iacono, 1999). As with every technology, there is a need to gain users' acceptance, which is greatly impacted by issues such as lack of user-friendliness of the devices, or issues with their wear. Currently, there are limited options to personalise devices to meet individuals' needs, and not all of them cater for the varying abilities of individual users. For example, patients find it difficult to use weighing scales if they have balance problems. Patients expect devices to be visually pleasant and wireless for easier use and manoeuvres (Horton, 2008). To enable data transmission to monitoring centres, devices have to be connected to the internet. One of the immediate problems for some of the patients is the lack of such connection at home, which has to be taken into account by telehealth providers before enrolling them to telehealth programmes.

From the perspective of staff, the interoperability of telehealth systems plays an important role (Giordano *et al*., 2011). If not compatible with the electronic health record system used by the NHS, telehealth creates new tasks for personnel, as they have to copy manually telehealth data into patients' records for it to be available to other clinicians. Also, the equipment breaking down and other technical issues are among the most often reported reasons why patients and staff felt discouraged to use it. Studies show that staff acceptance is influenced by concerns about the lack of face-to-face contact with patients, data safety, user-friendliness of devices and other technical

issues (Brewster *et al*., 2014; Taylor *et al*., 2015). Above all, previous experiences with telehealth have an impact on attitudes towards technology and dominate the decision whether to continue to use it (Brewster *et al*., 2014).

An existing barrier is also the lack of a nationwide scheme that funds telehealth services. Clinical Commissioning Groups (CCG) who procure telehealth have to use their existing funds to roll out telehealth services in their local areas. In many cases, telehealth is implemented as part of research programmes, and when they are completed, NHS organisations fail to secure recurrent funding. Another issue is also the lack of economic and financial information for service planners and Commissioners, that would allow to determine whether telehealth will be cost-effective within any particular service (The MALT Study Consortium, 2014).

Barriers at the organisational level include efforts to restructure existing practices to integrate telehealth into care pathways and managing change within NHS organisations. Telehealth does not always fit with traditional working protocols, and to be perceived as a vital part of health delivery, it has to be widely promoted. This requires the engagement of all organisations involved in the provision of telehealth and championing by enthusiastic individuals (Joseph *et al*., 2011).

1.4 National policy

The Government has long been an advocate of technology-enabled care and perceived it as a solution to a growing demand for healthcare. Tracing back the national initiatives, the first paper that set out the plans to use technology was '*The New NHS. Modern, Dependable*' White Paper (1997). The Paper expressed the wish for greater investment in technology and building health services' capacity around peoples' needs. More importantly, it set out the plans, that some of the conditions might be dealt with at home or in the community. The resulting implications of this paper were that the 24-hr telephone advice line (NHS Direct) was established and GP surgeries

have been connected by NHSnet to enable on-line booking for outpatients' appointments. These plans were backed with £1.5 billion investment.

The *'Information for Health'* strategy (1998) then followed. Its main aim was to support the drive for efficiency and quality in the NHS by the use of technology. The strategy set up the plans for integrated patient care and was backed with £1billion for the duration of the strategy. One of the key objectives of the strategy was to '*eliminate unnecessary travel and delay for patients by providing remote on-line access such as telemedicine'(Department of Health., 1998, p. 19).* That is how the role of telemedicine and telecare has been presented as the means for modernising the NHS. At the time, the Government set out the priority to develop a framework to guide and support the development of these technologies.

In March 1999, the *'Modernising Government'* (1999) White Paper was published, that put in place a number of initiatives, and directly impacted the launch of an ICT research and development programme to support the multidisciplinary health services research to evaluate the use of technology and its supporting evidence base. The budget for this strategy was £2.5 million for 3 to 4 years.

In 2001, the *'Building the Information Core: Implementing the NHS Plan'*(2001) document was published, which gave priority to self- care and the resources needed to treat people at home. The Plan was backed by the Government £700 million investment in IT. The funding was used to establish technology for carrying out tele-consultations between GPs and hospital consultants, equipping ambulance crews with similar technology to ensure that patients were under specialists' care while being conveyed to hospital, and developing electronic health records that enable patients to book appointments, see their tests results and connect with medical professionals.

In 2002, the Government published *'Delivering 21st Century Support for the NHS'* (2002) which included the plans for telecare to be available for all patients who require it by December 2010. It was from that point onwards the Government changed focus from 'telemedicine' to 'telecare'.

Following on that, the Government had launched a Preventative Technology Grant, with the aim of increasing the number of people on telecare, especially those with long-term conditions. The Grant provided local authorities with £80 million for the development of telecare programmes (Department of Health 2006). Also with the telecare in mind, in 2006 the second edition of ICT research was launched.

'Building the National Care Service Paper' (2010) further presented the Governments plans to use technology in social care, with the emphasis on supporting prevention and independent living. It stated that telecare technologies, such as fall sensors to prevent the elderly from falling, should be made available to all. The document also included plans to integrate health and social care services, to form National Care Services.

NHS Next Stage Review promoted telehealth as a 'core' preventative service for long term conditions, (2009).The practical implications of this paper were establishing the *Quality, Innovation, Productivity, Prevention* (QIPP) Framework, which supported staff with tools and techniques to work for quality improvements and innovation. Telehealth and telecare still play an important role in QIPP, as a mean of promoting self-care and world-class commissioning initiatives.

In a parallel development, in 2007, the Technology Strategy Board launched the Assisted Living Innovation Programme (ALIP) to support various telehealth projects, with a budget of £47.1 million. The main focus of the programme was to establish economic and business models for

telehealth and to test and develop product and services (Technology Strategy Board., 2011). As part of ALIP, Delivering Assisted Living Lifestyles At Scale (DALLAS) programme, was launched in 2011 to equip several communities across the UK with telecare (Worsley, 2013).

As a way to provide much needed evidence to these various initiatives, the Government launched the Whole System Demonstrator Trial. With over 6000 participants taking part, it was regarded as the largest telehealth clinical trial ever conducted. The results suggested impressive reductions in hospital admissions, reduction in the use of resources and mortality (Steventon *et al*., 2012), but delivered inconclusive results on cost-effectiveness (Henderson *et al*., 2013). The implications of these findings had a profound impact on the expansion of telehealth, as they were less promising than the Department of Health expectations (Department of Health, 2011). NHS Commissioning Managers felt discouraged from pursuing telehealth initiatives, as there are expectations that their decisions are evidence-based (Clarke *et al*., 2013).

In 2012 Department of Health had launched a 3 Million Lives programme that aimed at increasing the number of people on telehealth and telecare to three million by the year 2017 (DoH 2012), but despite having the financial backup from industry, the programme has been suspended in 2014 by the NHS England's Technology Enabled Care Services Directive (DoH 2014).

To sum up, over the period of 1997-2013, UK government delivered a number of policies to support the uptake of telehealth technology. During that time, funding in excess of £175M was made available to social and health care bodies to trial various types of home monitoring technology. The funding was not ring-fenced which resulted in discrepancies between the numbers of undertaken projects (Barlow *et al*., 2012). The Government has then diverted from providing funding into supporting research projects to generate evidence on clinical and cost-effectiveness of technology, as this was seen as having a slowing impact on the uptake of

technology. However, as the Whole System Demonstrator Project delivered mixed results; the need for evidence has become even more appealing.

The decision whether or not to invest in telehealth is now the Commissioners' responsibility. However, to procure technology, they have to ensure funding from their existing budgets that require them to assess whether the investment in technology will bring expected benefits for their local populations. In the light of inconclusive evidence on the cost-effectiveness of telehealth, this gap in knowledge becomes an important issue. To complicate things further, because of the organisational and financial split between social care and healthcare services in England, most of the regions' commissions run telecare services, but not all of them telehealth (Clark & Goodwin, 2010).

1.5 Economic evaluation and its use in decision-making

With ever-growing demand for healthcare but limited resources, there is a need for making informed decisions about resource allocation. Economic evaluation compares the costs and the consequences of health technologies and it is now a well-established method within the UK and worldwide (Briggs *et al*., 2006).

All economic evaluations assess the cost of health interventions but differ in the methods used to measure and value their consequences. Cost-consequence analysis assesses cost and effects of interventions without isolating single or aggregate consequences into one measure. Cost-minimisation analysis (CMA) assumes that effects of interventions are equal, therefore only compare costs. It allows identifying the least costly among interventions, using simple monetary units and is usually undertaken mainly for the purposes of technical allocation. Cost-effectiveness analysis (CEA), which is among the most often used frameworks, estimates effects in natural units, such as clinical cases detected or life-years saved. It demonstrates the difference in costs of

interventions together with their difference in effectiveness. It is used for technical efficiency purposes, allowing identifying suitable interventions within given budget or disease types. Cost-utility analysis (CUA) which is a variant of CEA estimates effects using preference-based measures of health such as disability-adjusted-life years (DALY) or quality-adjusted life years (QALYs). Cost-benefit analysis (CBA) estimates consequences of interventions in monetary units. It indicates which interventions will gain more benefit. All of the above methods are discussed in detail in (Drummond *et al*., 2015).

There are two approaches to undertaking economic evaluation. The first uses primary data collected alongside clinical trials and observational studies to evaluate the effectiveness of treatment strategies within the trial. The second uses existing secondary data to build mathematical models to analyse costs and benefits of these treatment strategies under different assumptions. Decision analytical modelling is discussed in Section 1.7.

Whilst economic evaluation is undoubtedly a valuable support tool, little is known about the precise role it plays for different decision makers. At the national level, the National Institute for Clinical Excellence (NICE) uses economic evaluations to inform decisions about allocation of public funds to health technologies and interventions. In the appraisal process, NICE follows the objective of health maximisation under a limited budget and within England and Wales, technologies within the threshold of £20 000 - £30 000 per QALY gained are generally accepted (Buxton, 2006).

The research on the use of economic evaluations on the local-level decision-making suggest that there are several barriers that prevent its widespread use (Eddama & Coast, 2008; Williams *et al*., 2008), which is further discussed in Chapter 4.

1.6 Weaknesses of trial-based economic evaluation of telehealth

Systematic reviews of telehealth found that the majority of the economic evaluations of telehealth were conducted alongside RCTs (Ekeland *et al*., 2010; Mistry, 2012), but some argue that RCTs are inappropriate for evaluations of this technology for several reasons (Ohinmaa *et al*., 2001; Williams *et al*., 2003).

Firstly, telehealth is perceived as a '*complex intervention*'. As such it may include multiple components, involve several groups or organisations, deliver a number and variability of outcomes, and allow flexibility and tailoring of the intervention based on learning by feedback loops (Medical Research Council., 2008). Importantly, complex interventions rely on participants to change their behaviour. That means that randomisation in such cases is a highly debatable issue, as it might have a confounding effect on the results of the study (Davalos *et al*., 2009; Reardon, 2005). Randomization is also not feasible as very often participants make a conscious decision about the uptake of technology.

Another common limitation of telehealth RCTs is their sample size and duration, which makes it challenging to detect differences in health outcomes that take time to emerge. Immediately observable benefits are improvement of care, better quality of care, faster diagnosis or transport related cost savings, which are difficult to incorporate into a standard evaluation. Many trials use surrogate measures such as hospitalisations avoided or exacerbations avoided that are not generalizable and complicate comparisons between studies. Other benefits of telehealth are disparate, applicable to broad range of users and spread around whole healthcare and social care system.

Secondly, as previously mentioned, telehealth ranges from basic systems that allow telemonitoring of vital signs, to services tailored to improve self-management, awareness of the

symptoms of the disease and medication compliance. The functionality of the devices differs as well, with some offering options to communicate with the health professionals via teleconferencing, receiving daily quizzes or questionnaires. Currently on the market there are over 12 bespoke telehealth devices offered as part of managed systems (Adeogun *et al*., 2011b), but there has been little recognition which system works best for what condition. Simply, one system given to different populations or deployed differently might create completely different benefits (Drummond *et al*., 2005; Taylor & Iglesias, 2009). Above all, the efficacy of the device is dependent upon methods of use. Some patients might be struggling to make use of it, due to technological or other reasons, and so create what is called *'false alarms'*, when the device detects a problem with vital signs readings that are not emergencies. Anecdotal evidence suggests that there are a high number of these types of alarms when patients are learning how to use the system, which have been just installed in their homes. Also, the range of readings which should trigger the alarm has to be calibrated for each new patient on the system.

When effects of telehealth interventions on healthcare system are considered, it needs emphasising that telehealth efficacy does not rely solely on technology, it requires significant changes in work processes and reconfiguration of existing practices. The effectiveness of telehealth services could be altered by changing the number of service specifications. For example, increasing the number of people using technology at a particular point of time, re-arranging turnover times, so that the equipment could be disinfected and installed in another patients' home faster are commonly used rules of economy of scale. The effectiveness of the service might be also altered by re-organisation, for example by employing less staff or shifting their roles. To achieve that, ideally, technology needs to be integrated into care pathways. Above all, organizational changes require time and understanding of key issues that can be a very complex task.

There is also a significant question of how to measure benefits of telehealth. Quality Adjusted Life Years (QALY) which is recommended by NICE as a unit of effectiveness for health technologies, has the advantage of capturing gains from reduced morbidity and reduced mortality, and integrating them into a single measure (Drummond *et al*., 2015). The concept of the QALY is also useful when changes in quality of life are being traded with survival but each QALY gained is assumed to have the same weight regardless of the characteristics of the individuals receiving the health benefit.

When adopting QALY in evaluations of telehealth, a few problems arise. The effects telehealth has on individuals might depend on the severity of their condition, their age, socio-economic status, or other factors, and QALY might not be sensitive enough to capture these (Whitehead & Ali, 2010). Bergmo (2014) has conducted a systematic review of economic evaluations of telehealth and concluded that disparities between methods used to derive QALYs might lead to bias when results of telehealth studies are compared.

When costs of telehealth are considered, there are also a few problematic issues. Telehealth is ultimately a network of several organisations – it is challenging to attribute costs to certain bodies. Also, the unit price of devices is constantly changing as it is greatly impacted by the market. Technology can become obsolete after 4-5 years and new products enter the market each year, which have an impact on overall prices. Rapid technological development means that the task of conducting telehealth evaluation can be compared to '*hitting a moving target'* (Gagnon 2004).

To sum up, although RCTs are perceived as the '*gold standard*' in case of evaluations of pharmaceuticals, they might not be the best solution in case of complex interventions due the specific character of these interventions. The coexistence of multiple components and the fact that

the effects of the intervention depend on the behavioural or organisational change of those involved make the search for an alternative approach necessary.

Telehealth researchers have long expressed concerns about the rigour and consistency of methods used in the field, raising questions about the quality of generated evidence (Ekeland *et al*., 2010; Mistry, 2012). There were several attempts in the literature to improve the methodology of telehealth evaluations. McIntosh and Cairns (1997) proposed a cost-consequences framework linked to a set of evaluative questions. Sisk and Sanders (1998) suggested cost-benefit approach with a focus on whether the health benefits of technology are likely to outweigh the investments needed to implement it. Le Goff-Pronost (2010) attempted to address issues of telehealth evaluations by adding a set of calculations (break-even point, discount benefits, social benefit, probable social benefit and economic result) to a standard cost-effectiveness framework. Ultimately the consensus on the best analytical framework for the evaluations of telehealth has never been reached.

1.7 Challenges of decision modelling in telehealth

Decision modelling is a technique that represents real world using mathematical equations in order to conduct experiments and mimic the behaviour of real systems (Brennan & Akehurst, 2000). Modelling is widely used in different disciplines, and in healthcare, it is a major method to inform decisions about resource allocation.

Discrete event simulation (DES) and Markov modelling are the most often used modelling approaches although alternative techniques such as System Dynamics and agent-based modelling are also prominent (Mielczarek & Uziałko-Mydlikowska, 2010). An excellent typology of decision modelling techniques has been published by Brennan *et al* (2006), and characterises

models according to whether they are concerned with a population as a whole (cohort) or the individuals.

For the introductory purposes, I provide short description of Markov and DES models and advantages and disadvantages of both approaches. Markov models are used to represent stochastic processes that evolve over time, thus they are especially useful in modelling the progression of the diseases (Briggs & Sculpher, 1998). The principle of Markov modelling is that the disease in question is divided into mutually exclusive states, and transition probabilities attached to each state to represent how a cohort of patient progresses in predefined regular time intervals, e.g. monthly. The model is also characterised as '*memoryless*' that means that the health states the cohort occupies in the future depend only upon the present state, not on the sequence of events that preceded it (Drummond *et al*., 2015). When the model is run over a large number of cycles, the estimates of long-term cost and effects can be obtained.

DES models are also built for stochastic processes, but used to represent systems where set of activities occur in specified time intervals and individuals have to queue in order to progress from one activity to another (Karnon *et al*., 2012). These can be as complex as the onset of a disease of interest, the progression to a new health state or resource use or clinical decision. A good example of such is a schedule of hospital bed allocation, where a new patient has to wait until the bed is free in order to be admitted. The important difference between DES and Markov is that the individual's characteristics determine what happens in the model and how quickly he progress. DES allows detailed comparison of scenarios, making predictions and optimisation thus has become one of the major methods used in healthcare modelling for example for scheduling patients flows, estimating allocation of resources such as hospital beds, or staff (Fone *et al*., 2003; Jun *et al*., 1999).

The conceptual model in DES is the activity diagram, that '*shows the logic of the flow of entities through queues and activities'* (Brailsford and Hilton (2001), page 9) depicted as a series of circles and rectangles. Conversely, in Markov models, the conceptual model consists of a list of the possible states in the system, the possible transition paths between those states, and the rate parameters of those transitions. These are usually depicted as circles connected by arrows. This thesis focuses predominantly on Markov and DES models when referring to '*modelling'* or '*models'*.

Main features of both types of modelling approaches are compared based on (Standfield *et al*., 2014) in Table 1 below.

Table 1. The comparison of Markov and DES approaches

	Markov process	**Discrete event simulation**
What is being modelled?	Cohort	Individuals
Memory	Markovian assumption	Can account for patients history of future events
Interaction between individuals	No	Yes, can simulate interactions between individuals and individuals and environment, e.g. queuing

	Markov process	**Discrete event simulation**
Risk	Does not account for competing risks as transition probabilities are derived for each of the mutually exclusive health states and are exhaustive.	Well suited for modelling competing risks
Time and occurrence of events	Model progresses according to fixed cycle lengths. The cohort is in various mutually exclusive states at any one time therefore does only experience one event at the time.	Model progresses according to the timing of events. Events can occur simultaneously. Time between cycles can be of any duration.
Development requirements	Less demanding of data, shorter development and validation time.	More demanding of data, requires more time to be developed and validated. Computational burden – requires simulation software.

Considering the differences listed in the Table 1, DES models are better suited to model processes where it is important to account for individual history, and various, often competing risk factors. In order to model similar process using Markov, enormous number of health states would have to be created (Standfield *et al*., 2014). Besides, DES has the ability to model complex systems, in the way it allows to account for interactions between patients and patients and environment (Brennan *et al*., 2006). This feature is often used to analyse supply and demand or queuing. For example, it has been successfully applied in the studies to improve operation of emergency services (Brailsford *et al*., 2004; Connelly & Bair, 2004; Duguay & Chetouane, 2007).

The most often cited disadvantage of DES includes the fact that DES requires extensive data, and its developmental time is significantly longer (Caro *et al*., 2010; Karnon, 2003; Karnon *et al*., 2012). Karnon (2003) conducted a comparative modelling study using both Markov and DES approaches. He proved that the accuracy of modelling was not affected by the method chosen as both methods predicted similar cost-effectiveness ratios. Markov however, was more favourable in terms of computer processing time needed – Markov model running time was one hour in comparison to three days needed for DES. Also, there is a consensus that in overall, Markov models are more transparent, as there is a risk of making the model unnecessary complex, when using DES approach (Karnon, 2003). DES usually requires advanced simulation packages, such as Simul8, as opposed to Markov that can be successfully built and run even using Excel spreadsheets (Drummond *et al*., 2015).

Having considered the advantages and disadvantages of both Markov and DES, the practical recommendations given by Briggs and colleagues (2006) on the choice of the approach is to consider what level of approximation is needed for a given problem. Authors recommend starting by considering how well Markov will capture the critical elements required for the decision problem. The options to overcome some of the shortcomings of Markov include using additional states, time-dependent probabilities and building a semi-Markov process. These have been well explained by Briggs *et al*. (2006).

Among the benefits of modelling, is the fact that it uses the evidence from many sources, such as RCTs, observational studies, meta-analyses or expert opinion to explore the cost and the effects of particular interventions, especially when relevant clinical trials have not been conducted (Buxton *et al*., 1997). It delivers much clearer information about costs and benefits of health interventions than any single trial would ever be able to. In such a way, modelling allows to imitate a trial in a specific context, extrapolate findings to other patients or countries (Buxton *et*

al., 1997). Moreover, it allows extrapolating surrogate measures such as the number of COPD exacerbations avoided, into final endpoint measures such as QALY. The latter is more useful for decision makers as allows comparing different health interventions.

Developing a decision model for telehealth is associated with several challenges. In a standard Health Technology Assessment (HTA), decision analytical models are developed to represent mathematically clinical pathways for alternative treatment strategies and to quantify the probability of patients following it (Briggs *et al.*, 2006). Defining the pathway has implications for the choice of the model structure. However, a standard pathway for telehealth has not been established yet, and substantial variations exist in the way telehealth services are provided.

Data poses another challenge. In the process of developing the model it is essential to synthesise data on several aspects of the service, for example, the resource use, costs and patients' quality of life. Many telehealth trials conducted up to date did not report on these costs and benefits in a sufficient way that would enable reusing it for the purposes of modelling. Moreover, the unit costs are constantly changing due to the advances of technology. This means that such model has to be flexible to accommodate frequent change of various service parameters, which in standard HTA would mean rebuilding the model every time there is a need to do so. This in turn is time-consuming and costly.

Additionally, the benefits of telehealth include a range of non-health benefits, such as improvement of quality and access to care, faster response in case of emergencies, savings of time and cost of travel for medical staff, reassurance for patients; which very often outweigh the benefits of reduced mortality and morbidity. There are ongoing discussions on the best approach to quantify the benefits of telehealth (Bergmo, 2014; Polisena *et al.*, 2009; Snoswell *et al.*, 2016), with the general consensus that QALY which is usually used in evaluations of health technologies

might not be appropriate. QALY might not be sensitive enough to detect small changes in quality indicators (Bergmo, 2014), and the health states utilities to estimate QALY are the hardest to obtain.

Another issue specific to telehealth, is the fact that the analysis of costs and resources need to be much more detailed to support operational decisions about the implementation and service delivery. In tandem with the need for business cases to support investments in the NHS, detailed financial modelling is required alongside the traditional economic modelling that is centered on an ICER for a single cohort of patients.

Although the use of modelling for telehealth evaluation has been greatly recommended, (Bergmo, 2009), the recent (December 2016), search of the literature identified limited number of published studies that applied the method (Bergmo, 2015; Dixon *et al*., 2016; Jean *et al*., 2015; Thokala *et al*., 2013).

1.8 Rationale and justification of the present study

 Academics and practitioners have considered simulation and decision models to be a valuable decision support tools for healthcare interventions (Brailsford, 2007; Jun *et al*., 1999; Mielczarek & Uziałko-Mydlikowska, 2010); however, despite the growing interest within Operational Research to examine behavioural issues of model use and the advent of Behavioural OR, little is known about how exactly models are used in such processes. Two issues relating to the use of models are worth considering at this point; the mode of model development and how this interplays with the type of decision process the model is applied to.

The literature distinguishes two modes of model development and use: the expert mode and the facilitated modelling (Franco & Rouwette, 2011). The expert mode has been executed in majority

of OR and health economics interventions and it is based on the premise that the modellers develop the tool to answer clients query, and present the findings usually in a form of written reports. These clients have therefore little or no contact with models and rely on analyses conducted entirely by the experts (Franco & Montibeller, 2010; Robinson, 2011; Robinson *et al*., 2004). Franco (2010) points to the fact that the expert mode is suitable in situations when clients expect the recommendation of an optimal solution to a problem rather than when they need to consider a range of possible solutions. Clients' involvement is therefore not necessary, as the modeller is ultimately a person who makes all the decisions on every step of the model development process, for example on how to frame the problem, how to construct the conceptual model and how to analyse the results.

Alternatively, facilitated modelling that is sometimes referred to as the '*group model building*' (Rouwette *et al*., 2011), '*the decision conferencing*' (Phillips 2007), or '*the journey making'* (Ackerman 2001), gives the clients the opportunity to be involved in the development process. It is usually delivered in a form of workshops, during which the group works to define and to structure the decision problem (Franco & Montibeller, 2010). Once the structure of the model is agreed upon the modellers populate it with the data, and the group meets again to find a solution to their decision problem. They test various what-if scenarios and conclude with an agreement on the solution.

Another issue to consider is the type of decision-making processes the modelling might be applied to in organisations, especially in the NHS. The management literature considers two types of decision-making processes that occur in organisations: routine decision making and problem-solving (March & Simon, 1958). Routine decision making occurs when there is a need for a solution to a recurring problem (March & Simon, 1958). This usually involves analysing practises or routines established over time in order to identify the optimal solution. As noted by Luoma

(2016), in routine decision making the benefit of using modelling is that it increases the effectiveness and efficiency of the decision-making process. Therefore, it can be assumed that relying on a results delivered by experts via the '*expert mode*' in such cases is sufficient to find an optimal solution.

Nonetheless, the problem-solving is characterised by messes with the scope of the problem not well defined and the process itself being time consuming and prone to error due to dealing with novelty and limited data. It happens in '*non-recurring situations*' or '*reoccurring situations*' but requiring a novel way of thinking (March & Simon, 1958, pp. 177-182). In such cases the role of modelling is to support and to inform the process of identifying the optimal solution that might involve considerable negotiations of stakeholders involved (Luoma, 2016).

The problem-solving is usually concerned with not well-defined messes or puzzles and the considerable efforts have to be spent to define the problem. Also, the fact that a number of stakeholders might be involved makes it possible that conflicts will arise due to existence of competing interests. There is an agreement that the facilitated modelling might resolve some of these issues as it leads to a decision that all stakeholders are more likely to accept. The practical implications of organising the workshops might be the deterring the people who consider it.

If we consider the fact, that the decision-makers demand for information might change over time or they might have further queries requiring the access to the model, both the expert mode and the facilitated modelling might not be flexible enough to facilitate such requirements. In case of the expert mode it is unrealistic to expect that the modellers would deliver new reports every time there is a change of data, as every change of service parameters requires models to be rebuilt which is time consuming and costly. Also, in case of the facilitated modelling once the model is completed there is no room for further amendments and changes of the code.

With all of these shortcomings of both approaches to model development and use, this thesis investigates an alternative approach based on the premise of not only engaging with the stakeholders during model development process, but also giving them the opportunity to work more closely with the model, via a web-based interface, or a platform. It is assumed that stakeholders would be given unlimited access to work with the model whenever needed. Throughout this thesis, the approach is called the end-user mode and is further explained in Chapter 7.

Although the OR researchers have called for development of flexible, and user-friendly modelling (Bowers *et al*., 2012; Harper & Pitt, 2004), there are limited examples of empirical studies that investigated this approach. An example would be the work of the health economics consultancy BaseCase, who are pioneering cloud-based platforms for models. Gijs Hubben, who founded the company, dedicated his doctoral research to explore the use of web technology to allow interaction with cost-effectiveness models via the Internet, to support decision making. He calls them '*user-friendly dashboards for models*' (Hubben, 2009). The company develops apps and web-based tools which could be used for example, by pharmaceutical company representatives when discussing the results of evaluations with reimbursement bodies. The software was successfully applied in case studies on vaccines (Hubben *et al*., 2007), and blood transfusion (Van Hulst, 2009). In both cases enhanced transparency of results was reported.

The other example is the study on adoption of a simulation model Scenario Generator by Brailsford *et al* (2011). Scenario Generator was implemented to facilitate the development of a patient pathway and resource allocation planning in the NHS. It was routinely used as a decision support system in several NHS Trusts.

In other disciplines that use modelling including the manufacturing industry, military, or environmental research, the trend is even more visible (Bastin *et al*., 2013; Morris *et al*., 2014; Murphy & Perera, 2002). However, although the method would remove the technological complexities associated with modelling there is little understanding of the appropriateness of this approach for certain types of modelling. Little is known if such approach makes modelling more useful when it is used in real-life decision making, what are the barriers and facilitators of such dissemination and about the development process undertaken.

This research will provide insight into some of these issues by testing the value of 'end-user mode' in a development of a decision model for telehealth COPD. Decision making relating to telehealth can be considered as problem solving. It occurs in highly complex environments, and involves several stakeholders with different, often competing interests. The examples of telehealth decision-making include: evaluating different alternatives in terms of the choice of equipment, service arrangements or resource allocation (Broens *et al*., 2007; Goodwin, 2010; Hendy *et al*., 2012; Joseph *et al*., 2011).

Furthermore, I believe that the end-user approach might remove some of the accessibility and acceptance barriers to modelling uptake in the NHS, commonly reported by the literature. This literature is further discussed in Part One of Chapter 4.

1.9 Research aims and objectives

This PhD study is concerned with the development and evaluation of a decision analytical model for COPD telehealth. The study researches the question: 'How to facilitate the use of economic decision analytical models in telehealth'.

The aims of the current research are threefold. Firstly, the study will investigate the impact of adopting conceptual modelling framework and usability testing during the model development process. This relates to the need for better information on the benefit of modelling in the evaluation of healthcare technologies and the need for greater use of modelling in telehealth field. The arguments in this thesis are not whether or not economic modelling is appropriate method for evaluating telehealth, but how current practices can be enhanced using the telehealth as a case study.

Secondly, given the evidence of the limited use of modelling and simulation in the NHS, the research attempts to reconsider the way people access and work with decision models. In particular, the research proposes '*the end-user mode*' of disseminating the decision model as an alternative to currently used practises.

Finally, the research will address the under-researched issue of usefulness of decision models in decision-making processes. In the standard HTA, feedback from users is rarely gathered. In case of telehealth, exploring the usefulness of models in real-life decision making and user satisfaction with these tools is important to gain acceptance for modelling by the telehealth community.

The specific objectives of the study are the following:

a) To elicit requirements for the economic model from telehealth stakeholders;

b) To assess usability of a web-based interface of the model;

c) To assess the end-user satisfaction of the model once it is used in real-life decision-making.

1.10 Links with Operations Research and Behavioural Operations Research

The proposed research is positioned closely to Operations Research (OR) and Behavioural Operations Research (BOR). With the great focus on improving decision making, Behavioural Operational researchers seeks to explore behavioural issues in model-based problem solving. They foresee their work as complementing traditional OR and facilitating better use of models especially by highlighting the behavioural issues that need to be taken into account when models are used for decision making (Hämäläinen *et al*., 2013).

Behavioural OR studies range from research that incorporates behaviour as a variable in the models, to studies that focus on behavioural aspects of model use. For example, Zohar (2002) incorporated behaviour into a modelling chain, Robinson (1998) studied how to assess the success of simulation projects, and Tako (2010) how to teach novices to model. These are only a few examples from a wide range of studies, but growth in the number and scope of behavioural OR studies has been observed (Crosnon 2013).

There is general consensus that simulation and modelling are not routinely embedded in healthcare decision making and that despite the evolution of software packages, the use of modelling is limited or moderate at best (Eddama & Coast, 2008; Williams *et al*., 2008). The barriers to simulation and modelling adoption in the NHS will be explored in more detail in Chapter 4.

The current study addresses the call to develop '*methods to remove barriers to modelling and simulation uptake*' by exploring the potential of using a web-based interface to for modelling telehealth COPD.

1.11 Thesis structure

The remainder of this thesis is structured as follow. Chapter 2 presents the Mainstreaming Assistive Living Technologies (MALT) study, and explains the extent of overlap between the MALT project and the PhD research. The chapter also presents the conceptual modelling framework and how it was adjusted to address the requirements of the current study. Within the conceptual modelling activities, the results of two literature reviews are presented: the review on NHS business cases and the results of the systematic review on cost-effectiveness model structures for COPD interventions. Furthermore, background information on COPD epidemiology, disease modelling, and COPD care pathways are also discussed.

Given the interdisciplinary nature of the topic area and the wide breadth of the literature relevant to issues of use, adoption and satisfaction with technology, the decision was made to conduct a review of theories related to the subject. Chapter 3 presents the methods and the findings from that review, as well as how they impacted the design of this study.

Moreover, in Chapter 4, the findings from a review of the OR literature that was conducted to provide a comprehensive insight into the existing barriers to simulation and modelling use in the NHS are presented. Additional literature review to identify appropriate methods to test the usability of a decision model, is also presented here (Section B).

Chapter 5 discusses research methodology and how it evolved during the course of the project. In particular, the chapter introduces the action research framework and how it was applied in this research.

Chapter 6 builds on the work presented in Chapter 2 and presents the results of the qualitative interviews to elicit requirements for the MALT model. The subsequent steps that involved

refining the data to develop a set of requirements that were later used by economic modellers are also presented.

In Chapter 7, the technical details of the MALT model are discussed. Also, a hypothetical scenario to demonstrate its capabilities is presented. In Chapter 8, methods identified in Chapter 4 are applied during model's interface usability evaluation. The iterative process of improving and subsequently refining the model's interface is also presented.

In Chapter 9 the findings from the evaluation of the user satisfaction are presented. Due to the iterative nature of this part of the research, the collection of the quantitative data was followed by subsequent qualitative interviews. The findings from both strands are presented and discussed in this chapter.

In Chapter 10, a summary of the principal findings drawn from qualitative and quantitative analyses and a discussion of their contribution is be presented. Additionally, strengths and limitations of the study are considered, followed by a recommendation for future research.

2. Conceptual modelling

In this Chapter, I further discuss the background and context of MALT model development, firstly, by describing MALT project aims and objectives, and secondly, by presenting the range of activities that were undertaken as part of the conceptual modelling. This included identifying relevant stakeholders, and how they were working with each other in four study sites, understanding care pathways, building a causal diagram of the main variables and reviewing previously published economic models of COPD interventions.

2.1 MALT study

Mainstreaming Assistive Living Technologies (MALT) was a collaborative project, conducted between 2011 and 2014 by researchers from the University of Sheffield, the University of Leeds and the Advanced Digital Institute. The project was funded by the Technology Strategy Board under its Assistive Living Innovation Platform and was used as a vehicle for this PhD research.

The overall aim of the MALT project was to examine barriers and facilitators to technology implementation by engaging with users, carers, managers and NHS staff in three workstreams:

1) Economic and financial modelling

2) Business modelling

3) Patient and staff acceptance

The MALT study adopted an action research framework (Reason & Bradbury, 2001), which is a methodology used by health services researchers in studies which seek to find a practical solution to a problem and to build an agency for change. As Action research is usually an emergent process, the study used a flexible research plan to respond to changes as researchers went along.

Four research sites were established for the study: sites A, site B and site C in South Yorkshire area of the United Kingdom and site D in West Yorkshire area of United Kingdom. The project involved a large number of stakeholders from both the study sites, and the telehealth industry. As part of the project, an Industry Panel was set up to promote research findings within the industry and facilitate their use in practice. Telehealth systems in these sites varied in terms of the number of patients, type of units used and the scale of telehealth systems. The overview of the main stakeholders in each study site is discussed in Sec 2.2 B.

In phase one of the MALT project, the in-depth case studies were conducted to understand business models and processes involved in telehealth provision in the study sites. Another project team investigated patients' and staff acceptance for telehealth. In total, 105 semi-structured interviews were conducted with the members of the staff, patients and carers. The results were validated with stakeholders during workshops held in each of these sites.

In phase two, the MALT team worked with sites as they developed or implemented telehealth to construct an understanding of what constitute a successful telehealth deployment. Some of the tools developed during the project, including an early version of the MALT model were tested in this phase. Additionally, a patient survey to examine technology adoption was also distributed to 2500 people. Finally, in phase three the findings from the research were disseminated through workshops and conferences to academia, industry, and NHS practitioners.

I participated in the work conducted by the modellers team, that set up as an objective to develop an economic model and financial planning tool that could be used by the NHS managers, Commissioners and the industry. The model was planned to be released in Open Access in the form of a web-based interface. My role in the project was to assist in the development of this tool, by conducting qualitative interviews with stakeholders and eliciting data on cost of telehealth

services in the study sites during Phase One of the MALT project. I was also responsible for conducting the usability evaluation of an online interface of the model during phase three of the MALT project. This work is further described in Chapters 6, 7 and 8. The relation of this thesis to the MALT project's phases one and two is illustrated in Figure 2, and the timescales of the MALT project are described in detail in Appendix A.1.

Figure 2: Relation of this Thesis to the phase one and two of the MALT project
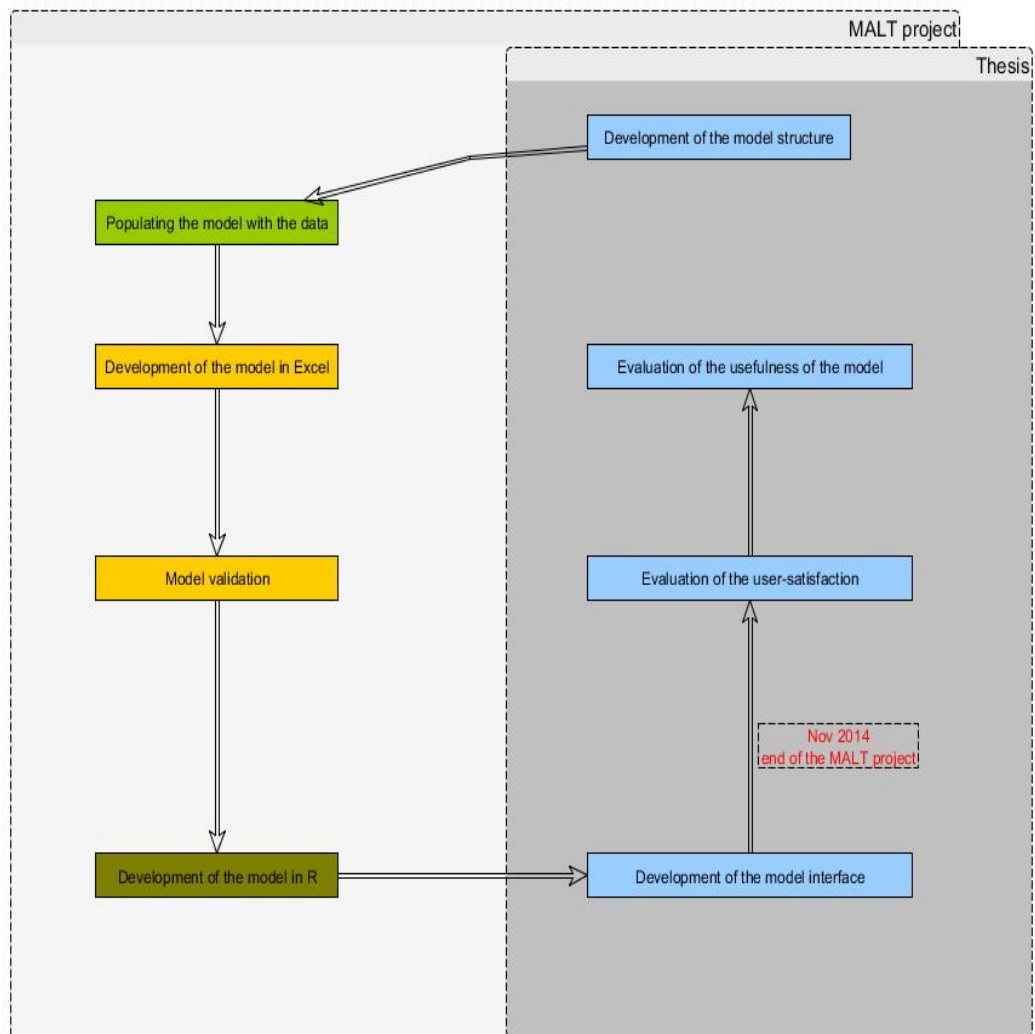
2.2 Conceptual modelling

A decision analytical model is a representation of reality, which uses several assumptions to translate real life processes into a set of mathematical equations (Eddy, 1985). A typical model requires information on natural history of the disease, current care pathways, resources required in the treatment process and expected benefits. There are a number of ways to depict the understanding of a decision problem, and there is no one perfect model. On the other hand, there is a limitation of how much information should a model include, with a common understanding that its purpose is to replicate the real world, not to be one (Tappenden, 2012). As explained by Robinson (2008), the problem situation ('real world') informs the modelling by feeding into the conceptual model that precedes the development of a computer model. Later on, through experimentation with the model the solutions are provided that may be verified when implemented in 'real world'. These relationships are illustrated in Figure 3 below.

Figure 3: Real world vs conceptual modelling (Robinson, 2008, p. 282)
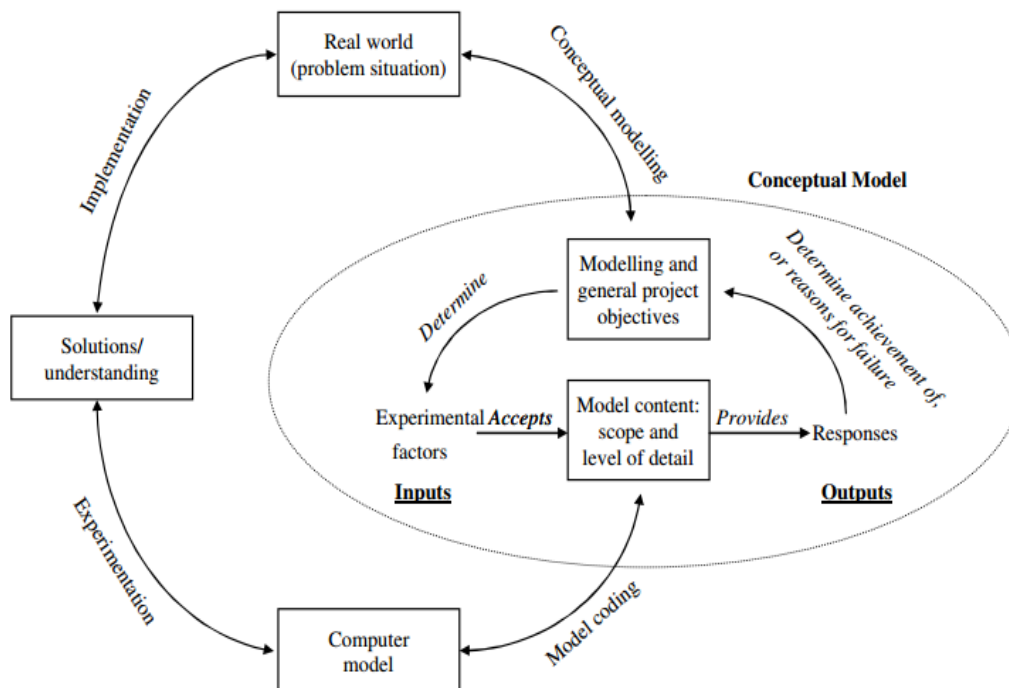
The conceptual modelling, takes on an important role as an '*activity related to translating the understanding of the decision problem towards a mathematical model-based solution*' (Kaltenthaler *et al*., 2011, p. 18), that directly informs the model inputs and outputs.

The role of conceptual modelling cannot be overlooked. It helps to develop, to share and to test one's understanding of a decision problem, and designing and specifying model structure (Kaltenthaler *et al*., 2011). Moreover, by providing a justification of abstractions, simplifications and omissions it helps to achieve transparency and strengthens models' credibility. Users can also gain trust in models results if its development process is clearly reported (Kaltenthaler *et al*., 2011) Most importantly, it helps to avoid errors early, as changes to the model structure once it is populated with the data usually require the model to be completely re-built.

In general, conceptual modelling frameworks provide a step-by-step approach to follow in order to select the appropriate model structure and the components of the model. Conceptual modelling usually takes place in the early stages of model development and is perceived as an iterative process, where modellers go back and forth between stages depending on the complexities of the task. Conceptual modelling is now recognised as a distinct method within the process of model development, but its use among many modellers is still rather informal and poorly reported (Kaltenthaler *et al*., 2011).

The importance of conceptual modelling was highlighted in a number of health economics guidelines, for example, the International Society for Pharmacoeconomics and Outcome Research (ISPOR) Task Force produced the general recommendations for modelling, where conceptual modelling was acknowledged and recommended (Roberts *et al*., 2012). However, the specific recommendations were limited to suggesting the use of influence diagrams and concept maps without any further practical advice.

A formal recognition of conceptual modelling in health economics has emerged relatively recently, with a number of health economics researchers such as Tappenden (2012) and Kaltenthaler *et al*. (2011) providing excellent background to the topic. In other fields which use mathematical modelling, such as Operations Research or Engineering, there is also a record of various frameworks developed for the specific types of models, for example, Robinson's (2011) framework for DES models.

The development of NICE guideline for health service configuration (2014) has largely addressed the issue of distinguishing between problem-oriented and design-oriented conceptual models, although the distinction was first developed elsewhere (Lacy *et al*., 2001). Problem-oriented conceptual models take on the form of a diagrammatic representation of care pathways and interaction between main processes and allow developing an understanding of the processes within modelled problem. On the other hand, design-oriented models allow justifying the choice of a model structure by specifying what choice is feasible in the light of available data.

Recently, Squires (2014) has researched conceptual modelling practises among modellers as part of her doctoral research. She developed a conceptual modelling framework for public health interventions based on the results of her qualitative research and a systematic review of published conceptual modelling literature.

Squires described her conceptual modelling approach as based on four main activities: (1) aligning the framework with the decision making process, (2) identifying relevant stakeholders, (3) understanding the problem and (4) developing and justifying the model structure. Under the framework, model development is understood as an iterative process, where modellers go back and forth between the stages as illustrated by double headed arrows in Figure 4 bellow.

Figure 4: Squires' conceptual modelling framework (Squires, 2014, p. 195).
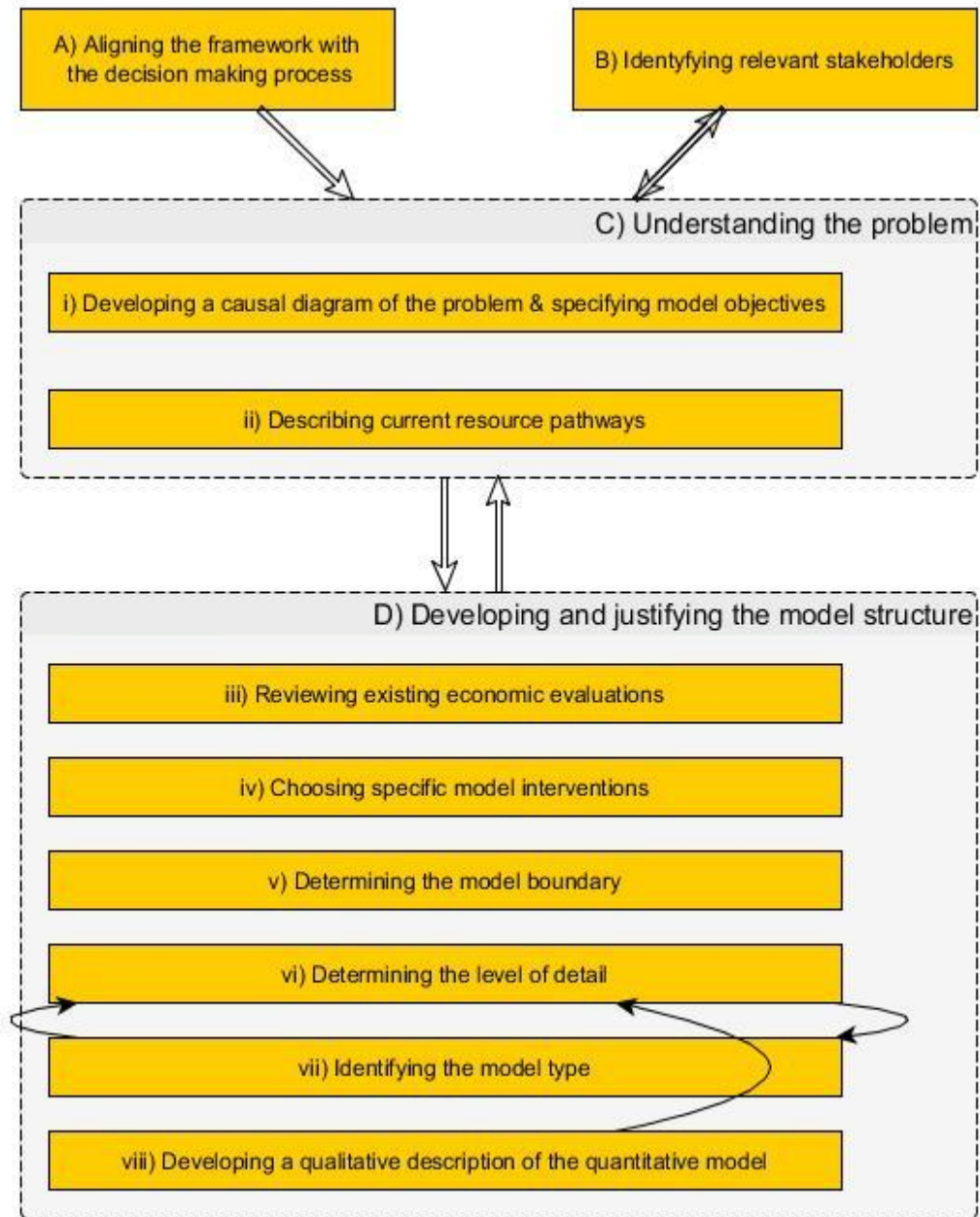
In stage (A) the decision of how to use the framework in a particular modelling project is undertaken. The next step (B) requires identification of all relevant stakeholders; defined as the people who benefit from the system, the people who perform tasks in the system, and the people

who own the system. Usually, clinical experts, decision makers, and patient representatives are considered here.

Activities undertaken to understand the problem include causal relationships and modelling objectives, as well as current resource pathways. Within the framework, Squires (2014) provides a set of questions that help to understand the intervention and its impact. In the later stage, when the model structure is being finalised, modellers can review existing economic evaluations, to choose model interventions, the model boundary and the level of detail. The last stage is the identification of the type of the model, as well as the qualitative description of the quantitative model.

The advantages of Squires' (2014) framework and its ultimate adoption in this study came from the fact that it has been developed for public health interventions, which similarly to telehealth are considered as complex interventions. For example, Robinson's (2008, 2011) framework was developed for modelling outside health and Kaltenthaler's (2011) for general HTA. Moreover, Squires' (2014) framework requires stakeholders' involvement, and this was perceived as crucial for this project having in mind the need to resolve issues around complexities of telehealth service provision. One of the main advantages of the framework over other published guidelines is that the choice of the type of the model, (for example, Markov or DES) is made as one of the last actions. This fact had the advantage over Robinsons'(2008, 2011) framework that strictly sets out that the model will be a simulation at the outset of the process.

The application of the framework in this study was seen as a way to minimise structural uncertainty and enhance the quality of the model. The framework facilitated MALT model development and scheduled the workload for the team members.

The following sections (2.2.1 – 2.2.4) describe the activities undertaken in the process of developing the conceptual model for the MALT study. The headings of each section refer directly to the stages described in the modelling framework in Figure 4.

2.2.1 Aligning the framework with the decision-making process

At the outset of the project, monthly meetings of the modelling team were set up to discuss and plan the workflow of the project. We discussed how the framework would be used in this particular study, taking into consideration the timescales of the MALT project set by the Technology Strategy Board and the available resources.

The aim of the project was to develop a model of telehealth interventions for two long term conditions (CHF and COPD), and my role in the project was to undertake the tasks related to the development of the conceptual model for COPD telehealth.

Because of the specific character of telehealth interventions, the project faced several challenges, which dictated that a number of additional activities were undertaken. For example, it was necessary to understand business models and some of the operational issues related to telehealth provision.

Due to the complexity of telehealth systems in the MALT study sites, the decision was undertaken to explore business and service provision models by conducting qualitative interviews with the main stakeholders. As this work overlapped with the research undertaken by other MALT team researchers the regular meetings were held to discuss the main findings and challenges in the field. The outcomes of this work are presented in section 2.2.2 of this chapter.

Furthermore, in order to develop a valid cost-effectiveness model a review of previously published economic models for COPD was undertaken to identify structures that could be replicated in this study as well as ones that should be avoided due to their poor quality. This was to be undertaken as stage D of the conceptual modelling framework and is discussed in Sec 2.2.4.

Additionally, due to the fact that telehealth very often imposes the re-organisation of the health services, I undertook a review of NHS business cases. Also, NICE costing tools were reviewed to understand the methods the NHS is using in evaluations of its services and commonly accepted standards for reporting financial data.

Finally, as part of the framework activities, I conducted qualitative interviews to elicit stakeholders' requirements for the model and to derive study-specific cost data for the model. The results of the requirement elicitation are presented in Chapter 6.

2.2.2 Identifying relevant stakeholders

Between August and December 2012, I assisted in nine qualitative interviews conducted by a researcher from the University of Leeds Business School (LM) with the operational and managerial staff in each of the study sites. This work provided vital background information about telehealth services. The topic guide included questions pertaining to the currently used care pathways and how these have changed following the introduction of telehealth. Moreover, we discussed the NHS procurement and commissioning processes that took place in each of the NHS Trusts; the relationship with the NHS providers and telehealth suppliers, and identified key people, who could be interviewed about the financial aspects of the service.

I participated in the weekly meetings with the MALT research team members, during which we discussed the findings from the field. That helped me to develop my understanding of the range

of stakeholders involved in the provision of telehealth in each of the sites, and to create 'stakeholders' maps' for each study site as shown in Figure 5.

Telehealth systems used in the MALT study sites consisted of a 'hub', a stand-alone device, which could be used with a range of peripherals such as a pulsometer, an oxygen saturation meter, a blood pressure meter and weighing scales. These were given to patients to allow daily monitoring. Usually, patients with COPD had oxygen saturation measured as a standard, and patients with CHF had their weight monitored as standard. The data from the device was transmitted for review from the patients' home to a triage centre where health professionals could take action if necessary. The intervention of health professionals could include a telephone consultation, home visit, or in case of an emergency, an ambulance call-out followed by hospital admission. Depending on a patient's condition and needs, they also received daily advice and filled in a symptoms questionnaire. Telehealth provision varied from 4 months to up to 2 years depending on individuals' needs and severity of the disease.

Study sites characteristics

In exception of site D, all study sites had population close to, or exceeding half a million people. Community matrons estimated the number of COPD patients in their localities as 400-700 people. These included patients of different severity of the disease and hence different care needs. The comparison of MALT study sites is presented in Table 2 below.

Table 2. The MALT study sites characteristics

	Site A	Site B	Site C	Site D
PCT registered general population[1]	570 792	420 000	309 141	821 310
Number of COPD patients per year[2]	400-600	700	550	650
Number of patients on telehealth[3]	82 (May 2010)	79 (Dec 2012)	180 (Dec 2012)	200 (June 2013)
Number of telehealth units available at the time of research taking place [3]	104	39	200	34
Year of introducing telehealth	2007	2009	2006	2006

[1] *source NHS Commissioning Board,* [2] *source Community Matrons Caseload;* [3] *These were estimated at the end of MALT study phase one. Also, the number of patients and telehealth units in use fluctuated over time.*

The overall aim of the telehealth services was to encourage independent living, support hospital discharge and prevent unnecessary hospital admissions. The eligibility criteria were similar in all study sites and focused on the following:

1. Patients with a diagnosis of COPD,
2. Patients who had between one to three previous hospital admissions,
3. Patients who are willing to use telehealth technology,
4. Patients who are able to communicate and read in English,
5. Patients who have telephone landline and telecommunication network at home.

Community matrons, case managers, specialist respiratory nurses or GPs further assessed patients for their eligibility. Cognitive impairment and existence of other co-morbidities were among exclusion criteria. COPD pathways and are discussed in detail in section 2.2.3.

The main stakeholders

The main stakeholders identified in the study sites were Clinical Commissioning Groups (CCGs), community care providers, acute care providers, primary care providers, telehealth providers and local authorities. In health services research patients are usually regarded as stakeholders, but in this study I decided that they were not to be taken into consideration, as the MALT model was designed with other users in mind.

Telehealth in site A was introduced in 2007 when a clinical trial was conducted, and in 2010 it become part of NHS community care services. One commercial company was commissioned to provide telehealth equipment, and to secure technical aspects of running of the system. Community nursing team was responsible for installing the equipment in patients' homes (technical triage) and screening daily readings. In 2013, a therapy assistant role was created, to carry out the equipment installation and removal from patients' homes. The local council did not take part in the provision of telehealth services in this region.

In site B, telehealth was introduced in 2009, when the NHS purchased first, multi-user hubs (the equipment that could be used by multiple patients) for residential homes. Since then, additional telehealth equipment was purchased from two telehealth providers. The equipment offered similar

functionality but had different specifications. Also, because of different contractual arrangements, these hubs were installed, maintained and stored differently. As part of the contract, one of the telehealth providers was responsible for installation, maintenance, removal and storage of its equipment. The second provider was only responsible for supplying the equipment and integrating it with the system. Initially, community matrons were responsible for installing and removing this type of equipment from patients' homes, as well as providing training to patients and on-going technical support. In 2012, the local council overtook that role. As the locally, community interest group was already providing a range of community care services, they also offered to take on the role of managing telehealth services for COPD patients.

Day to day operation of the system was complex. With two systems in place, admin team had to log into two different computer systems to review patients' readings and alerts. They had to manually copy the readings to the NHS system (SystemOne) in order for the information to be available to other clinicians. Community matrons reviewed their patients' readings daily and contacted the patients requiring follow-up. This had impact on the decision to improve the system by de-commissioning one of the telehealth suppliers, something that did not happen during duration of the MALT project.

In Site C telehealth service was first introduced in 2006. As there were varying practises within different community care teams utilising telehealth, a dedicated post of a telehealth coordinator was created to support and encourage the use of telehealth for COPD and CHF patients, and to oversee the process of installation and patient monitoring.

However, for a variety of reasons, the system was failing. For example, in case of telehealth for Heart Failure patients, clinical support workers usually installed the equipment in patients' homes and provided patients' training on the same day. They were responsible for solving technical

problems and for providing maintenance for these hubs. In all these roles, support workers had support from the telehealth company, and could only consult them on some matters.

Support workers were also responsible for monitoring patients' reading but needed input from community matrons to determine the most appropriate response in case of an alert. That meant that community matrons were sometimes performing both the technical and clinical triage on a daily basis, which doubled their workload.

In October 2012, the local council was contracted to provide installation, technical support and monitoring for telehealth patients, as they were already running similar services as part of the existing councils' telecare services. While the new system for installation has been successful, there were a number of problems with monitoring, and as a result, the contract for monitoring was cancelled in February 2013. The telehealth coordinator has taken on this role, as because of her clinical background she was able to provide both the technical and clinical triage.

In site D, telehealth has been used by community health services in this site since 2006 with varying degrees of success. Due to the structural changes in the NHS in April 2013 when three Clinical Commissioning Groups had been established in the region, the existing telehealth undergone major changes.

A private company in partnership with a social enterprise were commissioned to deliver the new telehealth service. They were already providing out of hours' telephone triage services in the region. In the case of telehealth, they had taken on patient assessment, technical triage and overall project management. In the process of tendering, they were introduced to a neighbouring local council, who was successful in delivering telehealth. As a result, they subcontracted this body to deliver installation, maintenance, repairs and storage of the equipment for a new service. A

different telehealth company was contracted for suppling telehealth hubs, as it offered equipment promising to overcome previous barriers with connectivity, portability and flexibility.

To sum up, the exact roles of stakeholders and the degree of the involvement in telehealth provision differed enormously between the study sites. In all study sites, Clinical Commissioning Groups (CCGs) (formally Primary Care Trusts) were responsible for designing, developing and authorising telehealth services. With the exception of one site, CCGs were the ones who covered the cost of purchasing telehealth equipment. They also financed the operational costs of running telehealth in these trusts.

Telehealth services were provided mainly by community care providers or private community interest groups. As observed in the study sites, local authorities have also taken on some roles in telehealth provision, such as providing installation of the telehealth kits, their repairs and storage. Some taken on the role to provide technical triage especially if they were already running telecare services in the region.

Telehealth industry representatives consisted of the companies that manufacture the equipment and 'middleware' companies – the ones that provide a range of telehealth-related services, such as, technical support, technical triage, or the installation and removal of the units from patients' homes.

The major differences between study sites as were noticed in regards to what entity provided installation and actual monitoring of patients records (technical triage). Also, various procurement scenarios were present in these sites, from payment for monthly managed service to payment for each connected patient. All telehealth providers informed me that their aim was to extend the provision of telehealth beyond the equipment and integration of the system, but provision of a so

called 'managed system' which includes all components of the service, was only seen in case of the provider in site B.

The comparison of the major actors in each of the MALT study sites is presented in Figure 5. The NHS providers are presented as green boxes, telehealth providers as yellow boxes, local authorities as purple boxes and community interest groups or similar as orange. The range of the roles undertaken by these entities are shown in grey.

Figure 5. The comparison of the stakeholders from the MALT study sites
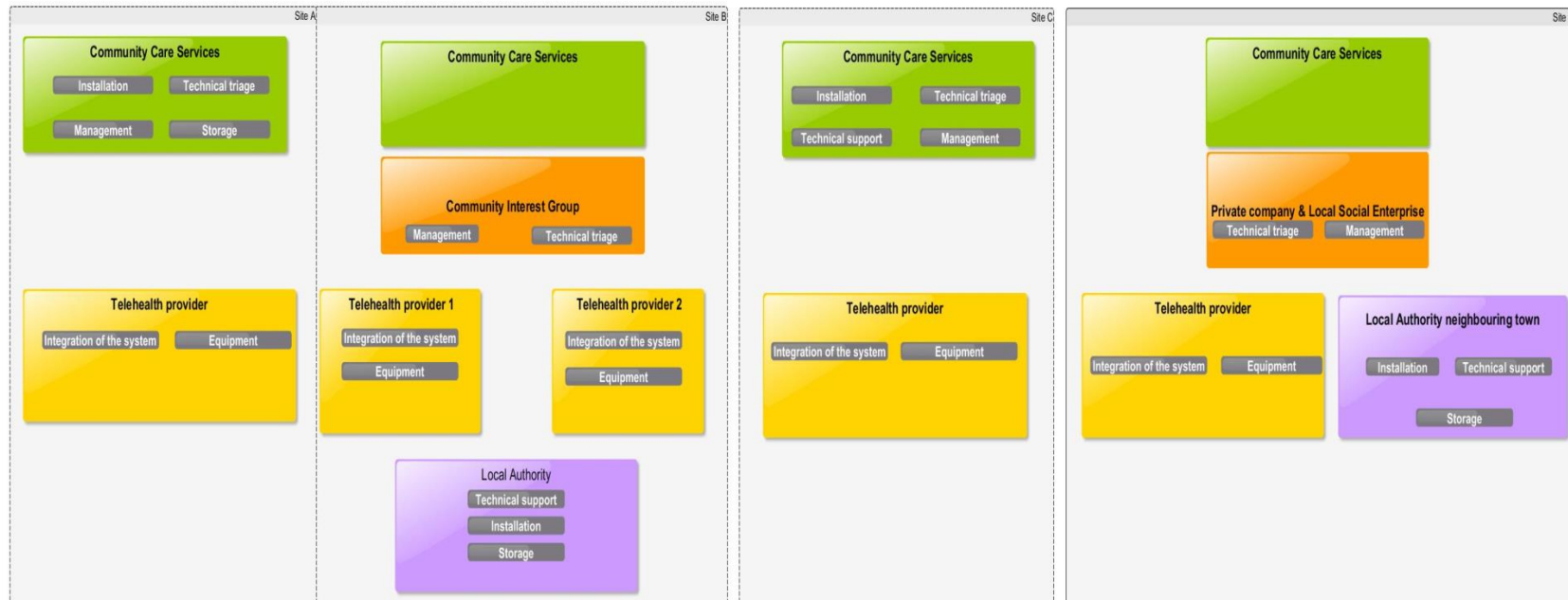
2.2.3 Understanding of the problem: Disease modelling

This stage of the conceptual modelling included reviewing the literature on COPD to understand the natural history of the disease; how telehealth affects its progression and impacts upon utilities and costs. The aim of this activity was to depict the causal relationships between the main variables, in the form of a causal diagram. Squires (2014) advised that the causal diagram should be developed with the focus on 'what is the problem', by depicting all variables and assessing their positive or negative impact. Based on these, the implications of the researched problem can be identified and assessed.

COPD Epidemiology

COPD is a chronic lung disorder, affecting nearly 18.5 million people in the UK alone, and its prevalence is estimated to grow to 20.6 million in the UK (Department of Health., 2011). COPD is the second highest cause of hospital admissions in the UK, with nearly one-third of patients being re-admitted within 90-days of admission (Department of Health., 2011). COPD is a very complex disease in which lung function worsens over time even with the best available care. From the perspective of the economic model, two factors from the natural history of the disease have to be considered: a decline in the lung function and episodes of the deterioration of the symptoms known as exacerbations. Exacerbations are defined as a worsening of the disease symptoms, requiring medical intervention lasting up to three days (Rabe *et al.*, 2007). There are two types of exacerbations: severe and non-severe.

In terms of the natural history of COPD, there is a relationship between lung function and symptoms, for example, if lung function worsens, symptoms, such as cough and breathlessness will usually deteriorate. Another relationship is between lung function and exacerbation severity; when lung function becomes more limited, the severity and frequency of exacerbations tends to

increase. This has impact on patients' quality of life. Also, exacerbations are associated with high risk of mortality, therefore their prevention and reduction is a target for management strategies (Rabe *et al*., 2007). Throughout the literature, Forced Expiratory Volume in 1 second (FEV$_1$) has been used as an indicator of the severity of the disease, one that drives both symptoms and exacerbations. From the perspective of the decision model, the fact that exacerbations constitute the highest cost in COPD care, (as they are the main reason for hospital admissions) was seen as important (Andersson *et al*., 2002).

COPD service objectives

I developed my understanding of the physiology and epidemiology of COPD with a consultant physician (Prof S Parker). Professor Parker stated that in his practise he noticed that if patients, even in very severe COPD stages learn how to 'self-manage' their condition and have standby rescue medications at home and support from community palliative team, they will not be admitted to hospital as often. On the other hand, anxious patients, even with mild COPD, tend to visit GPs and use A&E services regularly, purely because they have not mastered self-management.

In OR and health economics casual relationships between events or consequences of events are very often depicted using causal diagrams. These diagrams help to understand how main variables influence each other that in turn is essential to build a mathematical form of a decision model (Rodriguez-Ulloa & Paucar-Caceres, 2005). The causal diagram is structured in the way that variables are connected by arrows, with variables at the head of the arrow depicting those that depend on those at the tail. Also, a positive or negative sign is used to denote the increase or decrease in these variables. For example, a negative sign next to the variable 'exacerbations' and 'hospitalisations' means that the decrease of exacerbations leads to a decrease in the number of

hospitalisations. To present my understanding of the interconnection between the main variables

in case of a COPD disease I developed the causal diagram presented in Figure 6 below.

Figure 6: A simplified graph of the main causal relationships between modelled variables
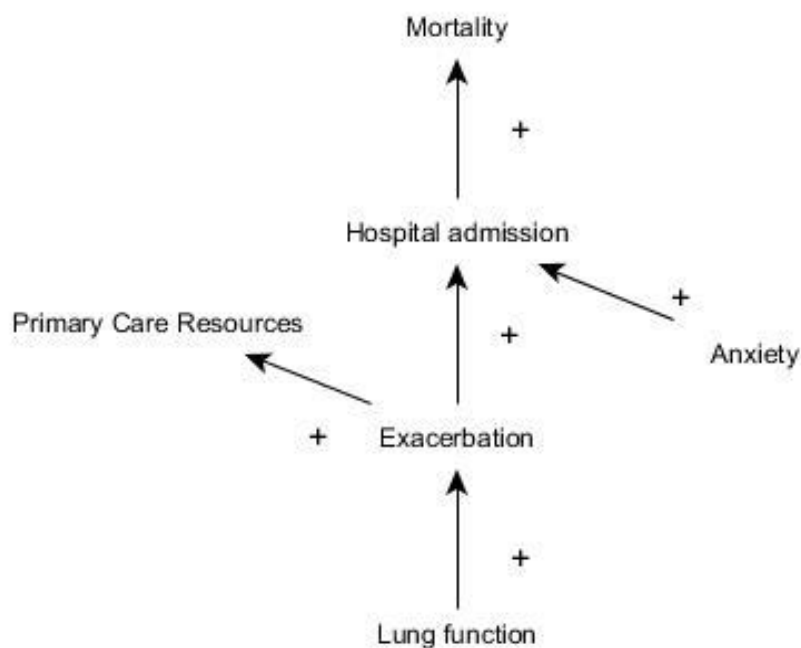
The core of the causal diagram in Figure 6 is the lung function, which if limited, affects the onset

of exacerbations. These in turn have a direct impact on hospital admissions. Exacerbations

followed by hospital admissions are associated with a significant risk of death. Exacerbations can

be also dealt with in the primary care, and in such case, the increase in exacerbations leads to an

increase in the use of primary care resources (the number of GP visits or nurse home visits).

However, the literature suggests that the increase or decrease in the number of visits depends on

the set-up of telehealth system (Bardsley 2013). For example, if telehealth system gives patients

an option to contact a consultant in a hospital, they would not need to visit their GPs. The

assumption that the increase in the number of primary care visits will occur was based on the

knowledge that in the study sites telehealth systems were set up to prompt visiting a GP in case of any non-severe exacerbation. Another variable in the causal diagram was seen as patients' anxiety where an increase in anxiety leads to increase in the number of hospital admissions.

Another way to depict how telehealth affects the main variables is by using a concept map. Concept map is a graphical representation of a problem, where concepts are represented in a hierarchical fashion: the most general at the top and more specific at the bottom of the map (Novak & Cañas, 2008).

My understanding of how the main variables are affected when a patient is provided with telehealth is presented in Figure 7 below. Telehealth enables early identification of worsening of patients' health status so that care is provided faster, preventing further complications or even death (Steventon *et al*., 2012). Telehealth also reduces patients' anxiety and improves medication compliance, which are both vital elements of self-management (Joseph, 2006; Taylor *et al*., 2014). In turn, better self-management leads to reduction in the use of healthcare resources. Telehealth has also an impact on quality of care because it facilitates faster response in case of an exacerbation and significantly improves patients' records as they are updated daily (Ekeland *et al*., 2010; Jaana *et al*., 2009).

Figure 7. The concept map of the impact of telehealth
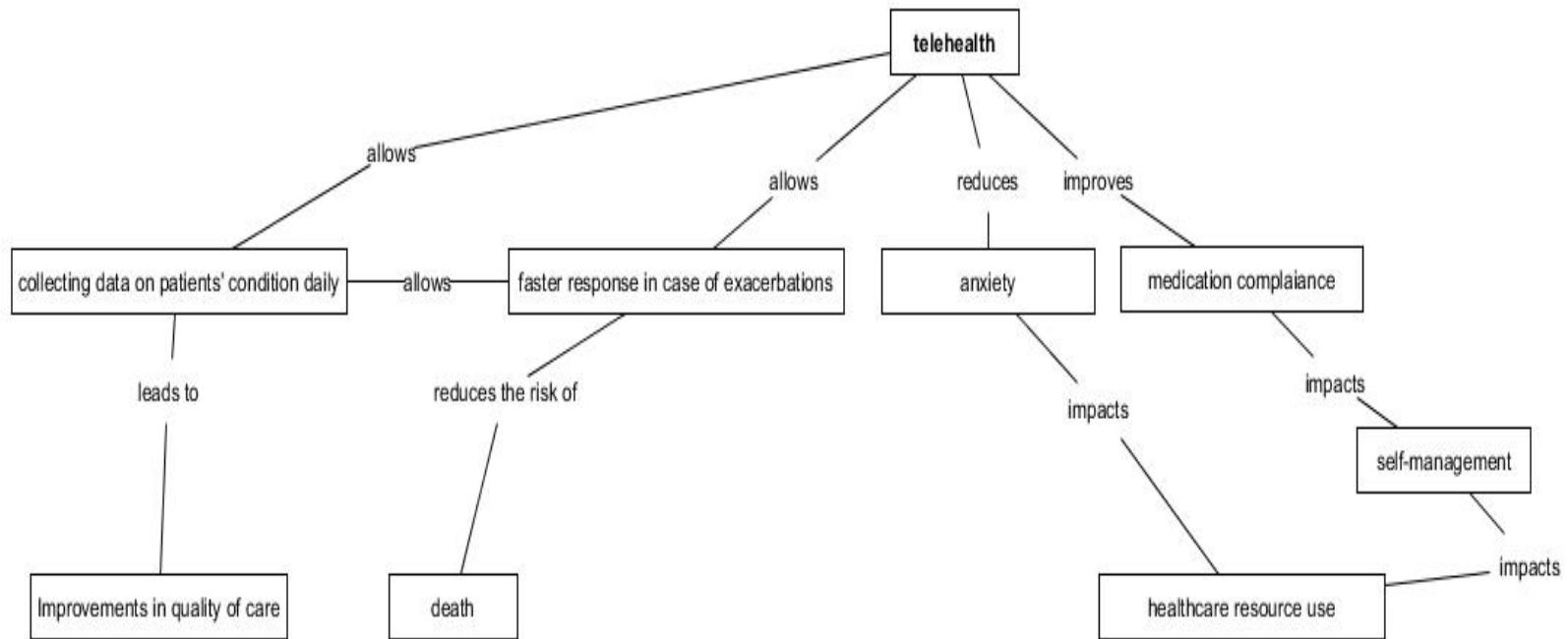
COPD care pathways

The next stage of the conceptual modelling involved developing an understanding of the COPD care pathways. The care pathways vary between NHS organisations, and a generic approach does not exist. I constructed my understanding of the usual care and telehealth care pathways based on four sources of information: a previous evaluation of a telemonitoring service in the region (Fitzsimmons *et al*., 2011), recent, high-quality randomized controlled trials of COPD telehealth interventions (Gellis, 2012; McLean *et al*., 2011; Vontetsianos *et al*., 2005), The Map of Medicine (2014) and NICE guidelines (2015).

In a usual care pathway, the patients are assumed to receive care mainly from community care teams or primary care. They undergo regular check-ups either in the GP practice or during consultations with consultants in secondary care. In case when patient health is deteriorating such that he is not able to cope at home, he might be admitted to the hospital. Hospital admissions might result in discharge or death as presented in Figure 8.

Figure 8: COPD care pathway
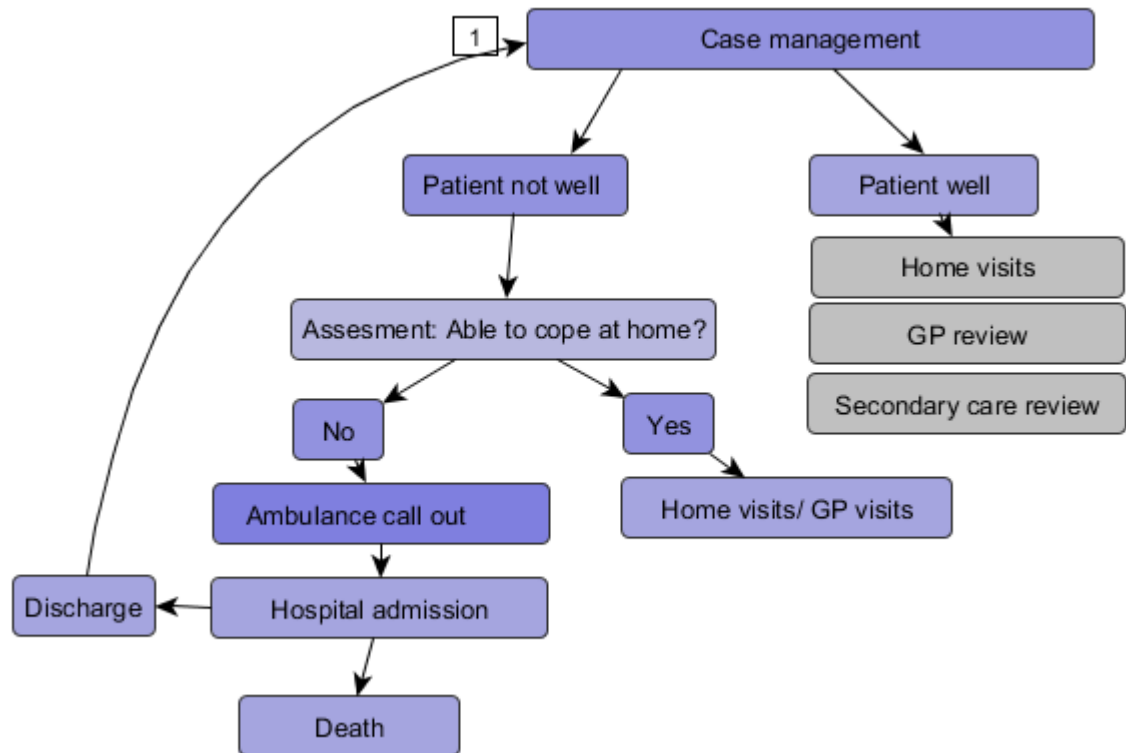
In a typical telehealth care pathway; if eligible, a patient is prescribed telehealth by either community matron or the case manager. Once he agrees, equipment is installed in his home. A patient has to undergo training, and there is usually a phase when a patient learns how to use the system. It is possible that he would generate what is called 'false alarms', which are flagged up by the equipment but are not associated with medical emergency, for example, when a patient uses equipment in a wrong way or there is an error of a reading. On the other hand, 'true alarms' indicate worsening of his health and require medical staff attention. After that phase, in case of triggering an alarm, patients' readings are assessed by a member of staff (technical triage), and in case of an emergency by a medical professional. The patient could be called to assess the severity of the situation, visited at home or sent to hospital. This is a phase generating the highest gains or

losses of QoL and also costs. I assumed that in the event of a hospital admission, a patient could either be discharged home and, therefore, carry on monitoring or being referred for a review by a Case Manager, marked on the diagram as (2). I also assumed that in some of the cases, a hospital admission would be followed by death. The telehealth pathway is presented in Figure 9 on the next page.

Figure 9: Telehealth pathway for patients with COPD
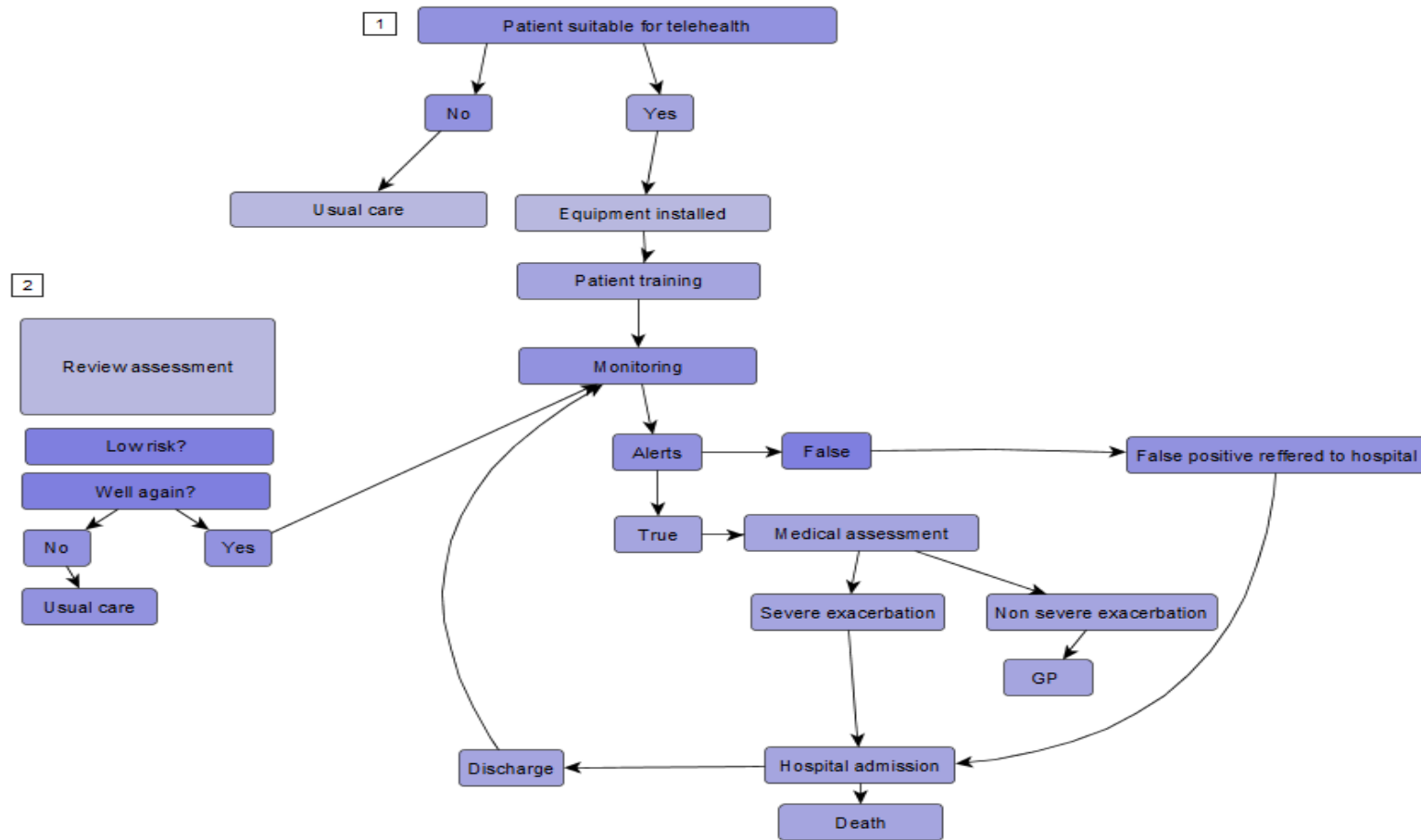
*False positives are classified as true, but should be recognised as 'false' at the medical assessment and returned to monitoring. False negatives are classified as false, but are likely to generate further alerts following further deterioration, until they are recognised as 'true'.

Risk Stratification

The next question arising in the project was the issue of how case managers decide which COPD patients would benefit most from telehealth. I held a meeting with an NHS Procurement Specialist (Dr Jeff Anderson) to discuss currently used risk stratification tools. Risk stratification tools use regression analysis based on a set of variables such as patients' age, co-morbidities, and the GP practice they belong to, to determine which patients are at highest risk of hospital admissions in the following year. This helps to identify patients who are most likely to be admitted to hospital and would benefit from health interventions such as telehealth.

I identified seven tools that are currently used by the NHS and health authorities, namely: Person-Based Resource Allocation (PBRA), Patients at Risk of Re-hospitalisation (PARR), Combined Predictive Model (CPR), Adjusted Clinical Groups, High Impact User, Health Numeric RISC and Population Health Toolkit; however, several commercial models are also available on the market. They differ in respect to the scope and data used, for example; CPR uses data from primary care; whereas the most commonly used PARR is based on the data on hospital admissions.

Risk modelling was initially considered as an option in the MALT model, but the task was dismissed due to lack of data and the use of different tools in different locations. Risk stratification has been incorporated in the MALT model via a simple measure – the number of hospitalisations in the previous year.

Additional activities I undertook in this stage included review on NHS business cases and NICE costing tools in order to understand NHS approach of presenting financial information. These are discussed in the sections below.

Understanding of the problem: Financial modelling

Review of the NHS Business Cases

The review of the NHS business cases and NHS guidelines for business cases was conducted to understand common NHS practices for reporting financial information.

A business case is developed for any project in the NHS that involves business or service development. The document is typically developed over time in three distinct stages before it is presented for an approval of a decision-making body. Firstly, a preliminary business case, which sets out the need for the project; is developed. It is followed by the outline business case, which presents what are the options for the project, and if successful, the full business case is developed. In case of capital projects exceeding £35 million, the standards set out by the Treasury have to be followed.

The scope of this review was to identify examples of different types of business cases in order to provide contextual information rather than identify all cases that match the inclusion criteria. The search was conducted in the web and grey literature databases and from the key documents references, to capture a sample of publicly available documents. The terms included in the search were: 'Full business case NHS', 'NHS business case', 'business case'.

Inclusion criteria:

- NHS Full business cases for health service development or restructuring

Exclusion criteria:

- Outline business cases (OBC) and cases for capital development, for example, merging hospitals, building new wards.

The searches were undertaken in September 2012. The number of the results by the source included:

1) Google and google scholar (35)

2) Open grey and greynet.org (0)

3) Key documents references (3)

The number of identified Full Business Cases was 35, from these, 7 were excluded as they only provided a summary of the venture (Outline Business Case). 18 documents were excluded as related to capital developments, either investments in new hospital buildings, or merging hospitals as this would be quite distinct from telehealth projects.

The review included seven Full Business Cases and three guidelines for developing NHS business cases. The overall low number of cases could be caused by the fact that this type of document is circulated internally and the majority are not made available for general public. Search on the Intranet of certain NHS organization, could help with the results, but was beyond the scope of this project. The details of the included and excluded cases are given in Appendix A.2.

The cases included in the review considered a variety of services, for example: neonatal services (Oxford University Hospital, 2012), A&E services,(Hull and East Yorkshire Hospitals Nhs Trust,

2012), cardiology services (The Ipswich Hospital, 2012b), sterilization services (The Ipswich Hospital, 2012a), pathology services (The Royal Wolverhampton Hospitals NHS Trusts, 2011), and general hospital restructuring (The Shrewsbury and Telford Hospital NHS Trust, 2012). The most detailed case referred to the plan for re-designing services for elderly patients (NHS Lothian, 2010).

The NHS Business Services Authority requires all business cases to be in a certain format, centred on five key principles: (1) that sufficient information is provided to assess why the project is desired, (2) that a preferred option is identified on the basis that it brings value for money, (3) that sufficient details are provided on how the project is going to happen, (4) that longer-term implications are considered, and lastly, (5) by presenting that the project is achievable (Mersey Care NHS Trust, 2012).

All the cases in this review followed that format and adapted additional analyses if necessary. The level of detail presented in each of the stages (1-5) depended on the scale and complexity of the project.

Option appraisal was usually performed and reported as using a cost- benefit analysis of at least three options, although these would be better described as cost-consequence analyses. At the outset, this was very often presented with a comparison of physical, technology and operational requirements, which were then scored and ranked against the benefits to identify the most viable of options. In one of the cases, an economic model was used to perform a cost-benefit analysis of strategies for the development of new cardiology services. Using the model, they estimated the net present cost of each of the options and analysed the predicted growth of the population and the incidence of cardiovascular diseases. This was used to justify the benefits and forecast future

activity (The Ipswich Hospital, 2012b). Interestingly, this included intangible benefits that cannot be quantified in financial terms such as patients' travel time.

Looking specifically at the financial information included in the business cases, all the analyses were conducted in accordance with the Green Book (HM Treasury., 2003). Once the preferred option was identified, the majority of authors carried out a forecast analysis of its financial implications. A detailed phasing of the capital expenditure and the source of the funding were demonstrated. In the majority of the cases, the discounted cash flows were presented.

In summary, the findings from the review suggest that in order to be in line with NHS practices, the following details should be reported:

I.   The analysis of costs should include capital costs, opportunity costs, and the revenue costs for each of the considered options. The later usually includes the cost of staff, direct expenses, overheads and capital charges. Costs are always expressed as full lifetime costs. In case of a proposal that involves equipment, the option of leasing has to be explored. Optimism bias of 5% is considered for each of the options.

II.   Net Present Value (NPV) analysis for costs and benefits is important. NPV allows demonstrating the present value of future expenditures. The technique involves discounting a stream of cash flows over time, reflecting the principle that the value of costs and benefits in present day terms decline over time. This helps to realise what funding is required for the project and prevents the projects form being abandoned when the 'true costs' are revealed.

III.   For the financial information to be robust, it has to include the sources of the revenue and required capital. In the case of projects designed to improve efficiency, of particular importance is the demonstration that they can deliver the required savings over an acceptable period of time. However, it is sufficient to present the savings in monetary terms, but it is not required to indicate whether that would allow redeployment of resources or disinvestments. Cost savings and benefits to other parts of the public sector were also not considered.

NICE costing tools

A separate category of reviewed documents included NICE costing tools for COPD (National Institute for Health and Clinical Excellence, 2010, 2011). These documents are available from the NICE website for each disease and treatment type and provide basic guideline on how to estimate total costs of implementing the guidance, given the population size and costing assumptions. For example, the COPD tool analyses what would be the provider cost of implementing the guideline over three years'. The cost of medication, hospital admissions and cost of alternative treatment options, and complications, are considered (National Institute for Health and Clinical Excellence, 2011).

The review of costing tools identified the features that determine the end-user character of such tools. For example, a drop-down menu to choose the population of interest out of entire UK population, and a detailed summary description of costing calculations were deemed useful. Additionally, the clear description of assumptions and information on how to operate the tool, are something that I considered to replicate in the MALT model.

2.2.4 Developing the model structure

Developing the economic model for telehealth is considered complex as it includes both modelling the progression of the disease and service planning issues. The modelling of disease progression underpins the assessment of cost-effectiveness while modelling service planning underpins any exploration of different implementation packages (e.g. speed of uptake). The need for both cost-effectiveness and service planning modelling was identified in the MALT proposal as important to reduce barriers to increased uptake of telemonitoring (Coates *et al*., 2013).

The tasks reported in this section are concerned with modelling of the disease and include reviewing previously published economic models for COPD interventions, choosing model interventions, determining the model boundary and the level of detail. The application of this approach to cost-effectiveness modelling has been developed elsewhere, although not applied widely, nor specifically to this topic. The development of the service planning model is explored in a subsequent chapter as it was undertaken with a new set of methods, and as such, represents a unique research application.

Review of previously published economic models for COPD interventions

In order to develop an appropriate model structure; I conducted a systematic review on previously published economic models for COPD interventions, focusing mainly on pharmacotherapies. Surgical procedures and organ transplantation were excluded. The aim of this activity was to identify the structure for economic models that could be replicated in this study, and also the ones that should be avoided due to the issues of quality.

Methods

A computerised literature search was performed using Medline, and Web of Science databases. Additionally, key papers were searched for further references. As shown in Table 3 below, the literature searches were conducted at the time of developing the MALT model, (September 2011), and they were updated at the time this thesis was written (March 2015). Search strategies are presented in full in Appendix A.3.

Table 3: Databases searched in the review of economic models

Database	Coverage	Date searched
Medline	1948-present	September 2011
Web of Science	1980-present	September 2011

Inclusion criteria:

1. Modelling studies on cost-effectiveness of COPD interventions

Results

The literature search was performed using RefWorks 2.0. A total of 6349 articles were retrieved. After exclusion of duplicates, 1562 titles and abstracts were screened for relevance. 40 studies were retrieved for full texts assessment. 16 studies were included in the final review. The process is illustrated in PRISMA diagram in Figure 10 below.

Figure 10. Prisma diagram COPD modelling studies
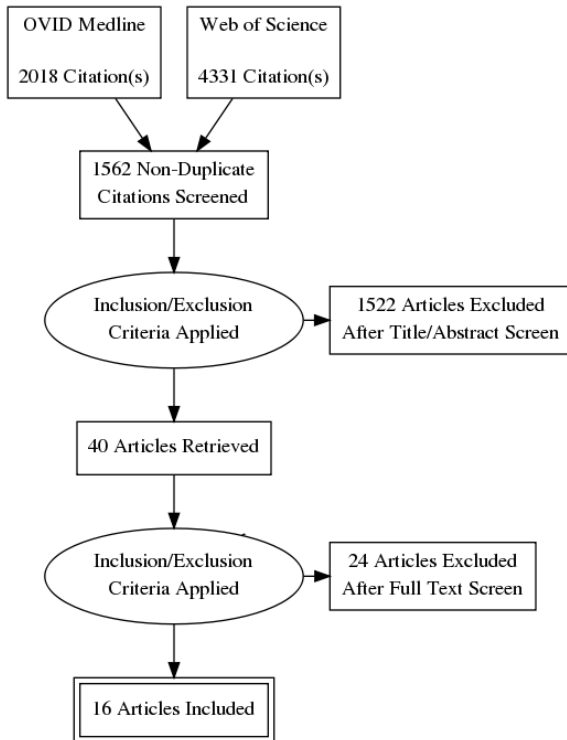
Diagram was generated using Toronto Health Economics and Technology Assessment Collaborative tool

http://prisma.thetacollaborative.ca/generator

The quality of the studies was assessed using the quality checklist for economic studies (Eddy, 1985), however, the focus of the review was to identify alternative model structures and their relative merits. At a later stage, it was recognised that the studies could also provide some model parameters that were not available from elsewhere (e.g., utilities) and so these were also extracted from the papers.

The review identified 14 Markov models and one patient-level simulation model. The technical details of Markov and DES modelling were discussed in Chapter 1.

The review identified several types of model structures for COPD interventions. The first type of model structures was based around the decline of lung function measured as FEV_1 percentage of predicted normal (PN). For example, Sin *et al*. (2004) used that classification to develop a three-state model (mild, moderate and severe COPD states) for an evaluation of inhaled corticosteroids. They assumed the decline of lung function of 47ml per patient per year in each of the severity stages.

Oostenbrink *et al*.(2005) also developed their model based on lung function decline. They used the Global Initiative for Chronic Lung Disease (GOLD) criteria to categorise patients into severity states of moderate, severe, and very severe COPD. Moderate COPD state was defined as FEV_1 greater than 50 %, and smaller than 80% PN; severe COPD state was defined as between 30 %, and less than 50% PN; and very severe COPD was defined as less than 30% PN. Mild COPD state and death state were not included in this model. The time horizon of the model was limited to 1 year, therefore the death state was not included in the model. Furthermore, patients were assigned a probability of experiencing severe or non-severe exacerbations during their transition between disease states. Also, backwards transitions were not allowed. The structure of the model is illustrated in Figure 11 below.

Figure 11: The structure of the Oostenbrink *et al.* economic model (2005)
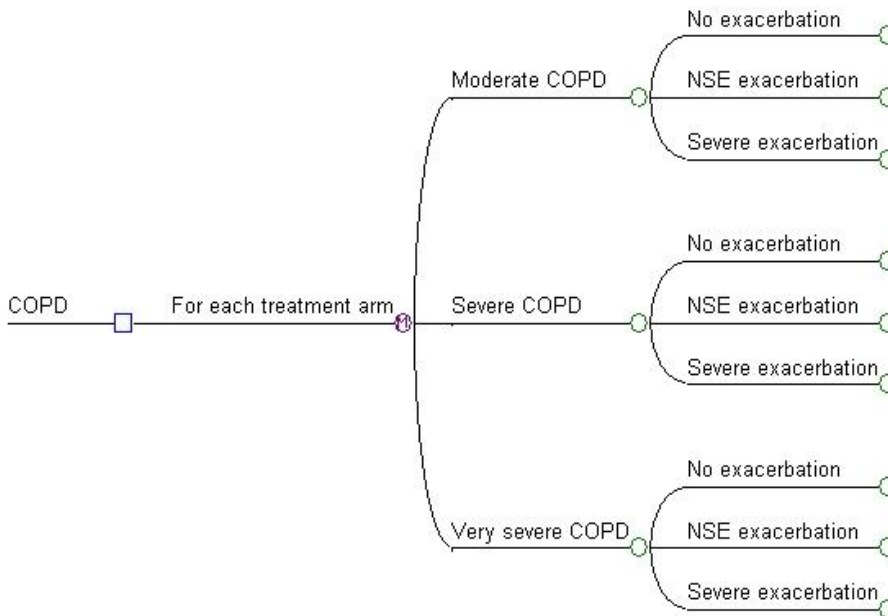
Oostenbrink *et al*. (2005) used two sets of transition probabilities, to reflect on the impact of treatment strategies (tiotropium, salmeterol and ipratropium). The utilities per disease state were derived from an observational study using EQ5D scores. For cycles where exacerbations occurred, utilities were reduced.

This is the most often replicated structure in the literature, but variations exist regarding the number and type of severity stages included. For example, Maniadakis (2006) adopted it to Greek settings, Rutten-van Molken (2006) to Spanish healthcare system and Gani (2010) to the UK settings. Within these, Rutten-van Molken expanded the time frame to five years, and added a 'dead' state in her model, while Gani added the possibility of transitioning to treatment states 'treated in primary care'; 'treated in secondary care' and 'untreated' following each of the types of exacerbations.

An example of a complex model is the one by Borg *et al.* (2004) who used a two-dimensional Markov model, first to assess the occurrence of exacerbations, and second, to model the decline in lung function. The model consisted of four health states categorised using GOLD criteria as: (1) mild, (2) moderate, and (3) severe COPD. The transition between states was based on the decline of FEV_1. It was assumed that in each of the disease states, patients experienced mild, moderate, severe or no exacerbation (Borg *et al.*, 2004). Figure 12 and 13 show the structure of the model.

Figure 12: The structure of the Borg *et al.* model (2004, p. 155).
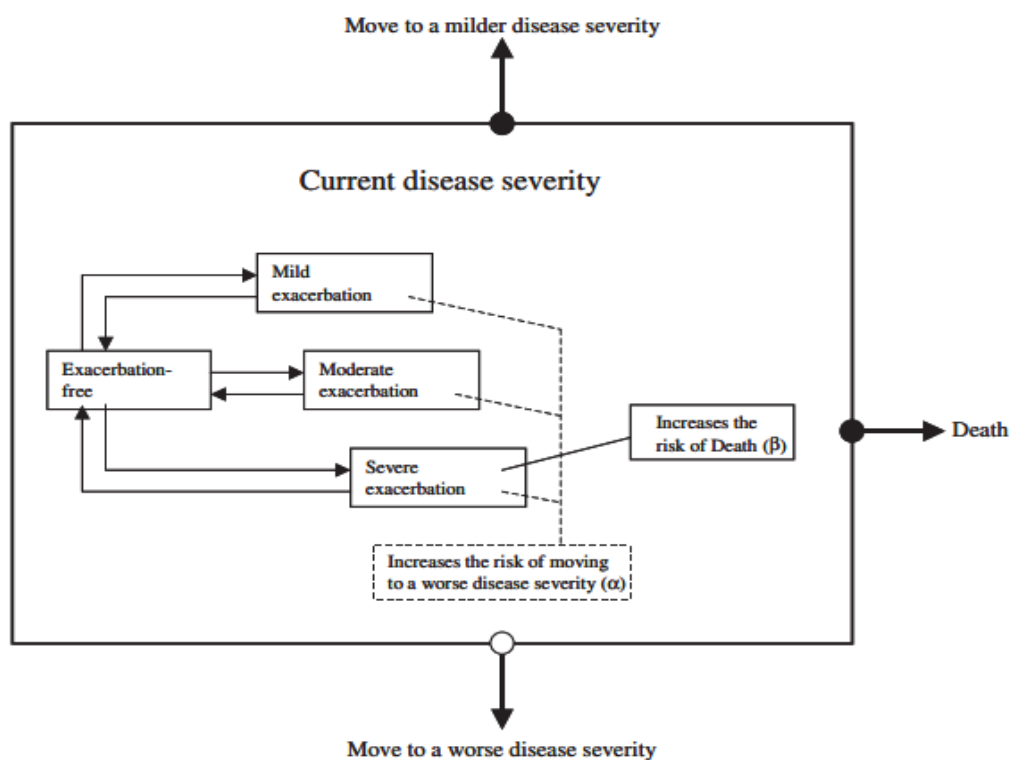
Figure 13: The second part of the Borg *et al.* model (2004, p. 155).
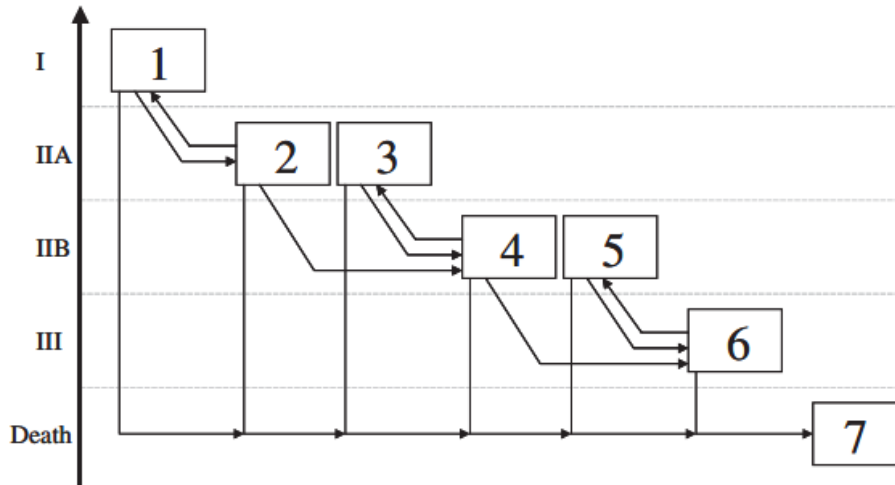
Within the model, it was assumed that patients could transition to previously occupied state, move to the next state, or remain in the current one or die. This is what differentiates this model from previously discussed studies, as backwards transitions were usually not allowed.

The second type of COPD model structures was based around exacerbations and resource use. A good example of such a model is Jubran's *et al* (1993) model, which was also the earliest Markov model developed for COPD interventions. It was developed to assess the cost-effectiveness of two treatment therapies: theophylline and ipratropium. The model had a horizon of one year and adopted a societal perspective. The model was structured around seven health states: (1) stable, (2) clinic visit, (3) consult, (4) ER visit, (5) hospital, (6) major toxicity, and, (7) minor toxicity. Moreover, it was assumed that each patient can experience one of these events: the exacerbation that would require hospitalisation, ER visit, unscheduled consultation or unscheduled clinic visit; toxicity; routine clinic visit, routine consultation or no event.

Oba *et al*. (2009) used an analogous structure in their model for assessing the cost-effectiveness of salmeterol, fluticasone and combination therapy for COPD. Patients' transitions were modelled depending on the severity of exacerbation (moderate, severe, very severe) and medical resources used as a result of an exacerbation (A&E visit, GP visit, nurse visit, hospitalisation). The model had four states: (1) no exacerbation, (2) non-severe exacerbation requiring GP or nurse visit, (3) severe exacerbation requiring hospitalisation and (4) death. The structure of the model is illustrated in Figure 14 below.

Figure 14: The structure of the model based on exacerbations and resource use by (Oba, 2009).
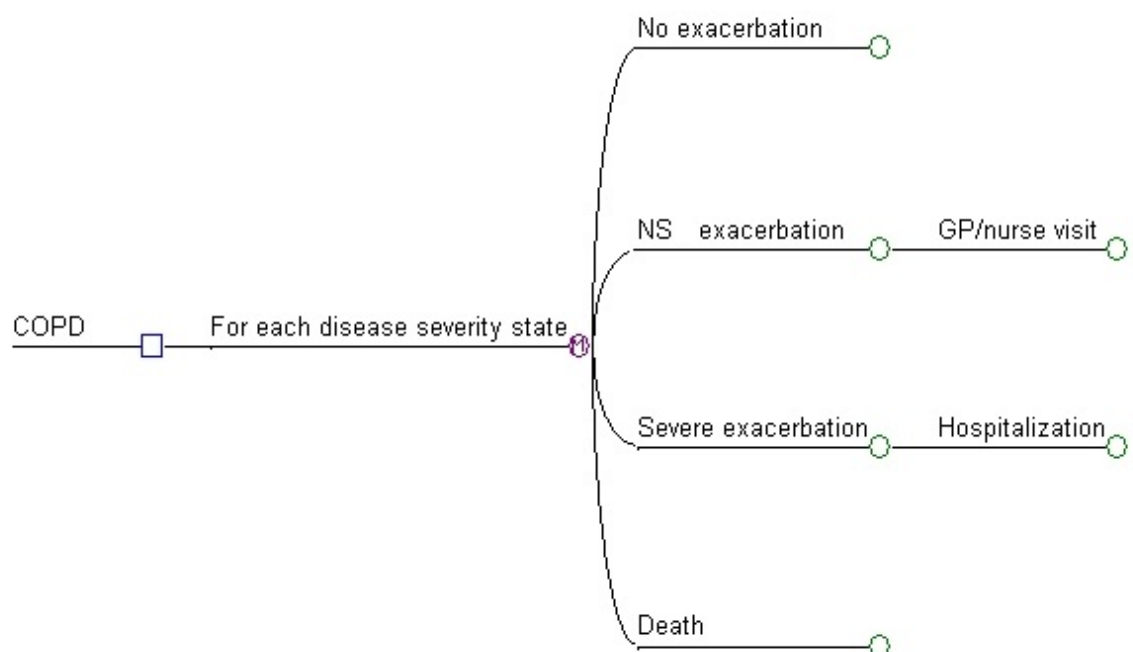
A similar model is Naik's (2010) two-state model for assessing treatment therapies for moderate COPD. The model used 'on treatment' and 'maintenance therapy' states with the assumption that patients 'on treatment' might either respond to the treatment or not, and that they would experience exacerbations. The assumption was that those who do not respond to the treatment would experience the side-effects or death. This was represented in the model by a chance node, as there was a probability of experiencing only one of these events.

A unique model in this review is Spencers' *et al*. (2005) simulation model. The model consists of four disease states: (1) mild COPD, (2) moderate COPD, (3) severe COPD and, (4) death. The model is presented in Figure 15 below.

Figure 15: The structure of Spencer *et al*. (2005) model
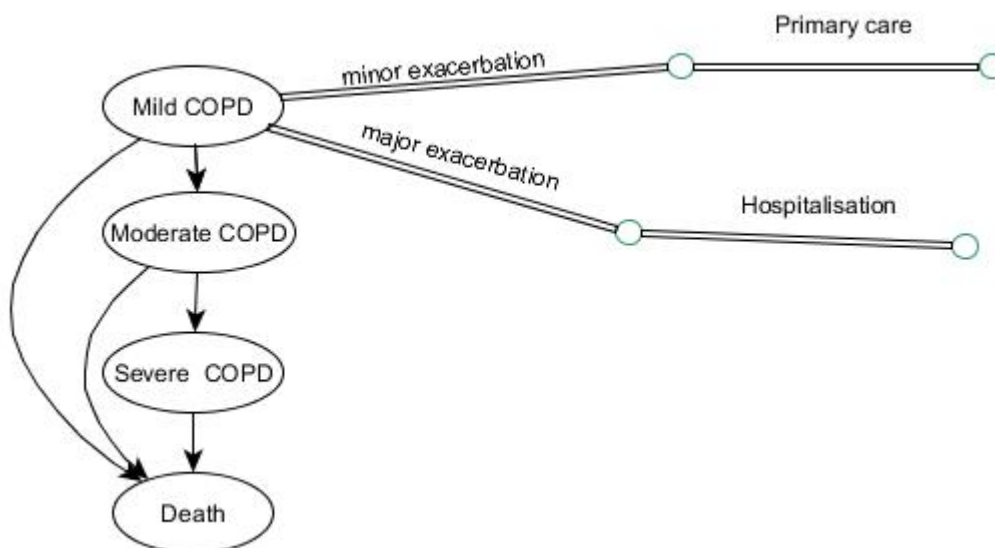
In this model transitions between states were determined using FEV_1 values. Exacerbations were considered either minor or major, by estimating their frequency for each of the disease states.

Additionally, a minor exacerbation was assumed to require a primary care visit, whereas the major exacerbation was assumed to require a hospitalisation. For a minor exacerbation, utilities were assumed to drop 0.61 in a mild state, 0.61 in a moderate state, and 0.05 in a severe state. In case of severe exacerbations, this fall was assumed as 0.26 regardless of a severity state. The comparison of the features of all identified models, based around the number of health states is shown in Table 4 on the next page.

							Utility weights by disease state				Exacerbation
Authors	**No of health states**	**Type of health states**	**Time horizon**	**Length of the cycle**	**Sensitivity analysis**	**Outcome measure**	**Mild**	**Moder**	**Severe**	**Very**	
(Naik *et al.*, 2010)	2	'on treatment'; 'maintenance therapy'	1 year	6 months	Deterministic: one way	Cost per exacerbation avoided per patient per year	-	-	-	-	-
(Sin, 2004)	3	Mild: $FEV_1 > 50\%$ Moderate: $FEV_1\ 35\% < 50\%$ Severe: $FEV_1 < 35\%$	3yrs	3months	Probabilistic	Quality-adjusted life expectancy	-	1.00	0.92	0.84	0.32 reduction for each state

Table 4: The studies included in the review of modelling studies

(Oostenbrink *et al*., 2005)	3	Moderate: 50 %<FEV$_1$<80% Severe: 30 %< FEV$_1$<50% Very severe FEV$_1$<30%	1year	1st cycle 8 days, subsequent 1 month	Probabilistic	Quality adjusted months	-	0.76	0.75	0.55	Reduction by 15% for non-severe exacerbation and 50% for a severe exacerbation
(Maniadakis *et al*., 2006)	3	Moderate: 50%<FEV$_1$<80% Severe: 30%<FEV$_1$>50% Very severe: FEV$_1$<30%	1 year	1st cycle 8 days, subsequent 1 month	Deterministic: one way	Quality adjusted months	-	0.76	0.75	0.55	Reduction by 15% for non-severe exacerbation and 50% for a severe exacerbation

(Chuck *et al.*, 2008)	3	COPD Stage 1: FEV$_1$ of 50% > PN, stage 2 FEV$_1$ 35%-49.9% PN, stage 3 FEV$_1$ < 35% PN.	Lifetime	3 months	Probabilistic	HRQL	-	-	-	-	-
(Gani *et al.*, 2010)	3	Mild COPD: FEV$_1$ 50%-80% Moderate COPD: FEV$_1$ 30%-49% Severe COPD: FEV$_1$<30% PN	1 year	1st cycle 8 days, 2nd 22 days, subsequent cycles 1 month	Probabilistic	QALY	0.787	0.750	0.647	-	Reduction by 15% for non –severe exacerbation and 50% for severe exacerbation

(Earnshaw *et al.*, 2009)	4	Moderate: FEV$_1$ < 50% Severe: FEV$_1$<30% <50% Very severe: FEV$_1$<30%	Lifetime	1 year	Probabilistic	QALY	-	0.755	0.748	0.549	Reduction by 15% for non-severe exacerbation and 50% for a severe exacerbation
(Spencer *et al.*, 2005)	4	Mild: FEV$_1$ ≥ 50%, Moderate: FEV$_1$ 35%-49%, Severe: FEV$_1$<35%.	25 years	3months	Probabilistic	Cost, number of exacerbations, survival, QALYs, cost-effectiveness	0.81	0.72	0.67	-	After minor exacerbation reduction to 0.61 for both mild and moderate states and 0.05 for severe state. After major exacerbation reduction to 0.26 for all states

(Hoogendoorn *et al.*, 2005)	4	Mild: FEV$_1$ ≥ 80%, Moderate: FEV$_1$ <80 % and ≥ 50%, Severe: FEV$_1$ < 50% and ≥ 30% Very severe COPD < 30%	lifetime	-	Deterministic: One-way sensitivity analysis	-	-	-	-	-	-	-
(Rutten-van Mölken *et al.*, 2007)	4	Moderate: FEV$_1$ <80 % and ≥ 50%, Severe: FEV$_1$ < 50% and ≥ 30% Very severe COPD < 30%	5 years	1year	Probabilistic	QALY, exacerbation-free months, mean cost	-	0.809	0.762	0.655	Reduced by 15% in case of non-severe exacerbation and 50% for severe exacerbation

(Zaniolo, 2010)	4	Moderate: FEV$_1$ <80 % and ≥ 50%, Severe: FEV$_1$ < 50% and ≥ 30% Very severe COPD < 30%	lifetime	1year	Probabilistic	Life years gained, QALY, number of exacerbations	-	-	-	-	-
(Oba, 2009)	4	Stable; Exacerbation requiring a physician visit; severe exacerbation requiring hospitalisation; death	3 years	3months	Deterministic: One - way	Hospitalisation, QALY	-	-	-	-	-

(Borg *et al.*, 2004)	4	stage I FEV$_1$ >80% PN, stage IIa 50% <FEV$_1$<80% stage IIb 30 %<FEV$_1$<50% stage III FEV1 <30%	30years	1 year	Deterministic	-	0.897	0.755 0.74* (state IIB)	0.549	-	Reduction: 5% mild, 15% moderate, 70% severe exacerbation
(Dal Negro *et al.*, 2007)	5	Mild, Moderate, Severe, Very severe, Death (GOLD 2004)	1,5,10 years and lifetime cohort simulation	1 year	Deterministic: one way	Number of exacerbations, symptoms free days, costs	-	-	-	-	-
(Jubran *et al.*, 1993)	7	Stable; Clinic Visit; Consultation; ER Visit; Hospitalisation; Major Toxicity; Minor Toxicity		1 year	Deterministic: One - way	Complication-free therapy month	-	-	-	-	-

BOLD Model (Buist *et al.*, 2005; Nielsen *et al.*, 2009; Rutten-van Molken & Lee, 2006)	10	Smoker, Never Smoker, Former Smoker, Non-smoker COPD stage I, Non-smoker COPD stage II; Non-smoker COPD Stage III; Smoker COPD Stage I; Smoker COPD Stage II; Smoker COPD Stage III/IV; and Death	20 years	-	Probabilistic	-	-	-	-	-	-	-

(-) Not included in the model

Comparison of alternative model structures

The approach to modelling in published cost-effectiveness studies varied in terms of the complexity of the modelling approaches and the underlying assumptions. With exception of three studies, (Borg *et al*., 2004; Buist *et al*., 2005; Hoogendoorn *et al*., 2005), the majority of the models adopted simplified assumptions pertaining to patients' transition between health states and efficacy of treatment.

The choice of the perspective varied between the studies, with the health systems' perspective being the most commonly used (Chuck *et al*., 2008; Gani *et al*., 2010; Hoogendoorn *et al*., 2005; Maniadakis *et al*., 2006; Oostenbrink *et al*., 2005; Rutten-van Mölken *et al*., 2007; Zaniolo, 2010). Two models used a third party perspective (Naik *et al*., 2010; Oba, 2009), and three used the societal perspective (Borg *et al*., 2004; Jubran *et al*., 1993; Sin, 2004). The study by Borg (2004) investigated the cost-effectiveness of hypothetical treatment strategies from both societal and healthcare provider's perspective. Lifetime horizon was applied only to two studies (Chuck *et al*., 2008; Hoogendoorn *et al*., 2005), whereas a one-year horizon was used in four evaluations (Gani *et al*., 2010; Naik *et al*., 2010; Oostenbrink *et al*., 2005; Zaniolo, 2010). Time horizon in other studies varied, for example; Sin (2004) and Oba (2009) used a three years' horizon, whereas Rutten-van Molken (2007) used the five years. There were two models that had longer time horizons: Buist (2005) of 20 years, and Spencer (2005) of 25 years.

Several outcome measures were used in these studies, such as the quality-adjusted months and years, cost per exacerbation avoided per patient per year or the cost of exacerbations. As reported by Starkie (2008), discrepancies can be found in the utility weights values used in these studies. Moderate COPD health state had assigned the values: 1.0 in Sin's (2004) model, 0.76 in Oostenbrink's (2005) and 0.72 in Spencer's (2005) model. Also the utility weight decline after

the exacerbation varied in these studies. In Oostenbrink's (2005) model utilities were reduced by 0.27, in Sin's (2004) by 0.52, and in Spencer's (2005) by 0.26. The majority of the models used constant transitions over time, with an exception of Oostenbrink *et al*. (2005) who used two sets of transition probabilities for each of the treatment choices and Spencer *et al*. (2005), who used different transition probabilities to account the effects of smoking.

Furthermore the approach to exacerbations varied considerably. Worth noting is the fact that up to date, there is no consensus on the definition of an exacerbation, so each author adopted his own. For example, Oostenbrink *et al*. (2005) based their understanding of the exacerbation severity on physicians' assessments and treated as an exacerbation each worsening of respiratory symptoms lasting for at least three days. Using this analogy, a non-severe exacerbation was defined as 'an awareness of a sign or symptom that was easily tolerated', or as 'discomfort enough to cause interference with usual activity'. A severe exacerbation was defined as 'incapacitating or inability to do work or usual activity'. In contrast, Sin (2004) classified exacerbations based on the type of care needed to improve the symptoms. Mild exacerbation was defined as 'worsening of symptoms requiring outpatient physician services and exacerbation therapy'; moderate exacerbation was defined as 'clinical episodes requiring emergency department utilization or urgent physician office visits (and institution of exacerbation therapy)'; and severe exacerbation was defined as 'symptoms requiring in-patient care (and institution of exacerbation therapy)'. A similar approach was adopted by Spencer (2005), who considered exacerbations as minor when a patient required primary care treatment and as a major when a hospitalisation was required. Exacerbations in Sin's (2004), model had an impact on the health outcome of the model (HRQL). In contrast, Oostenbrink *et al*. (2005), assigned a probability of exacerbation to each patient depending on treatment group and disease state the patient was in. Oba (2009), used exacerbations to define health states; (exacerbation requiring physician visit and severe exacerbation requiring hospitalisation) and Borg (2004), included exacerbations in a Markov chain (mild, moderate,

severe exacerbations and exacerbation-free states). Exacerbations were not taken into account in Hoogendorn's *et al*. (2005) model.

The impact of smoking on COPD disease progression had been demonstrated in various studies (Celli *et al*., 2008; Han, 2010; Scanlon, 2000). Spencer *et al*. (2005), accounted for smoking by developing separate transition probabilities for smokers and ex-smokers, with different FEV_1 decline rates (62 and 31ml/year) . Another method for incorporating smoking status was using it as a cohort characteristic (Sin, 2004; Zaniolo, 2010). A unique approach was presented in the Burden of obstructive lung disease (BOLD) model, where smoking status was incorporated into the model as a health state (Buist *et al*., 2005). For example, apart from 'smoker', 'never smoker', and 'former smoker' health states, the smoking status was incorporated into each of the COPD severity states as 'non-smoker COPD stage 1' or 'smoker COPD stage 1' etc. Other economic models did not take smoking into consideration because of the lack of sufficient data.

Importantly, COPD pharmacotherapy does not affect the decline of lung functions and is only considered to control symptoms of the disease and severity of exacerbations (Nishimura & Tsukino, 2000). However, the choice of pharmacotherapy affected the way the effects of treatment therapies were modelled. In five models pharmacotherapy directly affected transition probabilities and thus the mortality and hospitalisation rates (Gani *et al*., 2010; Maniadakis *et al*., 2006; Naik *et al*., 2010; Oostenbrink *et al*., 2005; Rutten-van Mölken *et al*., 2007). In four models, the treatment had a direct impact on mortality (Dal Negro *et al*., 2007; Hoogendoorn *et al*., 2005; Oba, 2009; Zaniolo, 2010). The model by Spencer *et al*. (2005) is specific in this respect as the effect of treatment was considered to affect the risk of exacerbations, disease progression, mortality and health status.

One of the main features of Markov modelling is the inclusion of the 'death' state. In the current review only seven models incorporated death state (Borg *et al*., 2004; Buist *et al*., 2005; Dal Negro *et al*., 2007; Oba, 2009; Rutten-van Mölken *et al*., 2007; Spencer *et al*., 2005; Zaniolo, 2010). The reason for not including death in the remaining models was the small numbers of deaths observed in studies used to populate them.

Six models (Borg *et al*., 2004; Chuck *et al*., 2008; Earnshaw *et al*., 2009; Hoogendoorn *et al*., 2005; Oostenbrink *et al*., 2005; Sin, 2004) based their transition probabilities on data from The Lung Health Study (Anthonisen, 1994). The remaining models used data from RCTs (Borg *et al*., 2004; Gani *et al*., 2010; Oba, 2009; Oostenbrink *et al*., 2005; Rutten-van Mölken *et al*., 2007; Zaniolo, 2010), observational studies (Maniadakis *et al*., 2006), the Third National Health and Nutritional Examination Survey (Sin, 2004), and the National Health System Data Registry (Spencer *et al*., 2005).

The majority of the models fulfilled the main categories of the checklist (Eddy, 1985), pertaining to the scope and overview, data sources, base case results and sensitivity analyses results, therefore could be perceived as high quality. Several items were not discussed in these studies: the need for modelling versus alternative methodologies, impact of modelling assumptions on the results; applicability and limited ability of the results, and description of research in progress that could alter the results of the analysis.

One of the requirements of the modelling studies is also to perform a sensitivity analysis (Claxton *et al*., 2005). Probabilistic sensitivity analysis was used amongst 10 of the reviewed studies (Chuck *et al*., 2008; Earnshaw *et al*., 2009; Gani *et al*., 2010; Maniadakis *et al*., 2006; Naik *et al*., 2010; Oostenbrink *et al*., 2005; Rutten-van Mölken *et al*., 2007; Sin, 2004; Spencer *et al*., 2005;

Zaniolo, 2010). Two studies used deterministic one-way sensitivity analysis (Hoogendoorn *et al*., 2005; Oba, 2009).

Conclusions from the review

The findings of this review are consistent with previous systematic reviews, which showed the dominance of FEV_1 based models (D'Souza, 2006; Starkie *et al*., 2008). The review deepened my understanding of issues related to incorporating the natural history of the disease in the models and the methodological problems other authors encountered.

I understood that as the aim of any treatment for COPD, in the absence of the cure is to '*prevent and control symptoms, reduce frequency and severity of exacerbations and improve health status*' (Rabe *et al*., 2007), these should be included in the model, because of their impact on costs and effects. As seen in the review, most of the published models have used lung function as an indicator of progression through the model, with exacerbations occurring in each health state and its frequency worsening as disease progresses. There was however a problem with defining exacerbations and gathering the data on its occurrence and different authors adopted different approaches. In the reviewed models, symptoms have not been incorporated despite their effect on patients' quality of life.

After considerations, based on the review I identified four types of structures that could be replicated in this study: based on FEV_1, and based on exacerbations and resource use. These model structures are discussed in section 2.2.5 vi.

2.2.5 Conclusions from the framework

iv) Choosing specific model interventions

Based on the results of the literature reviews we decided that the effect of medical therapies and smoking were not to be included in the MALT model. We also discussed several additional factors that potentially have to be considered in an economic model for COPD interventions. These included the question of what level of detail on patient characteristics should be provided in the model and the choice of clinical outcomes.

v) Determining model boundary

The next stage in the framework required that the 'boundary' of the model, understood as a mark between what would be included in the model and what would be regarded as part of external environment was determined. Within the discussion of modelling team, we agreed that these would be as follows:

Population: The model would follow a hypothetical cohort of COPD patients, who have been offered home telehealth either on discharge from the hospital or as a part of their COPD management programme.

Intervention: Telehealth for self- monitoring as described in section B.2.2.2.

Comparator: The comparator would be 'usual care' in a way it is described in the care pathway diagram in Fig 10 and 11.

Time horizon: The model would use 40 years' horizon as this was expected to capture the full lifetime costs of patients with COPD.

Perspective: The perspective of the model would be from the NHS and Personal Social Services perspective, but costs would be disaggregated to individual organizations as the impact of telehealth will vary, which can generate different incentives across the stakeholders. For example, hospital costs are expected to reduce, while community care costs may increase or decrease depending on the cost of the telehealth technology and the structure of the new service.

vi) Determining the level of detail

Once the boundary of the model was determined, we started working towards determining the model structure based on the previously gathered information. As described in Sec C we identified that the model could be structured in four ways: (1) based on FEV_1; (2) based on the number of hospitalisations; (3) based on the resource use, and (4) based on severity of exacerbations.

In case of a model based around FEV_1, it would have been structured with health states ranging from moderate COPD to very severe COPD, including severe and non-severe exacerbations. In that sense, it would be a replica of Oostenbrink's (2005) model. The structure of the model is presented in Figure 16 below.

Figure 16: The first model structure considered
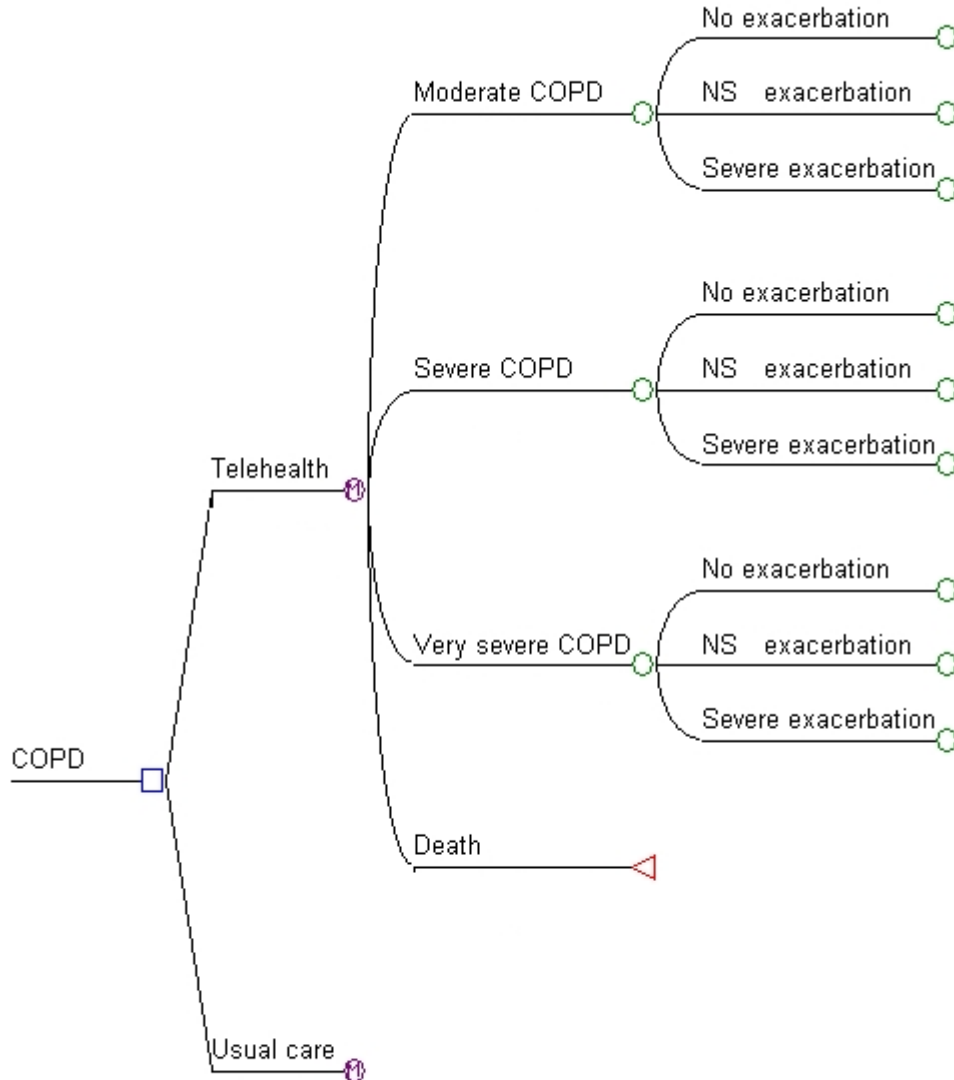
In this model, exacerbations were understood as *'as a complex of respiratory symptoms such as cough, sputum, dyspnoea, or wheeze; lasting for at least 3 days'* (Oostenbrink *et al*., 2005, p. 34). Because in the original paper additional criteria were used to distinguish between mild and moderate exacerbations as non-severe exacerbations, I developed my own definition based on

(Rodriguez-Roisin, 2000). For the purposes of this model non-severe exacerbations (NS) were understood as '*the worsening of the symptoms that can be managed with medication*'.

Having identified an FEV_1 based model as the most common in the literature, further meetings with consultant physicians were held to establish what model structure would be most useful from their perspective. Prof Parker pointed out that FEV_1 as a measure of severity in COPD is clinically viable, but the availability of the data would need to be verified before pursuing with such a model. That led me to a meeting with another Consultant physician (MK) to verify current practice concerning collecting FEV_1 data. In this particular NHS Trust, FEV_1 data was routinely collected in all GP surgeries every 15 months with the maximum interval of 2 years. GP were collecting FEV_1 data because of requirements of the Quality Outcome Framework (QoF). However, this was seen as a constraint from the perspective of the MALT project, as the study did not have access permission to primary care data.

Another important argument came from recent COPD clinical guidelines and suggested that there is an ongoing debate on optimal staging system for COPD. It is argued that a multidimensional system that takes into account body mass index, airflow obstruction, dyspnoea and exercise capacity is more accurate in categorising patients and in predicting death and hospitalisations than FEV_1 alone (Celli *et al.*, 2004). As such, FEV_1 is not ideal and not the only way of categorising severity of the disease.

The team also agreed that this structure would not capture the effect of telehealth very well, as telehealth is usually prescribed to patients in severe or very severe COPD, as these patients seem to benefit the most from technology. Consequently, the decision was undertaken to ultimately rule out this structure from further considerations.

The second considered model, as presented in Figure 17, had health states based on hospitalisations. One of the issues discussed with the team, was that a hospitalisation is not an ideal health state in the way health economics understands it. For example, 'moderate COPD' is better reflecting a state patient is in and transition required to leave it. The main disadvantage of this structure was also seen in the fact that it would not be able to capture adequately the severity of the disease.

Figure 17: The second model structure considered
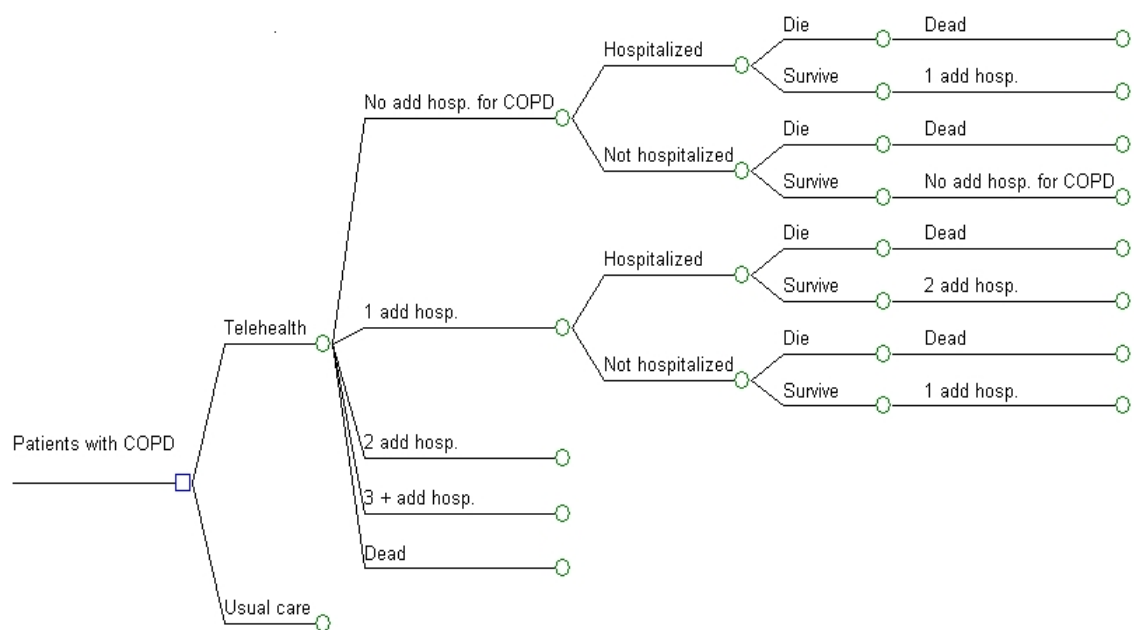
A model based on hospitalisations was perceived as viable, as this is commonly used outcome measure in COPD studies that have a direct impact on costs of the healthcare system. However, for individual patients, healthcare use does not necessarily reflect the severity of the condition, nor the prognosis of the disease. As previously mentioned, patients with very severe COPD tend to have more hospitalisations and GP visits, but they can avoid hospitalisations by mastering self-

management. Moreover, there are further factors that determine hospitalisations. Firstly, patients with several co-morbidities are more likely to use health care resources irrespective of the severity of their COPD. Also, the decision to hospitalise is often based on the assessment of patients' social support. A patient with a long-term condition, who lives alone, is more likely to be admitted and stay longer in hospital irrespective of his chest condition if he does not have the confidence to go back home. Also, anxious patients even with mild COPD tend to visit GP and A&E departments frequently, so their psychological status plays a role.

The third considered model (Figure 18) would be structured around resource use with health states: no exacerbation, mild exacerbation requiring GP or nurse visit, severe exacerbation requiring hospitalisation and death. The model would be a replica of Oba's model and a more detailed version of model no 2. The important feature of this model was seen in the fact that resource use defined the categorization of exacerbations. This was seen useful from the perspective of financial modelling.

Figure 18: The third model structure considered
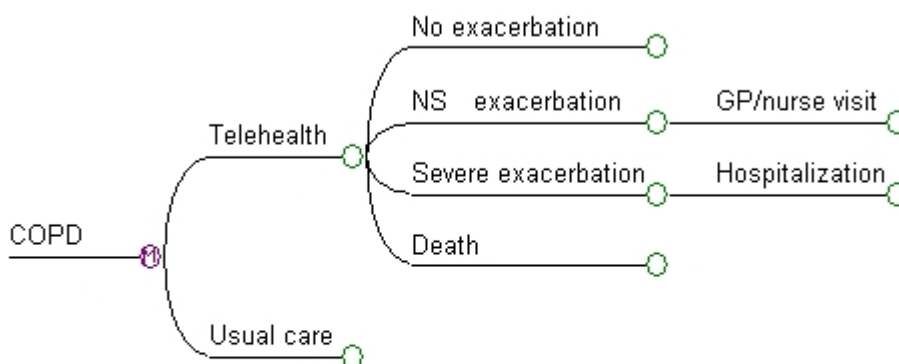
Two constraints for using this structure were identified. Firstly, there was limited access to data of sufficient quality in the study sites. Secondly, I identified variations of the way the care pathways were depicted within the NHS. This was supported by the literature, that shows that not all severe exacerbations lead to hospitalisations, and not all non-severe exacerbations are treated in Primary Care (Bach *et al*., 2001; Rodriguez-Roisin, 2006).

The fourth considered model was exacerbation based (Figure 19). The model would reflect the pathway and resource use similarly to model based on exacerbations. The only constraint was seen in developing appropriate transition probabilities. The advantage of the model was seen in the fact that the severity of exacerbations was defined clinically.

Figure 19: The fourth model structure considered
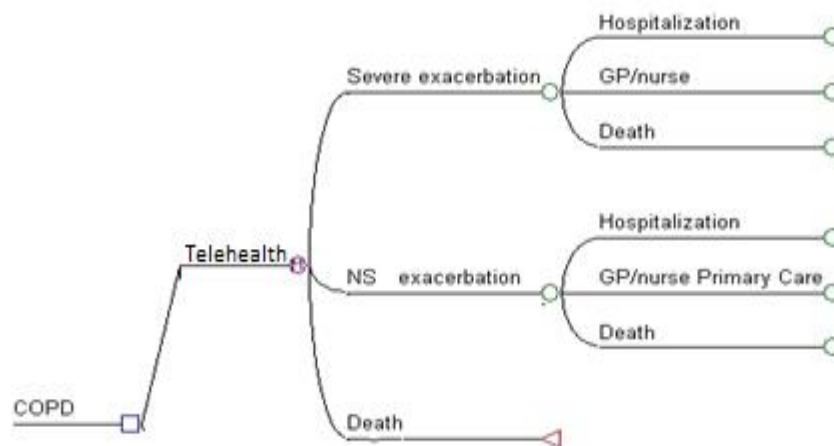
This model was seen as acceptable as well, as long as a definition and differentiation between non-severe and severe exacerbations would be clearly defined. From the GP point of view, an exacerbation needing hospitalisation may be viewed as a 'severe' exacerbation but as it was pointed out earlier, the decision to hospitalise is complex. In my discussions with clinicians it was even revealed that sometimes in usual practice yet another distinction is used. Exacerbations are defined as either 'infective' or 'non-infective', depending on whether an infection is present or not.

Most importantly, I learned that from the perspective of clinicians, changes of symptoms are more significant to capture than changes in FEV_1. This finding had the implications for the design of the MALT model, as that would mean that all variables affecting symptoms should be considered. One of such is smoking, which is known to impact the deterioration of health of COPD sufferers.

At this point in the project, the choice of the model structure remained an unresolved issue. The project team agreed that the decision would be postponed until there was information on the availability of local data on risk stratification, hospitalisations and care pathways that would be obtained through qualitative interviews with the members of the staff in each of the study sites. (The results of these interviews are described in Chapter 6).

vii) Selecting the type of the model

Using the taxonomy of model structures (Brennan *et al*., 2006), it was decided that a cohort model would be more appropriate than an individual level model. This was mainly determined by lack of data to populate individual level model and project timescales. Also, in the above-mentioned literature review, the majority of the models were Markov models, with only one published patient level simulation model (Spencer *et al*., 2005) for COPD interventions.

3. Theory review

In this chapter, I describe the results of a theoretical review undertaken to understand mechanisms of technology adoption, use and user satisfaction. This work establishes a link between the model specification developed via the conceptual model, the actual use of the model by end users and their satisfaction with the model in their own decision making context.

This review was seen as complex because it was necessary to refer to the literature of a range of disciplines: information systems research, behavioural sciences, psychology, decision-making sciences and management. It was considered that all of these disciplines may present useful approaches to guide the choice of methods and measures for the evaluation of a proposed model.

The aims of the review were to identify theories relating to the phenomenon of technology acceptance, use and user satisfaction; and to define precedents of usefulness.

3.1 Methods

There are substantial variations in the methodologies of reviews that consider theory. I followed Campbell's approach (2014), who utilised traditional systematic review techniques, such as developing a comprehensive search strategy, performing quality appraisal and synthesis of findings. Given the multidisciplinary character of this work, I did not consider quality appraisal as viable but adopted an approach where studies were selected as viable based on how many times they were cited by others. My intention was to map the range of theories available on the subject of acceptance, use and satisfaction from IT, not to generate a unified theory.

An alternative approach, where theories are identified based on the review of the empirical studies on the subject also exists. However, exploring every study which was published on the subject in the IT domain for various types of technologies was perceived as not viable.

A computerised literature search was performed using electronic databases (PsycINFO, Scopus and Web of Science). The search was limited to the literature published in English between 1960-present. Additionally, to locate relevant studies, techniques such as citation searching in key paper references, hand searching in relevant journals and snowballing were used. The journals considered were: MIS Quarterly, Management Sciences, Journal of Management Information Systems, Information System Research and Decision Sciences.

Inclusion criteria:

- To be included the study had to describe theory. Papers outlining speculations or hypotheses were rejected.

- To be included theory had to define mechanism linking the use and the adoption of technology (either software, or a decision support system or a model) with user satisfaction.

Exclusion criteria:

- The studies on mobile applications and media technology, as they were regarded too distant from the focus of the study

The following combination of terms were used to initiate the search: 'use', 'acceptance', 'satisfaction' and 'information system', 'information technology', 'software', 'decision support system'. These were combined using Boolean 'AND' with the terms to identify theory: 'theory', 'model', 'mechanism' 'concept'. Because of the high number of results, they were further refined by limiting them to the literature published in the fields of information science, computer science,

health services research and Operations Research and management science; and published in English.

3.2 Results

Web of Science database search:

- The initial search produced 26000 results, which were then sorted by 'most cited' and 'recent' to limit the number of records,

- The Top 2000 citations were exported to reference management software and de-duplicated for sifting.

Scopus search:

- The search produced 514 results.

PsycINFO search:

- After the initial search, because of the high number of citations (6624), an additional set of key terms had to be added to restrict to the most relevant literature (exp websites/ or exp computer software/ exp information technology or exp computer attitudes AND theory OR model OR mechanism.mp AND acceptance OR satisfaction).

- The search produced 768 results.

The results of all searches produced a total of 3282 citations. After the exclusion of duplicates, 2125 studies were screened for relevance. The abstracts of 1472 studies were then screened to identify studies that met inclusion criteria. At this point, because of the high number of records, the data extraction form, similar to one used by Campbell, was used to categorise studies as shown in Figure 20 below.

Figure 20: Data extraction form

General details of the paper: author, title, discipline

Type of technology: software/ information system/ decision support system

Type of study: critical review, meta-analysis, empirical study which replicates and validates previous theory, empirical study which attempts to extend previous theory or empirical study which applies the theory in different setting

This was sufficient to identify main theories and exclude surrounding or irrelevant literature. The total number of articles retrieved for full-text assessment was 106. From these, nine theories were included in the final review. The process is illustrated in Prisma diagram (Figure 21 below).

Figure 21. Prisma diagram the review of theories
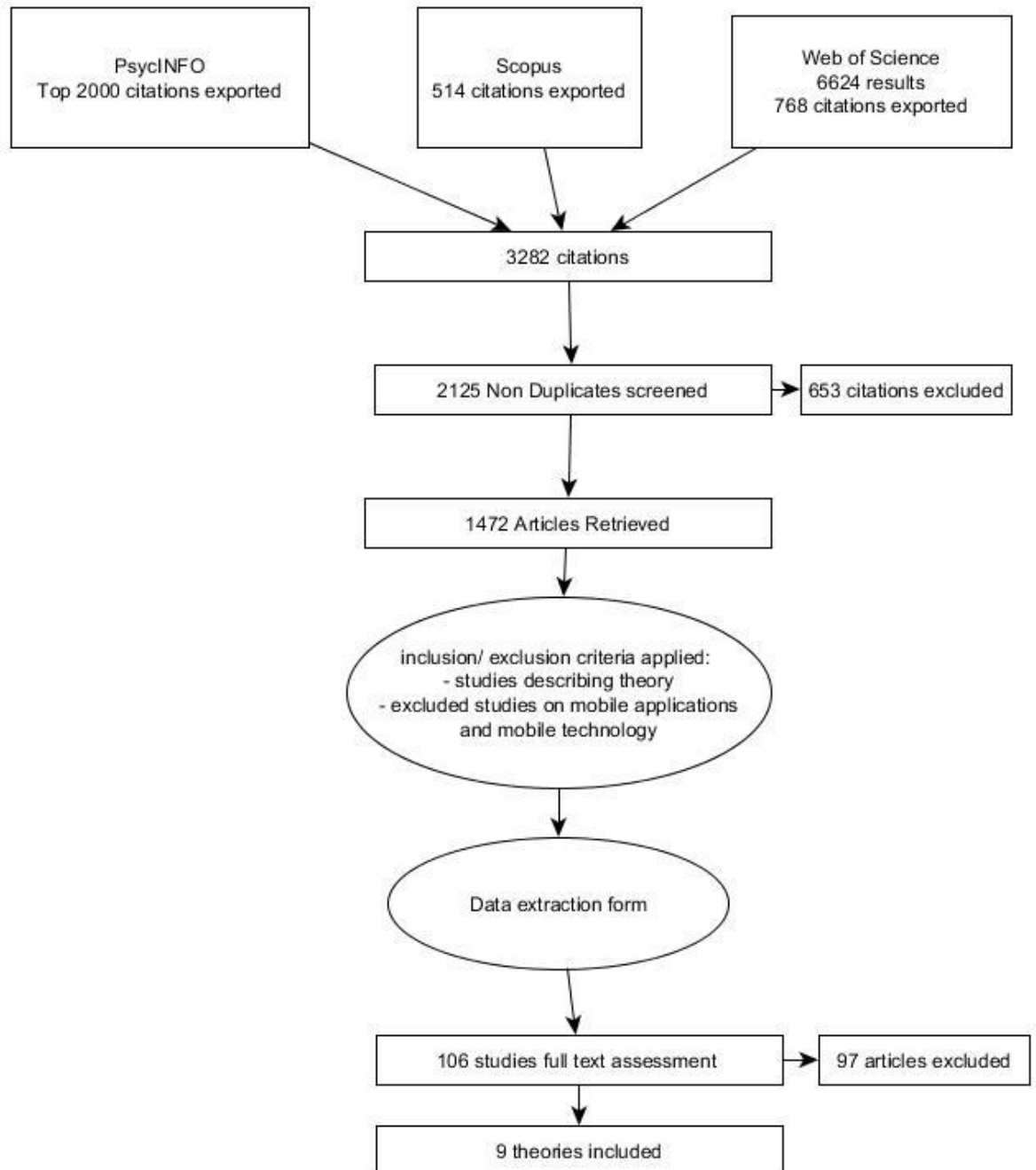
3.2.1 Characteristics of the studies

The first category of identified studies considered adoption and utilisation of technology. Developed in psychology and behavioural sciences, they were mostly used to explore barriers to technology acceptance or to predict use of technology, with the aim of preventing its abandonment or misuse before costly implementation processes take place. The following section presents a descriptive overview of these theories.

What are antecedents of adoption and use of technology?

Psychological theories seeking to explain and predict behaviour include the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) and the theory of planned behaviour (Azjen, 1985, 1991), which are general social behavioural theories, used to explain the relationship between attitudes and volitional behaviours. The theory of reasoned action defines persons' behaviours as determined by the intention to perform the behaviour, in contrast with the intention as a function of persons' attitudes towards the behaviour and subjective norms. The theory of planned behaviour also states that specific beliefs influence intentions and behaviour. In comparison to the theory of reasoned action, the theory of planned behaviour includes perceived behavioural control, which is the 'ability to act out of the behaviour' as a force that shapes individuals' intentions and behaviour.

In his doctoral research, Davis (1989) adapted the theory of reasoned action concept to study determinants of technology acceptance and concluded that two main forces drive the intentions of users: perceived ease of use and perceived usefulness. These affect attitudes toward using the technology and in turn actual behaviour. In other words, attitudes of a person who is not convinced of the usefulness of the technology and its ease of use; might impact his behaviour to the extent that he might not actually use it. These findings gave rise to the technology acceptance model ,

116

(illustrated in Fig. 22), which is now one of the most influential theories in the field, even reaching a status of a paradigm of sorts (Bagozzi, 2007).

Figure 22: The technology acceptance model (Davis, 1989, p. 987).
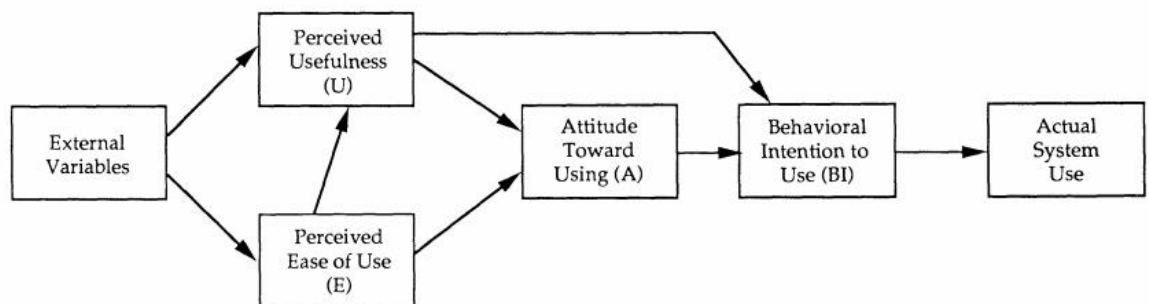
The technology acceptance model has been widely used in various disciplines, for example in the World Wide Web (Lederer, 2000), in healthcare (Holden & Karsh, 2010), online commerce (Koufaris, 2002) and web-based information systems (Mun, 2003). Two main published extensions of the model: the technology acceptance model 2 (Venkatesh, 2000), and the unified theory of acceptance and use of technology (Venkatesh, 2003), attempted to overcome shortcomings of the model by expanding determinants of perceived usefulness as illustrated in Figure 23 and Figure 24.

In the technology acceptance model 2, Venkatesh (2000) extended the original construct with social influence processes ('subjective norm') and cognitive processes. The term 'subjective norm' comes from the theory of reasoned action and is defined as *'person's perception that most people, who are important to him think he should or should not perform the behaviour'* (Fishbein & Ajzen, 1975, p. 302) In other words, the theory assumes that people might do something even if they do not favour the behaviour or its consequences, if they are motivated by people who think

they should. Subjective norm has been proven to directly impact a range of other factors, such as the image a person has of himself, and behavioural intentions to use. Additionally, the author considered the impact of cognitive processes such as the degree to which a system is applicable to one's job, the quality of the output it produces, and the demonstrability of the results, on perceived usefulness. For example, if the software or an IT system produces the results relevant to the task or character of the job, but in an obscure fashion, the user is less likely to understand how useful the system is.

Figure 23: The technology acceptance model 2 (Venkatesh, 2000, p. 188)
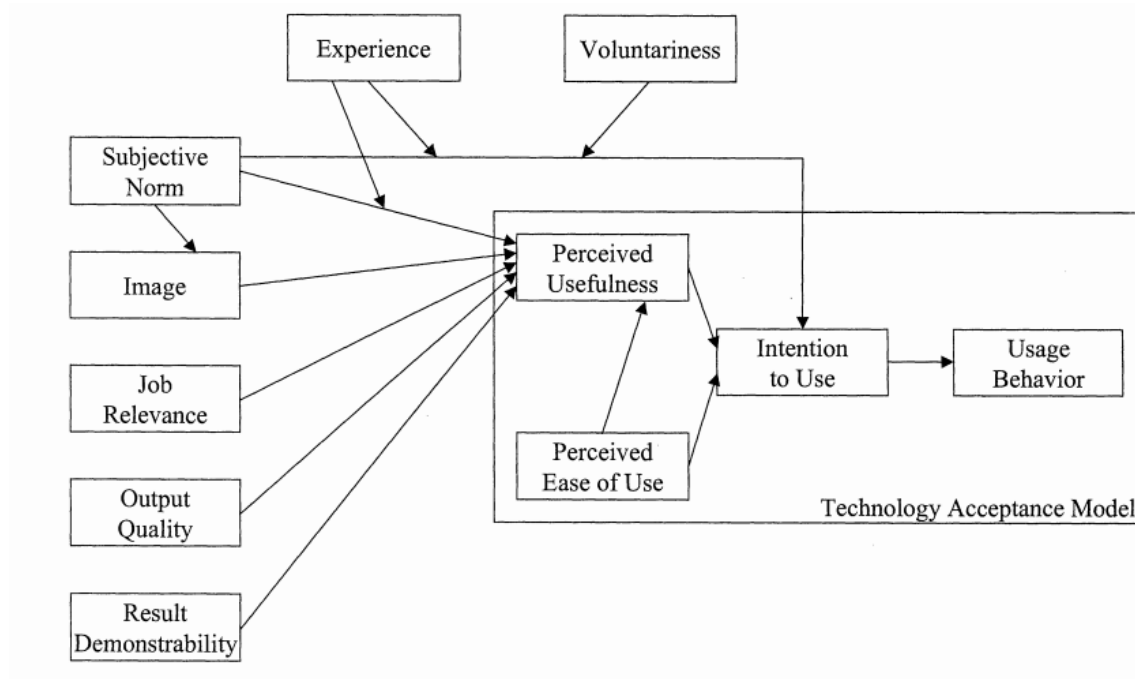
As the academic debate on the factors which influence an individual's behaviour continued, Venkatesh (2003) proposed a model which brought together the elements of eight models previously occupied with the concept of IT usage behaviour, in his unified theory of acceptance and use of technology, which explains determinants of intention and usage. The theory states that

performance expectancy, effort expectancy, social influence and facilitating conditions determine the intention and usage. Moreover, factors such as gender, age, experience and voluntariness of use have been demonstrated to impact on behavioural intention. The theory is illustrated in Figure 24.

Figure 24: The unified theory of acceptance and use of technology (Venkatesh, 2003, p. 447)
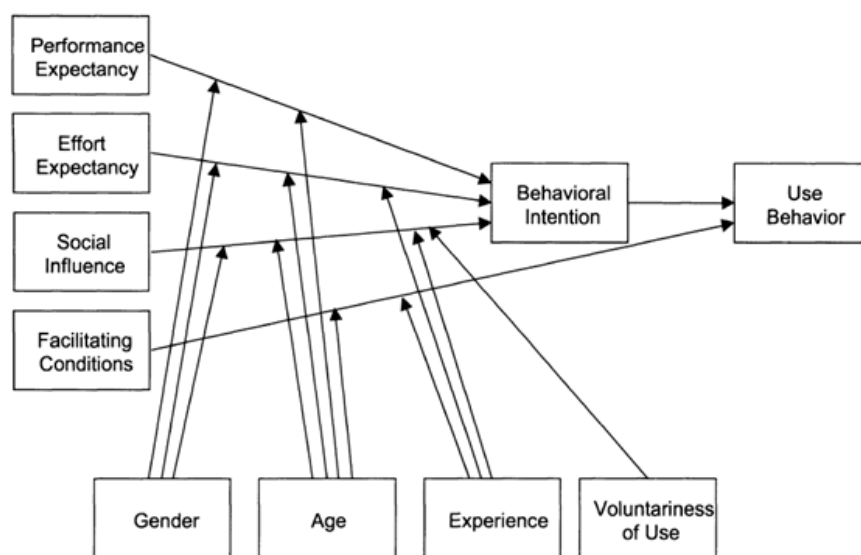
Goodhue and Thompson (1995) proposed the technology-to-performance chain model, that posits that in order for technology to be used, the capabilities of the technology have to match the demands of the task. Therefore, the evaluations that are solely based on measurement of use have several drawbacks, as in situations when use is voluntarily, the measure of use have little diagnostic value e.g. it is difficult to establish if greater use was triggered directly by systems' properties. Also, discovering that the performance is low require further actions to be undertaken to properly identify the problems with a particular technology.

The model has been applied in a variety of research studies in an online environment, very often combined with constructs from the technology acceptance model or other theories, e.g.((Gillenson & Sherrell, 2002), (Pagani, 2006)). Figure 25 illustrates the version of the model that has been empirically tested (Goodhue, 1998), with dotted lines representing the constructs where a moderate relationship exist.

Figure 25: Technology-to-performance chain model (Goodhue, 1998, p. 108)
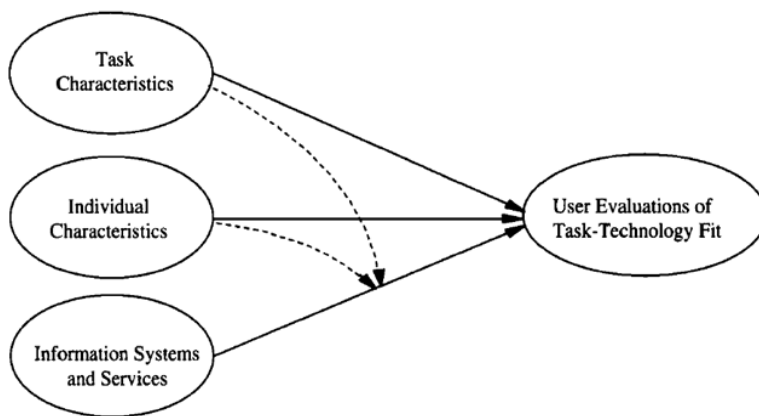
The users' perception of task-technology fit was found to be impacted by the task characteristics, individual characteristics and information systems and services (Goodhue, 1998).

What makes people continue to use technology?

While the significant number of theories identified in this review considered adoption of technology, two offered explanations of mechanisms which make people continue to use technology. Oliver (1980) studied users' satisfaction post-adoption and discovered that the extent to which a person is satisfied with a product in the longer term, can be explained using three constructs: expectations, perceived performance, and disconfirmation of beliefs. According to the

expectation confirmation theory, expectations influence both perceptions and disconfirmation of beliefs, which refers to the judgements an individual makes about the product. When the product outperforms original individual expectations the disconfirmation is positive, and satisfaction grows.

Rogers (2003) in his influential diffusion of innovation theory, argued that a spread of a new idea depends on the innovation itself, communication channels, time and the social system. He proposed that individuals adopt technology based on the perception about its complexity, compatibility with an existing system, and its observed effects. Rogers (2003) also studied the rate at which innovations are adopted and concluded that to be sustained, a technology has to reach a sufficient number of adopters.

How to define usefulness?

In the absence of the literature on 'usefulness' of technology, I paid attention to the concept of its 'success'. The information systems success model (DeLone & McLean, 1992), consist of six interdependent variables: (1) system quality, (2) information quality, (3) use, (4) user satisfaction, (5) individual impact, and (6) organisational impact. According to the model the use and the user satisfaction depend on the information quality, and the overall quality of the system and the service. As a result of using the system, an individual achieves certain benefits. These benefits will (positively or negatively) influence his satisfaction and impact on the decision whether or not to use the system further. Therefore, the model posits that any IS system can be evaluated in terms of its overall quality and the quality of the information it produces.

Ten years after the publication of their first model, based on its evaluation and many contributions to it (DeLone & McLean, 2002; Molla & Licker, 2001; Saarinen, 1996; Seddon, 1996; Seddon,

1997), DeLone and McLean proposed an updated model of information system success (DeLone & McLean, 2003), where they replaced individual impact and organisational impact with 'net benefits' variable as illustrated in Figure 26.

Figure 26: IS success model (DeLone & McLean, 2003, p. 24)
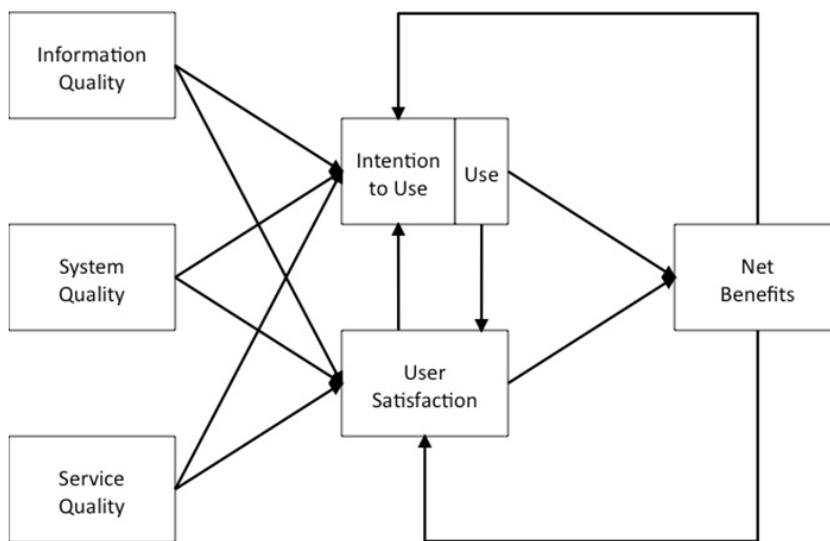
Interestingly, user objective satisfaction has always been considered as an appropriate measure of usefulness of technology. Defined broadly as the opinion of the user about a particular computer application, which they use (Doll & Torkzadeh, 1988), the concept focusses upon the views of end-users, and initiated the development of several instruments for assessing user satisfaction in a number of subgroups and environments. For example: end-users (Doll & Torkzadeh, 1988), decision support applications (Sanders, 1984), or organisations (McHaney *et al.*, 1999). Mahmood (2000), based on the results of the meta-analysis suggested that in an organisational context user satisfaction is impacted by: (1) the quality of information software produces, (2) features of the user interface, (3) support provided by developers, and (4) user attitudes towards the software.

3.3 Discussion

The need to understand what impacts the way individuals perceive IT systems and what determines their decision to use it, was the motive behind this review. From a range of bodies of literature, I identified nine major theories focused on use, acceptance and satisfaction of technology. These theories were to be used as a guide to research design, data collection, and data analysis.

One of the main theories identified in this review is the technology acceptance model and related theories, which were initially built on the theory of reasoned action (Ajzen & Fishbein, 1980), and were concerned with predicting the acceptance and the use of technology. The original technology acceptance model suggests that user salient beliefs - perceived usefulness and perceived ease of use are the main factors affecting subsequent use. On the contrary, DeLone and McLean (1992, 2003; 2002), who studied the concept of 'success', suggested that information quality, system quality and service quality are the main determinants of use and user satisfaction. Looking from the different perspective, Goodhue (1998), proposed the concept of task-technology-fit, as the primary determinant of use.

Several similarities have been identified in the constructs of these theories. Firstly, the majority of theories considered the influence of society and labelled the constructs as 'subjective norm' or 'image'. Secondly, the meaning of some constructs although labelled differently remained the same, for example, effort expectancy construct that was used in the unified theory of acceptance and use of technology is similar to perceived ease of use in the technology acceptance model. Similarly, 'perceived usefulness' in the technology acceptance model and the technology acceptance model 2 can be related to 'relative advantage' in the innovation diffusion theory.

The contribution to knowledge made by all of these theories is not to be underrated. However, they also have drawbacks, which had to be carefully considered before the decision upon adopting any particular theory in this study.

Over the years, the technology acceptance model has been used as a 'gold standard' in IS research, even though reviews confirmed that it only explains 40% of acceptance (King, 2006; Ma & Liu, 2004). Legris *et al* (2003) suggested that the model has some significant limitations. For example, it does not assess the use over time; for example, post-implementation, during the first few months or after one year. Also, in predicting use the model ignores the role of the management in the organisational context, and the fact that ultimate decisions whether or not to implement a particular technology is very often made by a group of key stakeholders.

Also, although the technology acceptance model helps us to understand why individuals use technology, it fails to show which components of a particular technology are useful and which are not, and specifically what makes technology useful. In this respect, it has limited value for the purposes of this study. Moreover, the model assumes that more utilization enhances performance, whereas in case of technology poorly fitted to the task, more use will not improve performance, on the contrary it may reduce it (Bagozzi, 2007).

There is an argument that the technology acceptance model has outlived its usefulness, as the IT/S environment has changed dramatically during the past decade and it is possible that in the current environment, there are many more factors that could potentially impact on the use of technology (Bagozzi, 2007). It is also argued that the development of other theories was halted, as over the years the attention of IS researchers focused on replicating the model original construct instead of testing new concepts (Benbasat, 2007).

Nevertheless, even though to date there is little guidance on the appropriateness of the model for particular types of technology, the model has remained the most often cited theory in the field (Shaikh, 2015). A recent review of the use of the TAM in healthcare showed that the theory has a widespread application in explaining determinants of technology acceptance in health (Holden & Karsh, 2010).

In the case of the IS success model, one of the main criticisms comes from the fact that it simplifies use and intention to use (Gable, 2008). It ignores the fact that the use might fluctuate over time and might change, for example, as the result of learning. Moreover, the model does not explain how the main variables change in situations when the use is either mandatory or voluntary. Another criticism comes from the fact that, although the model has been validated (Seddon, 1996), mixed support for some of the relationships posited by the model has been found, for example that information quality leads to greater use, or that system quality leads to greater use. That suggests that the construct as a whole might be inconsistent (Petter, 2008).

The major criticism of the task-technology fit theory comes from the fact that, it does not provide a standardised measure of fit that could be applied in the research process. Moreover, the use of this theory is very much unguided. In the literature it is often combined with the technology acceptance model, as the two complement each other, for example by Klopping and McKinney (2004) and Larsen *et al* (2009).

3.4 Impact of reviewed theories on the design of the study

The unique nature of an online economic model makes some of the reviewed theories more appropriate than others, though all have been very influential in the design of this PhD study. During the course of the work, the technology acceptance model and information success model became central to my understanding of the requirements for a useful economic model and the

factors that can impact decision-makers when using it. My understanding was that there is a risk of end-users being discouraged from using the model for reasons related to its ease of use (the technology adoption model).

The review allowed me to identify the ways in which decision analytical models (DAM) are analogous to decision support systems (DSS). Both consist of three fundamental components: (1) the database, (2) the model, and (3) the user interface (Alter, 1977); and DSS are validated by some form of comparison with the real world rather than against a set of requirement specifications (Alter, 1977), with DSS aiming to develop an understanding of specific queries.

According to a taxonomy used for DSS, DAM can be perceived as (1) 'representational DSS' (Alter, 1977), as it performs 'what if' analyses; (2) 'passive DSS' (Haettenschwiler, 1999), as it aids decisions but does not explicitly suggest solutions; or (3) 'model-driven DSS' (Power, 2002), as it uses data provided by users.

I therefore adopted a theory of IS success which posits that IT/S success could be assessed based on the analysis of its use and user satisfaction. This theory has been very influential in the design of the methods for MALT model evaluation.

A review of relevant theories drew my attention to the importance of the quality of the proposed model, and several factors have been considered to ensure the quality of each of the models' components. To ensure information quality, established health economics evaluation methods were applied in the process of model validation, such as sensitivity analysis on the models' results, and comparisons of results with those of existing models (McCabe & Dixon, 2000). System and service quality was ensured by a series of usability tests (as discussed in Chapter 7). The only dimension that was considered as being beyond the empirical abilities of this research was 'net

benefits'. As discussed previously, a long-term follow-up of the model users was deemed beyond the scope of this research.

Although the concept of success seems to primarily occupy the domain of computer studies research, there are now indicators that it may be applied to the fields of operations research and health economics. Two researchers who pioneered the concept of success in operations research literature are Balci (1994, 2002) and Ulgen (1991). Balci suggested that for a model to be regarded as successful, its results must be credible and accepted by decision-makers. Ulgen proposed a set of criteria for judging the success of modelling projects: completion on time, implementation in real life, and achievement of measurable financial savings due to use of the model. In my view, modelling literature is slowly adopting 'success' as an overarching term to include the concepts of 'validity', 'credibility', 'acceptability', and 'confidence'.

'Task technology fit' highlights the factors that may influence the way the model is perceived by its users. For example, the characteristics of individuals, in particular their skills, may affect the way in which a technology is assessed, with a person assessing the software as 'not useful' simply because he is struggling to unfold its true potential. Task technology fit therefore has practical implications for the evaluation of the proposed model, thus I collected demographic data on model users, their modelling experience and software preferences, and their reasons for use of the model. This additional data was to be used in the case of contradictory findings from the model evaluation or poor feedback.

At the outset of the project, I decided to use the data on model usage as a surrogate indicator of its usefulness, as suggested by the technology acceptance model, the technology acceptance model 2 and the unified theory of acceptance and use of technology. I hypothesised that if users considered the model to be unreliable or its data inaccurate, their usage would reflect this. The

voluntary nature of the model meant that it could be avoided or abandoned if it did not reflect the needs of its users. For this reason, I collected the data on model use in the form of logs (as discussed in Chapter 9).

The concept of 'end-user computer satisfaction' (Doll & Torkzadeh, 1988) was useful for this study, as it guided the choice of the instrument for measuring users' opinions about the model, and in an early stage of the project an alternative concepts such as model '*success*' was considered for this purpose.

Exploring issues of user satisfaction led me into the minefield of behavioural theories, with studies addressing matters such as cognitive dissonance (Festinger, 1962), and confirmation bias (Nickerson, 1998). Festinger (1962) states that individuals seek consistency between their expectations and reality, and this is significant in the modelling context because users are likely to reject results that contradict their expectations or beliefs. Similarly, confirmation bias (Nickerson, 1998) refers to the tendency to interpret evidence in terms that are in line with one's beliefs, expectations, or pre-defined hypotheses. Both concepts have been used extensively in experimental psychology, and have been applied in operations research to explain why the results of simulation projects have been rejected or disbelieved. For example, Sterman (2002) relied on it to explain the results of modelling studies in system dynamics.

I was also influenced by the findings of Robinson (1998), who suggested that perception of the 'success' of a modelling project, in his case simulation software, may change over time. He suggested that a successful model is one in which the results have been accepted, implemented, and then confirmed to be correct in real life. He distinguished between four stages of modelling project success, with a range of possible benefits of each. In the first stage, a modelling project can be regarded as successful if it improved understanding of the key issues. Success in the further

three phases is determined by (2) stakeholders accepting the results, and the study results being (3) implemented and (4) confirmed in real life (4). Robinson suggested that success in the first two stages does not necessarily mean that a simulation study will be deemed successful, but failure in the early stages does mean overall failure for the project. This staging system provided me with a better understanding of how modelling studies could be assessed in practice.

This study was designed to reach Stage 2, as defined above, with the findings of the model validation and following discussions revealing whether stakeholders accepted the model results for their telehealth services. The extent to which the results were implemented and confirmed in real life lay beyond the abilities of this research, hence net benefit could not be measured.

3.5 Conclusions

To identify measurable constructs of economic model usefulness, I looked firstly into how people accept and use technology. Central to this is the concept of 'perceived usefulness', present in the technology acceptance model and related theories, as well as the information system (IS) success model. Secondly, I paid attention to factors affecting users' perception of software and computer technology overall, reviewing theories such as technology fit and computer user satisfaction. Lastly, as a model was to be embedded into the organisational structures of the NHS and telehealth industry, I investigated into how organisations adopted innovations.

4. Literature reviews

4.1 PART ONE: Review on the use of modelling in the NHS

This chapter presents a review of the literature on the barriers to simulation and modelling uptake in the NHS. There has been a recent increase in research examining barriers to the use of simulation and modelling, and implementation of modelling results, and this chapter draws together the evidence from health economics and Operational Research studies.

4.2 Methods

The general principles of literature review were applied for this review (Grant & Booth, 2009). The literature search was conducted in Web of Science and Scholar databases, considering English publications from 1990 to the present. Key search terms included: healthcare modelling OR modelling OR simulat* OR markov* AND use. Journals including the Journal of the Operational Research Society and Value in Health were searched manually for further relevant publications. The inclusion criteria were that studies were to be published in English and use qualitative methods. Those using other methods were excluded.

4.3 Results

The search identified 35 articles, titles, and abstracts, all of which were read. Whilst the issue of modelling use was discussed in all 35, only nine studies were qualitative inquiries and thus included in this review. The summary of these is provided in Table 5.

The reviewed studies considered the use of modelling on macro (national), meso (regional), and micro (local) levels of decision-making, and made observations on existing barriers to its use. These barriers are discussed in detail in the following sections. The studies employed a variety of

methods, including qualitative interviews and focus groups, observation, and analysis of documents. Two studies using survey methods were also identified.

To consider the use of evaluations on the national level, Bryan and colleagues (2007) conducted qualitative research on the use of cost-effectiveness analyses in technology appraisals with members of the NICE appraisal committee. Unlike the local-level decision makers, NICE has the advantage of being able to commission cost-effectiveness research for every topic considered. However, even for NICE, access is a major barrier to modelling use. Bryan's (2007) understanding of this was that there was not only a barrier to physical access to modelling tools, but also to the information they contained, when interpretation of the results or understanding of modelling principles is limited. For example, some members of the appraisal committee reported limited understanding of the evaluation concepts and called for better presentation of modelling results. The study also noted that there were problems with acceptance of the results, especially when QALY was used as an efficiency indicator and the results were presented using ICERs and cost-effectiveness threshold.

Early research into the use of economic evaluation at the local and regional level of the NHS include studies conducted by Drummond *et al* (1997) and Duthie *et al* (1999). In 1997 Drummond conducted an extensive survey with UK decision-makers on the barriers to economic evaluation uptake, and asked the respondents to rank barriers to modelling use from a provided list. The results of the survey suggested that the critical barrier to modelling uptake was the fact that the modelling does not inform operational decisions. Moreover, NHS managers stated that budgetary constraints, especially inability to move resources between primary and secondary care budgets, were one of the main reasons for not using simulations. This issue was also highlighted by Van Gool *et al* (2007) in their study with Australian decision-makers, where they called for incentives which would force decision-makers to align healthcare practice with the evidence from cost-

effectiveness research. Drummond *et al* (1997) also pointed to methodological barriers to modelling uptake, with models suspected of bias due to the number of assumptions used or their funding sources. Moreover, authors pointed to the lack of long-term view in published cost effectiveness research as a factor that negated its usefulness.

Consistent with these findings were the results of a qualitative study by Duthie *et al* (1999), which comprised interviews with 34 decision-makers from regional health authorities, including GPs, clinicians, NHS managers, hospital pharmacists, and health authority directors. The primary finding of the research was that stakeholders were choosing not to use cost-effectiveness evaluations because of the mismatch between them and the way in which contracts were set up and money flowed in the NHS. For example, budgetary constraints meant that when a decision was made to fund a new technology, stakeholders found it difficult to relocate the resources needed. In the study of European decision-makers, these budget constraints were found to be a major barrier to modelling use (Hoffmann *et al*., 2002). Moreover, authors noted that economic jargon was deterring clinicians from using models and recommended that it should be avoided (Duthie *et al*., 1999; Hoffmann *et al*., 2002).

Hoffman *et al* (2002) met with the decision-makers from two health authorities. During a focus group, they asked consultants, GPs, and other decision-makers to review a sample of economic studies from the NHS Economic Evaluation Database. Methodological issues around the quality of the evidence used in these evaluations, and the generalisability of the results were among the reasons reported for refusing to use modelling. Decision-makers stated that the majority of the available studies were conducted in USA and therefore had limited transferability into the UK setting. Consistent with previous research on the subject, the study found that poor uptake was due to economic evaluations having limited usefulness in the case of operations and management decisions, as they answer the question of cost and the effects of health technologies, rather than

those of operational issues such as bed allocation and staffing. The other finding of this research was that understanding of health economics principles among decision-makers was limited, which further impacted their trust of modelling results. Focus group participants experienced problems with understanding economic jargon.

Discussing barriers to the implementation of models, Brailsford (2005) emphasised that the NHS is culturally different to all other industries where modelling is routinely used. Healthcare organisations deal with human beings, and staff struggle to adopt techniques that view humans as 'widgets'. She also discussed the high cost of simulation and poor quality data as impeding factors.

Investigating the use of CE analysis in local formulary decisions, Williams and Bryan (2007b) found that access was still one of the major barriers to modelling uptake. Other findings were consistent with previous research, and included the acceptability of simulation and modelling due to concerns over bias, both from the number of assumptions used in these studies and sponsorship of the research. Studies funded by private organisations and/or pharmaceutical companies were treated with particular caution. Williams *et al* (2008) further investigated these barriers, conducting a systematic review of the literature and five in-depth case studies with NHS decision-makers. Documentary analysis, observations of committee meetings, and semi-structured interviews confirmed previous findings that access and acceptability were the main barriers to modelling use. When considering access to modelling, they distinguished again between problems with obtaining cost-effectiveness studies and problems with interpreting results. The latter was especially present at both the national and local levels of decision-making, and was explained by poor understanding of cost-effectiveness principles.

The authors recommended two strategies for overcoming issues with access. The first was to develop a standardised approach to the presentation of cost-effectiveness results, and the second

to provide extensive training in health economics for those involved in decision-making. In terms of acceptance, the study found that decision-makers did not trust evaluations because of the assumptions used in models, the concerns about the robustness of the analyses, and the poor quality of these studies. The research also reported that some stakeholders, especially on the national level, had ethical objections to a strongly utilitarian approach to health economics principles. They stated that cost-effectiveness failed to consider the opportunity costs of adopting a new drug, especially, 'The sacrifice that would be required for the additional resources to be made available', and its use instead of cost-effectiveness threshold is problematic. The interviewees from NICE also questioned some of the tools and methods of health economics, in particular the concept of QALYs and the failure to consider equity issues in analyses. On both the national and local levels, concerns about the robustness of analyses and the quality of data were reported.

The same findings were reported in the study conducted by Eldabi (2007), which focused on academic and simulation experts and the future of simulation and modelling in healthcare. There was a consensus that the poor quality of data and lack of whole system approach were the main barriers to simulation use. The absence of appropriate, easy-to-use tools was also seen as a barrier. The research suggests that, as healthcare continues to adopt tools and techniques from the industry, there is a need for greater integration between health economics and OR. This issue was also noted by Harper and Pitt (2004) when devising a framework for successful implementation of OR models. The key to successful modelling was seen as ensuring the quality of the data, as this would help to build confidence in modelling among decision-makers. Other recommendations included considering appropriate software tools and/or toolkits for modelling.

Another study by Eddama *et al* (2009) on the use of economic evaluations among local level decision-makers found that, despite developments in health economics and OR, the previously

identified barriers were still present. Data collected during qualitative in-depth interviews, observations of decision-making meetings, and analysis of documents produced at these meetings revealed that there was still a mismatch between local decision-makers' information needs and the available evidence. Stakeholders stated that they disregarded economic evaluations as they needed tools that would support management and operations decisions. They also had limited understanding and awareness of health economics approaches. With the exception of NICE guidelines, the study found no evidence of use of economic evaluation on local level decision-making.

Finally, a study which stands out in the identified literature is that by Brailsford *et al* (2011), which comprises a case study exploring barriers to simulation adoption and use in the NHS. The study was conducted when modelling software Scenario Generator was made available for a number of PCTs in the country. The staff from participating NHS organisations were interviewed about their experience with the software, and the reasons behind either their adoption of it or refusal to use it. The group included people who had not used it, people who had attempted to use it but given up, and active users. The study provided an insight into the experience of using Scenario Generator, and identified several barriers to modelling use in the NHS. These included data availability and quality, a lack of understanding of the data embedded in the model, a lack of time and resources for working with the model, and the need for senior management support. A summary of all of the reviewed studies is presented in Table 4 below.

Table 5. The summary of studies included in the review of barriers to modelling and simulation use in the NHS

Author	Study population	Methods	Identified barriers	Type of decision analytic model	Mode of use	Type of decision making
Bryan *et al* (2007)	NICE appraisal committee	Qualitative interviews	Accessibility Acceptability	CEA and CUA (mainly Markov models)	Expert	Decision-making
Drummond *et al* (1997)	Directors of public health, directors of hospital pharmacy	Survey	Budgetary constraints Methodological barriers	CUA	Expert	Decision-making

Author	Study population	Methods	Identified barriers	Type of decision analytic model	Mode of use	Type of decision making
Duthie *et al* (1999)	GPs, clinicians, NHS managers, health authority directors	Qualitative interviews	Budgetary constraints Economic jargon Discrepancies between decision problem and the outputs presented by the model	Assessing the relevance of sentences referring to outcome of analysis of the effectiveness of treatment A over treatment B	Expert	Decision-making
Hoffman *et al* (2002)	GPs, consultants, health authority staff	Questionnaire Focus group	Methodological barriers Economic jargon The understanding of health economics principles	Assessing the abstracts from NHS EED database	Expert	Decision-making

Author	Study population	Methods	Identified barriers	Type of decision analytic model	Mode of use	Type of decision making
Williams and Bryan (2007b)	Hospital pharmacists, local formulary committee members	Documentary analysis Observation Qualitative Interviews	Perceived bias due to the industry sponsorship Inability to realise savings	NICE guidelines Economic evaluations	Expert	Decision-making
Williams *et al* (2008)	The NICE technology appraisal committee, primary care management committees, hospital medicines management committee, country wide priorities committee	Case studies: documentary analysis, observations and qualitative interviews	Access Acceptability Methodological	Markov	Expert	Decision-making

Author	Study population	Methods	Identified barriers	Type of decision analytic model	Mode of use	Type of decision making
Eddama *et al* (2009)	Local level decision makers	A systematic review of the studies using qualitative interviews, observations, documentary analysis	Economic evaluation did not support management decisions Lack of awareness of economic evaluation approach and its usefulness for local level decision making	CEA	Expert	Decision-making
Eldabi *et al* (2007)	Experts' opinion	Survey	Poor quality of data Need to adopt whole system approach	Simulation	-	-

Author	Study population	Methods	Identified barriers	Type of decision analytic model	Mode of use	Type of decision making
Brailsford *et al* (2011)	NHS staff	Focus group Interviews	Data quality Lack of time and resources to work with models Need for champions to promote the use of models Lack of senior management support	Simulation	End-user mode	Problem-solving

(-) not included

4.4 Discussion

A review was conducted of the literature on the use of simulation and modelling in the NHS, covering studies of various healthcare decision-makers at the national, regional, and local levels of the NHS. The review highlighted a number of important barriers to simulation and modelling use in the NHS, which could be broadly categorised as being related to acquiring and working with models, and problems with modelling methodology and assumptions. Research into decision-making in the NHS has reported limited or moderate uptake of modelling and economic evaluation techniques (Eddama & Coast, 2008; Eddama & Coast, 2009; Hoffmann *et al*., 2002; Jahangirian *et al*., 2015; Taylor *et al*., 2013b, p. 1271; Williams & Bryan, 2007a; Williams *et al*., 2008). The lack of relevant models was one frequently mentioned barrier, and access to evaluations was considered to be limited in the case of all stakeholders, except NICE (Williams *et al*., 2008), with the main reason given being the cost of commissioning new models and the need to employ health economists.

Williams *et al* (2007a) explained this phenomenon using Weiss's classification of research utilisation (1979). Weiss believed that research findings could be utilised by rational or interactive models. In the rational model, findings are always treated as a primary source of information, and therefore they influence decisions and get implemented. NICE is a good example of this model. The only reasons why modelling may not be used are access to evaluations at the time they are needed, and difficulties with understanding the results due to the language used or the presentation of the results.

Although Williams *et al* (2007a) noted that simply increasing the number of evaluations would not resolve the problem of limited use of modelling, the literature is replete with recommendations to improve the availability of models. For example, Robinson *et al* (2004) discussed the idea of

simulation model reuse, ranging from reuse of parts of the code, to model components, and complete models. The idea was discussed in relation to simulation modelling in general, and its major benefits were seen as the reductions in time and cost of developing models. Discussing the benefits of simulation in healthcare, Bowers *et al* (2012) suggested that it might be beneficial to provide NHS managers with a simulation model with a robust interface, allowing them to perform the analyses themselves, for example in case of hospital bed allocation. He developed this idea further by also suggesting that adding real-time forecasting features to these models would give them the chance to be adopted as routine decision support tools.

Examples of such implementations are limited, with the models developed by Brailsford (2011) and Proudlove (2007) being the exceptions. In Weiss's (1979) interactive model of decision-making, decisions are based on 'negotiated compromise and the balancing of competing interests, rather than solely on the available evidence' (Williams & Bryan, 2007a, p. 138). The decision-making on adoption of new technologies at the local level of the NHS is a good example of this model. The decision is usually taken by committees comprised of various stakeholders (prescribing committees, priorities networks, and representatives of local health organisations) and the process is prone to conflict due to the multitude of competing interests and agendas. The use of modelling in such decision-making processes is challenging, and requires convincing others of the results of the modelling, or sometimes even of the method itself. A good explanation for the conflicts in model-based group decision-making is given by Franco (2013), in the discussion of models as boundary objects.

Organisational barriers were identified in several studies as significant factors affecting modelling uptake (e.g (Duthie *et al*., 1999; Williams *et al*., 2008). The inability to transfer resources between different health budgets, and the fact that they are fixed were factors cited in several studies of NHS managers (Duthie *et al*., 1999; Eddama & Coast, 2009). In health economics, the common

understanding is that the resources from substituted alternatives can be used to fund new health technology. However, in reality the healthcare budgets are fixed, so that in order to fund a new technology managers are faced with a task of finding the resources. Further barriers are imposed by resistance to adoption of tools in the industry, and healthcare is an environment characterised by constant change (Brailsford, 2005).

In the literature, considerable attention is given to the way in which models are accessed by their users, including discussion of the importance of user-friendliness of interfaces, toolkits, and platforms (Bowers *et al*., 2012; van Gool, 2007). For example, Eldabi *et al* (2007) pointed out the lack of easy-to-use tools which can be readily used by clinicians and managers, and suggested that developing the methods to 'glue models together' or provide disposable modelling might be a solution. He also called for the development of a generation of models that end-users could set-up and test themselves. Additional recommendations were to avoid economics jargon in the model, with several authors proposing that specialists' language deterred use of the models (Duthie *et al*., 1999; Eldabi *et al*., 2002; Hoffmann *et al*., 2002).

Related to the issue of barriers is that of acceptance of economic evaluations and trust in modelling results (Hoffmann *et al*., 2002; Williams *et al*., 2008). Williams (2008) and Bryan (2007) note that 'acceptance' pertains to both the issues of accepting the results of modelling and assumptions used in particular models, but also to accepting the principles of health economics. There is a consensus in the reviewed studies that understanding of health economics principles among decision-makers is limited (Eddama & Coast, 2009; Hoffmann *et al*., 2002) and that this further impacts on trust in modelling results. The proposed solutions included greater training in OR and modelling methods (Drummond *et al*., 1997; van Gool, 2007; Williams & Bryan, 2007b) and involving stakeholders in model development process (Brailsford *et al*., 2009, pp. 1840-1849; Harper & Pitt, 2004; Robinson & Pidd, 1998).

Investigating stakeholder engagement in healthcare simulation, Brailsford and colleagues (2009) suggested that people who did not take part in developing the model were less likely to believe and use it. She noted that, especially in the NHS, there was limited trust in generic models, or models developed for other organisations, because of the 'not invented here syndrome' (Taylor *et al*., 2013b, p. 1271). Other supporters of stakeholder involvement claim that because NHS managers and clinicians are unfamiliar with modelling methods, involving them in the development process can ensure model buy-in (Brailsford *et al*., 2011; Harper & Pitt, 2004). Throughout the development process, stakeholders gradually gain trust in the model when they see that all issues that matters to them and their organisation are considered important and included in the model (Harper & Pitt, 2004).

Further barriers highlighted in the reviewed literature were related to methodology. The quality of data and generalisability of results were among the most often reported barriers for NHS stakeholders (Hoffmann *et al*., 2002). This was also noted in the studies with European decision-makers (Hoffmann & von der Schulenburg, 2000; Zwart-van Rijkom *et al*., 2000). Decision-makers expressed concern that the data used for some of the modelling studies were of poor quality, which further impacted the quality of results. Brailsford *et al* (2011) suggested developing better processes for data collection in the NHS and ensuring that NHS managers knew the importance of ensuring data quality. Other solutions included the development of a national body to perform quality assessment of modelling studies and economic evaluations (van Gool, 2007).

Other methodological barriers commonly highlighted in the studies were the fact that the models were prone to bias due to the large number of assumptions being used (Hoffmann *et al*., 2002). NHS stakeholders complained that the models were not transparent. Highly complex models were found to be deterring stakeholders, as they perceived them to be 'black boxes', and an

understanding of the inner workings of the models was found to be essential for building trust in the modelling results.

Within the modelling literature and good practice guidelines (Eddy, 2006; Eddy *et al*., 2012; Karnon *et al*., 2012; Philips *et al*., 2004), there exists an agreement that model users should understand how the models work and what methods are used for depiction of the decision problem. This literature, however, tends to emphasise the role of training in health economics methods as the only solution to ensuring transparency and creating understanding among model users (Fleischmann & Wallace, 2005).

There is an opinion that the reduction of the number of assumptions would always pose the risk of oversimplifying models (Eddy, 2006). Fleischmann and Wallace (2005) proposed that without compromising the complexity of the model, transparency might be achieved by documenting how the model works, using simple and understandable language, performing validity testing of the main assumptions, and giving the users direct access to the content of the model, so that they could examine the components of the model.

In light of reviewed literature, the need for complex models for the NHS should be revisited. Based on their experience of healthcare modelling, Taylor and colleagues (2013a) explained that unlike in academia, where new knowledge is constantly constructed, the NHS requires flexible, easy-to-understand models that do not require specialist software, and can be developed relatively quickly and updated without the input of modellers.

Lastly, one of the findings of this review is the value of conducting empirical or observational research in which participants are provided with particular models for routine use. There was an observable difference between findings from these types of studies and those of qualitative

studies, where the stakeholders were interviewed on the subject of modelling use. The empirical research provided more in-depth explanations of existing barriers and insights into the practical implications of model use in organisations. For example, a 'lack of time to use the model' was not reported anywhere other than the (Brailsford *et al*., 2011) study.

4.5 Conclusions

This literature review identified several common barriers to modelling uptake in the NHS, with solutions to some of these issues proposed. However, the limited number of studies on successful modelling in the NHS suggests that these barriers remain largely in place.

4.6 PART TWO: Review of methodology

Usability evaluation is now recognised as a necessity in the software development field. Having in mind the interdisciplinary character of these methods and the infancy of usability research in health economics, a systematic review has been conducted to identify methods which could be applied in the usability evaluation of an economic model interface.

4.6.1 Introduction and review objective

Usability is a core term in various fields such as human computer interaction (HCI), user centered design (UCD) and usability engineering. Usability is not a one-dimensional concept; and comprises of multiple aspects such as learnability, effectiveness, efficiency, satisfaction, and presence of errors (Nielsen, 1994; Rubin & Chisnell, 2008). Learnability of the system means that users can quickly adapt to it. Efficiency means that users should get maximum productivity after learning how to use it. Also high user satisfaction and low error rate are very often used as indicators of usability.

Although there are several usability standards e.g. (ISO, 1998) and conceptual models e.g. the metrics for usability standards in computing (MUSiC) (Bevan, 1995); the goals, operators, methods and selection framework [GOMS] (John & Kieras, 1996), the quality in use integrated measurement (QUIM) (Seffah *et al*., 2006), there is no unified definition of usability. The most concise concept defines usability as 'the effectiveness, efficiency and satisfaction with which users can achieve goals in particular environments' (ISO, 1998). The examples of how the construct evolved over the years are shown in the Table 6 below.

Table 6: Usability metrics

(Shackel, 1991)	(Nielsen, 1994)	MUSIC (1995)	(ISO, 1998)	(Constantine & Lockwood, 1999)	QUIM (2006)
-	Efficiency	Users' performance	Efficiency	Efficiency in use	Efficiency
Effectiveness	Error frequency	Acceptability of a system	Effectiveness	-	Effectiveness
-	Memorability	-		Rememberability	-
Learnability	Ease of learning	Ease of use	-	Learnability	Learnability
-	User satisfaction	-	User satisfaction	User satisfaction	Satisfaction
				-	Safety
	-	-	-	Reliability in use	Trustfulness
				-	Accessibility
Flexibility					Universality
Attitude					Usefulness
					Productivity

Some of the presented concepts define usability in a broad way, taking into consideration not only utility but also technical and ergonomic aspects of the application. In general, although there are different names for the usability attributes, they are used to describe similar if not the same concepts. Throughout the literature the ISO 9241 – 11 standard for usability (ISO, 1998) and Nielsen's (1994) definition are among the most often cited. The review by Hornbaek (2006) suggests that usability is context specific and there are trade-offs between usability criteria, with some being more important than others. Hertzum and Clemmenssen (2012) conducted 24 interviews with usability professionals to understand practises in the field. The results suggested that only 53% of usability constructs used by professionals were based solely on ISO 9241, and in the case of majority of evaluations experts made decision on which usability facets to include for the specific contexts. They suggested that usability has evolved from a narrow into a diversified concept with user satisfaction becoming a central concern of majority of usability evaluations (Hertzum & Clemmensen, 2012) .

As a general concept usability cannot be measured per se, but usability attributes can have measurable parameters assigned to them (Nielsen, 1994). Nielsen categorised them as *objective performance measures* indicating how capable are users in using the system – and *subjective preference measures* which show how users like the system (Nielsen & Levy, 1994). The examples of the following include:

Effectiveness

- Percentage of successfully completed tasks

- Number of errors made by users

Efficiency

- User effort

- Time needed to complete tasks

- Number of time users seek help

Satisfaction

- User satisfaction rating scales

- Percentage of users who indicated they were satisfied with the application

- Percentage of users who had positive attitude towards the application

This systematic literature review aimed to understand:

1. What usability evaluation methods have been applied by researchers to evaluate IT applications similar to the proposed model (decision support systems or web based software user interfaces) and how were they employed?

2. What are the existing approaches for data analysis from such studies?

4.6.2 Methodology

There are a number of handbooks that provide an excellent guide to the usability evaluation methods. For example, publications of Nielsen (1994; 1994) and Rubin (2008) are regarded as gold standards in the field. Nevertheless, it felt appropriate to conduct a review on usability case studies to generate more robust findings and identify a wide range of practical aspects of applying these methods in various settings.

Methodological reviews are challenging and not well defined. In developing a methodology for this review I adopted Hutton's (2000) approach. I developed comprehensive search strategies to identify usability case studies from information systems, human computer interaction, and management science disciplines. By appraising them I was able to identify a method which could be best replicated in the current study.

4.6.3 Methods

4.6.3.1 Search strategy

A comprehensive search strategy was developed in consultation with specialists from the User Centered Healthcare Design stream of the South Yorkshire CLAHRC. A list of following key terms was developed and used in the process: 'usability testing', 'user interface', 'usability evaluation', 'usability of a website', 'usability of an interface', 'iterative usability testing'.

In order to identify studies which would not denote the term usability in either the title or as a keyword, but measure aspects of it I decided to extend my search by using the terms 'effectiveness', 'efficiency' and 'user satisfaction'. However, the aim of the review was not to identify all literature on the subject but to conduct a comprehensive overview of methods used in the field on a representative sample of available literature.

The inclusion criteria were: 1) studies published in English, 2) evaluations of a web based or software applications, 3) the study should report qualitative or quantitative data on usability and justification of employed measure of usability.

The main criteria for the selection of studies were that they had to bear a resemblance to some of the characteristics of the MALT model interface. Firstly, the interface will be web based, so that

the studies on websites were included. However only two aspects of websites were of interest: navigation and content. Issues such as colour scheme, traffic, safety were not of interest of this study as software developers' assured fulfilling these criteria when they undertook the task to develop the website. Secondly, it could be regarded as bespoke software so that studies on software applications were included. Finally, it contains some characteristics of the decision support system as discussed in Chapter 3 section 3.4. The aim of the review was to overview available methods not to collect data on all usability research conducted up to date.

Excluded were studies focusing solely on describing experiences with user interfaces and comparing predictions of theoretical models with actual user behaviour. Studies on other technologies such as TV, mobile phones, tablets, including games were excluded.

A search was conducted in the literature published between the years 1980-2014, using subject databases, specific journals and grey literature sources.

   a) Subject databases

Published literature was searched using aforementioned terms in the Health Informatics and Computer Science and Information Systems databases, specifically in: IEEE database, PubMed, Science Direct and Web of Knowledge.

   b) Specific journals electronic search

Key journals in the field (International Journal of Human – Computer Studies, Information and Management, International Journal of Medical Informatics and Lecture Notes in Computer

Science) were searched to identify further relevant studies. Citations in key articles were searched for relevance.

c) Grey literature searches

A search of grey literature was conducted in Google scholar web and eTheses to identify relevant literature such as PhD theses, summaries of reports and projects.

d) The key publications

The key publications were used to identify relevant references by hand searching the reference lists. All retrieved references were entered into EndNote™ (version 7) to remove duplicates and select appropriate ones.

4.7 Data extraction

Studies meeting inclusion criteria were screened using data extraction forms, to identify the type of technology used, sample size, usability measures employed and data analysis method. The results of this screening are presented in Table 4.

4.7.1 Description of the studies

4.7.2 Included studies

The search resulted in 226 results, after applying the exclusion criteria 54 of them were included in the final review.

4.7.3 Excluded studies

Based on the inclusion and exclusion criteria, 92 studies were excluded after screening the citations. This includes studies published in languages other than English, studies conducted on technology other than websites or software applications and studies which did not present any data.

After full texts were retrieved and studies screened, a further 80 studies were excluded on the basis of not meeting inclusion and exclusion criteria. The details of the study selection process are presented in the Prisma diagram (Figure 27) below.

Figure 27: Prisma diagram usability evaluation studies
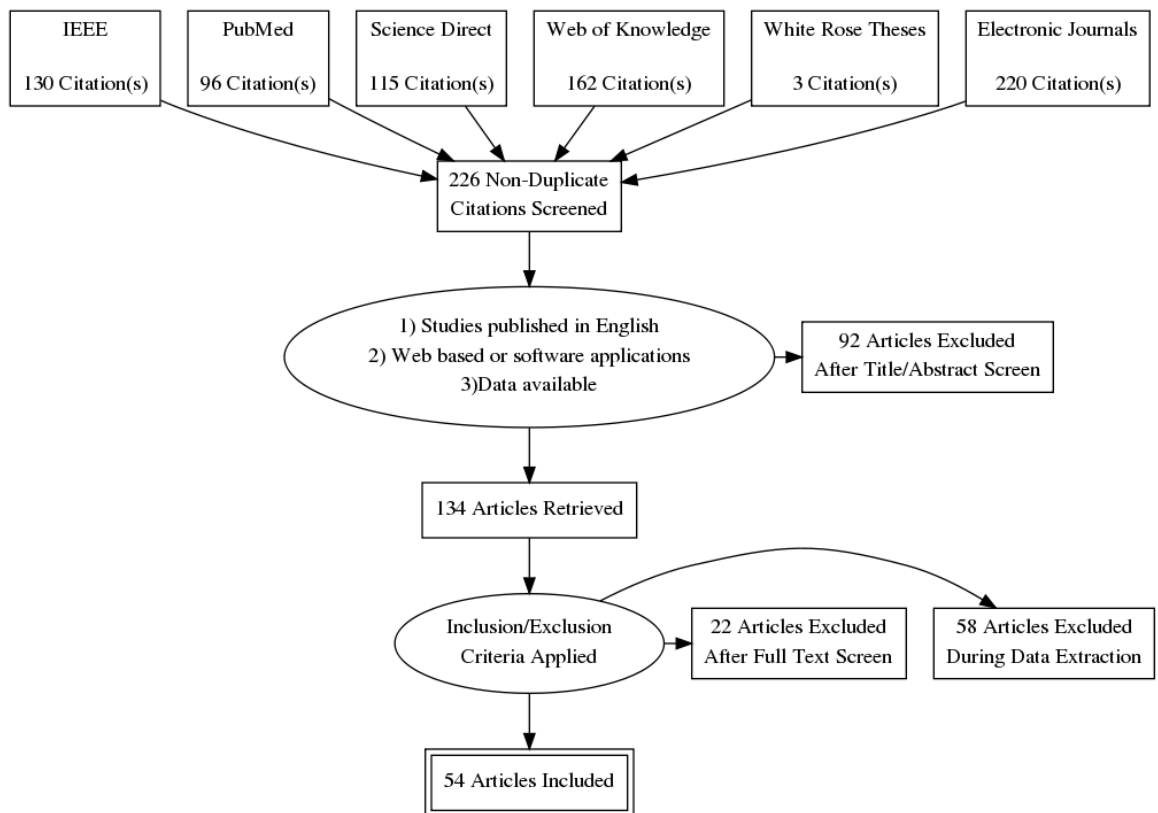
4.7.4 Characteristics of the included studies

Among the 54 included studies, there were 12 quantitative studies, 27 qualitative studies and 15 mixed methods studies.

52% of studies were conducted on websites or portals, for example, an electronic commerce website (Benbunan-Fich, 2001), library portals (Ward, 2005) or health portals (Ossebaard *et al*., 2012).

48% of studies were conducted on interfaces of software applications, for example modelling software (Paradowski & Fletcher, 2004), online disease management systems (Carroll *et al*., 2002; Leslie *et al*., 2006; Lin *et al*., 2009; Stinson, 2010), clinical data management software (Christ-Neumann *et al*., 2014; Fossum *et al*., 2011), electronic health records (Edwards *et al*., 2008). The key characteristics of the studies are summarised in Table 7 below.

Table 7: Studies included in the review of methodology

Author/Year	Topic	Design and sample size	Sample size*	Key Measures of Usability if stated
[1] (Abdinnour-Helm *et al*., 2005)	Learning Platform	Task Analysis Questionnaire (EUCS)	176 176	Completion time EUCS satisfaction score
[2] (Aggelidis & Chatzoglou, 2012)	Hospital information system	Questionnaire (EUCS)	341	EUCS satisfaction score
[3] (Allen *et al*., 2006)	Website	Heuristic Evaluation	4	-
[5](Avouris *et al*., 2003)	Website	Questionnaire (65) Heuristic Evaluation (11) User observation (5) Heuristic Evaluation (17)	65 questionnaires 11 evaluators users 17 evaluators	-
[6](Bairamzadeh & Bolhari, 2010)	University Website	Questionnaire	270	User satisfaction

[7](Battleson *et al.*, 2001)	Library Website	Task Analysis + Think Aloud	11	-
[8](Benbunan-Fich, 2001)	Commercial Website	Think Aloud	8	-
[9](Blackmon *et al.*, 2003)	Website	Cognitive Walkthrough	119	-
[10](Carroll *et al.*, 2002)	Clinical DSS	Interviews Heuristic Evaluation Cognitive Walkthrough Think Aloud	7	Effectiveness Efficiency User Friendliness
[11](Castilla *et al.*, 2013)	Website	Heuristic Evaluation + User test	10	-
[12](Chen & Macredie, 2005)	Website	Heuristic Evaluation	1	Strengths and Weaknesses of identified heuristics
[13](Chisman *et al.*, 1999)	Website	User test	10	
[15](Cooke, 2010)	Website	Think Aloud	10	

[16](Dupuits & Hasman, 1995)	Clinical DSS	Questionnaire (Bailey)	30	User satisfaction scores
[17](Edwards *et al.*, 2008)	Software	Heuristic Walkthrough	2 groups each of 4 evaluators	
[18](Fossum *et al.*, 2011)	Software (Clinical DSS)	Interviews	25	Ease of use
		Cognitive Walkthrough Observation + Graphical user Interface Evaluation	5	Usefulness
[19](Grady, 2000)	Website	User Test	2 groups of 6	
[20](Granic *et al.*, 2008)	Website	User testing + Interviews + Heuristic Evaluation	30	User satisfaction
[21](Hu *et al.*, 1999)	Software	Task analysis	715	Effectiveness
[22](Hu *et al.*, 2008)	Website	Think Aloud	11	
[23](Jaspers *et al.*, 2004)	Software (Clinical DSS)	Think Aloud + Questionnaire (QUIS)	8	

[24](Ju & Gluck, 2005)	Website	Task analysis	2 groups of 20	
[25](Kanungo *et al*., 2001)	Software (DSS)	Experiment	79	
[26](Kilsdonk *et al*., 2013)	Software (Clinical DSS)	Think Aloud	13	Effectiveness Efficiency
[27](Lee & Lin, 2014)	Software	Focus Group + Task Analysis + Questionnaire (QUIS)	30	-
[28](Leslie *et al*., 2006)	Software (Clinical DSS)	Heuristic evaluation Think Aloud Task solving + questionnaire	6 5 30	-
[29](Lewis, 1990)	Software Interface	Cognitive Walkthrough	4	-
[30](Li *et al*., 2012)	Software (Clinical DSS)	Think Aloud Clinical Simulations	8 8	Effectiveness

[31](Lin & Lin, 2009)	Website	Questionnaire + Task analysis	-	-
[32](Lin *et al*., 2009)	Software	Focus Group Think Aloud + Questionnaire Focus Group	Group of 6 and 4 9 Group of 4 and 5	-
[33](Lowry *et al*., 2006)	Website	Task analysis	120	Interactivity User satisfaction
[34](Ossebaard *et al*., 2012)	Web Portal	Interviews + Task analysis + Questionnaire	21	-
[35](Paradowski & Fletcher, 2004)	Modelling software	Task analysis	4	-
[36](Peute & Jaspers, 2005)	Software	Cognitive Walkthrough + Think Aloud	5	Efficiency User errors
[37](Pike *et al*., 2011)	Software	Questionnaire Interviews	2 rounds with 54 and 181 users Not stated	-

[38](Roberts & Fels, 2006)	Software	Think Aloud	2 groups of 9	-
[39](Ruland, 2004)	Software	Questionnaire	184	-
[40](Stinson, 2010)	Software	Interviews	11	-
		Observation	8	
[41](Swaak *et al*., 2009)	Website	Think Aloud	20	User Errors
		Questionnaire (Web evaluation)	20	Verbalizations
[42](Wang, 2005)	Website (Educational Platform)	Questionnaire	7	Learnability
		Observation		Time to completion
		Interviews		
[43](Ward, 2005)	Website (Library Portal)	Questionnaire	131	-
		User test		
[44](Wrubel, 2007)	Software	Think Aloud	18	-
[45](Wu, 2009)	Website	Interview + Questionnaire	40	-

[46](Yeh, 2008)	Website	Questionnaire	50	-
[47](Yen & Bakken, 2009)	Website	Heuristic Evaluation	5	-
[48](Yi, 2008)	Website	Questionnaire	147	-
[49] (McHaney & Cronan, 1998)	DSS	Questionnaire	442	-
[50] (Hasan, 2012)	Website	Heuristic evaluation User testing	-	-
[51] (Ebenezer, 2003)	website	Focus group User test Questionnaire	9 7 7	-
[52] (Cunliffe *et al*., 2001)	Website	Questionnaire Logs analysis User observation Heuristic Evaluation	5% 5 1	-

[53] (Trafton, 2010)	DSS	*Phase 1*	4	-
		Think Aloud	4	
		Observation	4	
		Interview	4	
		Questionnaire		
		Phase 2		
		Interviews	10	
		Observations	35	
		Log analysis		
[54] (Jankowski, 2006)	DSS	User needs analysis		-
		User observation	Not stated	
		Questionnaire	Not stated	

*valid responses

4.8 Findings of the review of methodology

The studies identified in this review presented an array of usability evaluation methods that could be broadly categorised as involving professionals or end-users. The first category includes heuristic evaluation and cognitive walkthrough. They both require the involvement of software developers or usability experts who either review the application step-by-step or serve as subjects.

In a heuristic evaluation, a group of experts evaluate the application by comparing its characteristics with a set of usability principles (heuristics). It is suggested that the method is only effective if the evaluation is conducted by a group of people, as it is believed that one person would not be able to find all usability problems. The number and the degree of severity of problems detected, also depends on the skills and the experience of evaluators. Each expert produces a list of usability problems in isolation, and only after all evaluations are completed, the findings are aggregated and discussed.

During the evaluation, each expert goes through the software several times and judges its compliance with usability principles. The findings from the evaluation are presented as a list of usability problems with references to violated heuristics. The heuristics most often used in the reviewed literature were Neilsen's (1994) and Molich's (1990), which include guidelines pertaining to issues such as language, consistency, shortcuts, error messages, and options available to the users of the particular software. For example, there is an emphasis on using simple language, allowing easy navigation through the application, and providing error messages and help options. This method is relatively easy to conduct, and it can be used in any of the software development stages.

In one of the case studies (Edwards *et al*., 2008), the method was used to evaluate an electronic health record (EHR) in a paediatric hospital. As a result, 172 usability problems were identified.

The authors concluded that the method was efficient, but challenging to use, mainly because of the nature of the environment in which the study was conducted. Many tasks completed with the EHR required the involvement of few people and evaluators found it difficult to simulate clinical workflow.

The method has been also well utilized in several studies on websites (Allen *et al*., 2006; Castilla *et al*., 2013; Granic *et al*., 2008) and is suggested to have similar effectiveness as other evaluation methods (Yen & Bakken, 2009). Apart from cost saving other benefits of the method include the fact that it does not require extensive data analysis, and that the results are available in a relatively short period of time.

Another Inspection method, the cognitive walkthrough is conducted by exploring the application by an individual or a group of experts, who attempt to simulate how a typical user would use it in order to accomplish a set of tasks (Lewis, 1990). The cognitive walkthrough is based on a theory of learning by exploration (Polson & Lewis, 1990), and on research on problem-solving.

The evaluators are not given any prior guidance on how to use the software. Users' background, the typical tasks they would perform, the procedure needed to successfully complete the tasks and application context of use are defined prior to evaluation. The analysis of the walkthrough typically consists of examining each task and telling a story of what was required to achieve a particular goal. The disadvantage of the method is that it does not assess functionality as it is only focused on one aspect of usability – ease of learning. The method is also prone to evaluator bias, as it is entirely based on evaluators' assumptions. That is why in the reviewed literature cognitive walkthrough was usually combined with another method, as in the studies by Kushniruk (1998), Carroll (2002), and Fossum (2011). Blackmon (2003) however successfully applied it in the

usability assessment of various websites. His work resulted in producing the cognitive walkthrough for a web (CWW) method.

End-users can be involved in all stages of software development, and among the methods that involve users are empirical and informal testing. Empirical testing includes the think aloud and field observation. Informal testing usually includes focus groups, interviews or questionnaire methods.

In the reviewed studies, the think aloud was one of the most often conducted. The think aloud originates from cognitive psychology, where it is used to study cognitive processes of humans. Lewis (1982) introduced this method to the HCI field but it was Ericsson and Simon (1984) who further propagated it.

In a think aloud session end-users are asked to continuously verbalize their thoughts while solving the tasks using a particular software/website. Participants are not asked to explain the reasons for their thoughts or provide any commentary but just report the information that they are currently thinking about. Such verbal protocols are used to gain insight into the cognitive processes that people use during problem-solving, decision-making or judgement tasks.

The think aloud can be conducted both concurrently and retrospectively. The difference between the two is that in a retrospective think aloud participants are video recorded when working on the task but verbalize their thoughts only after they complete it. The retrospective technique is regarded as more useful in the cases of heavy cognitive workload for example when participants use a foreign language.

Typically, users are audio or/and video recorded to collect verbalizations. The majority of the reviewed studies used video recordings (Benbunan-Fich, 2001; Hu *et al*., 2008; Wrubel, 2007), sometimes enriched by capturing eye movements as in the study by Cooke (2010). Data analysis proved to be complex and time-consuming in all cases. Verbal protocols provided rich and in-depth data that was usually transcribed verbatim. Video data was content analysed to capture discontent, moments of struggle and similar. Additional data, if captured was analysed alongside the recordings.

The most often mentioned disadvantages of the method include the fact that the feedback gathered from users is subjective and, therefore, prone to sample selection bias. There are various factors that have an impact on the type of feedback gathered, such as users' computer skills, their experience, and how articulate they are. This is the reason why it is recommended to recruit different types of users to match all types of tasks.

An often expressed concern with think aloud studies is also the sample size. In the nineties, Nielsen suggested that five people for each of the usability test rounds is sufficient for stable results (Nielsen, 1994). That point was later criticised by Lewis (1994), Spool (2001), Woolrych (2001) and Faulkner (2003) suggesting that in fact, more than 5 users are needed to reveal 85% of usability problems. The question of sample size in a usability evaluation is still far from being solved, and variations of 4 to 20 users were observed in the reviewed literature.

Researchers in the field also argue that thinking aloud could influence participants' performance and behaviour while completing the tasks. Rhenius and Deffner (1990) argued that thinking aloud slows users while completing the tasks; while Ingwersen (1992) argued against reliability of data generated using this method since there is no guarantee that it reflects real behaviour of the users.

Despite these concerns, think aloud remains one of the most frequently used methods in the field. Its advantages include the fact that it allows to obtain unique data on actual user behaviour that could not have been obtained via another method.

Another method revealed by the review, are the field observations. They were usually conducted in the end-users' natural environment and regarded as very useful in gathering data on real usage situations. In the reviewed literature, the method proved to be especially useful for studying clinical decision systems (Fossum *et al*., 2011; Stinson, 2010; Wang, 2005). Very often this method led to the discovery that the software was used in an unexpected way. However, one of the requirements of the method is that the observers have to be highly skilled and follow a structured process.

A number of the studies in this review adopted a questionnaire on different stages of software development, either to elicit user requirements during software development or feedback from the users on the actual use. For the purposes of evaluation, several validated instruments have been developed, such as Questionnaire for User Interface Satisfaction (Chin *et al*., 1988), Web Usability Questionnaire (Kirakowski & Cierlik, 1998), End-User Computing Satisfaction (Doll & Torkzadeh, 1988). The questionnaire is often used in conjunction with other usability methods, for example Avouris (2003) used both heuristic evaluation and two types of questionnaires (an online questionnaire to assess utility and WAMMI questionnaire to assess user satisfaction) in a study of an academic website. The overview of all usability evaluation methods and its advantages and disadvantages are presented in Table 8 below.

Table 8: Usability evaluation methods identified in the review

Method name	Stage of development	Requirements	Advantages	Disadvantages
Inquiry methods				
Questionnaire	To elicit user requirements and typical user tasks	Hundreds of users	Relatively easy to get feedback from high number of users	Validation of the instrument might be costly and time consuming. Questions and answers might be not well understood by participants.
Focus Group		6-9 users per group	Identifying high number of issues in relatively short period of time	Experience and skills of facilitator needed to gain feedback
Interviews		Few users	Useful information which enriches further stages	Relies on human memory which could be inaccurate.
Informal (Inspection) methods				
Heuristic Evaluation	All	A group of experts	Relatively easy to conduct. Low cost. It can be conducted early in design process. Allows to assess the application in a holistic way.	Does not identify real use problems. Results highly influenced by the skills and knowledge of expert reviewers.
Cognitive Walkthrough	All	A group of experts	Cost saving	Task specific, focuses mainly on ease of learning.
Empirical testing:				

Think Aloud	Design	5 users per round	Enables an in-depth understanding of users' attitudes and opinions about the software. Usability problems can be identified right away.	Time consuming to conduct. Data analysis is complex, especially if incudes video recordings. Facilitator needs experience and skills to conduct these interviews especially to maintain participants' verbalizations throughout the interview.
Field Observation	Final testing	3 or more users completing a set of predefined tasks	Provides data impossible to get in laboratory setting since it is conducted in users' natural environment.	Difficult to conduct as a skilled observer is required. Prone to observer bias.
Questionnaire	Final	Hundreds of users	Several validated instruments available. Recommended for assessing subjective opinions and attitudes of users.	The manner in which questionnaires are responded cannot be controlled. Sample bias might produce false results.
Analysis of logs	Final	At least 20	Data on software use is collected automatically from all its users	Relatively high volume of very refined data. Time consuming analysis. Possible issues of user privacy.

4.9 Discussion

By reviewing usability case studies, I aimed to understand how to conduct a usability evaluation and how to analyse usability data. The review provided sufficient guidance on how to conduct a usability evaluation and highlighted methodological issues the field is struggling with.

The results of the review indicate that there is no strict rule for choosing a usability evaluation method. One might argue that usability of a website differs from the usability of a decision support system. In case of a website it might be important to look at aesthetics and layout; in case of DSS factors such as safety might be more important. In both cases however, the core of the usability testing would be to ensure that users can execute their task without problem or discontent. The literature including the most often cited handbooks suggest that for better results inspection methods should be accompanied by user testing as all of these methods have different goals and measure different aspects of usability.

There is also an agreement that the employment of a single method is not efficient in discovering all of the usability problems, and combining several methods can bring best results (Koutsabasis *et al*., 2007). Well-recognized handbooks (Nielsen, 1994; Rubin & Chisnell, 2008) even suggested that at least one measure should be used for each of the usability criteria and that the choice of the criteria depends on the software's context of use.

This was observed in the reviewed literature, where the majority of the studies (23) combined two or more of the methods – with verbal protocol analysis (think aloud) and a user satisfaction questionnaires being the most often used.

The example of such is a study by (Ossebaard *et al*., 2012), who divided his evaluation of a health portal into three phases. First, he used semi- structured interviews to assess participants' computer

literacy. Scenario – based think-aloud followed, where users were asked to complete a series of tasks and were video recorded while doing so. Website navigation and content were then assessed using interview method, where a validated instrument (Website Evaluation Questionnaire) was used to gather data on subjects' experience of using the portal. The final phase was a focus group conducted to evaluate the main findings from previous sessions. Using such a mix of techniques resulted in several usability improvements recommendations for the website.

Another example is the evaluation of web portals by Granic *et al*., (2008), who used think aloud, satisfaction questionnaires and interviews with the users, and, in addition conducted a heuristic evaluation. This approach is in line with experts' recommendations (Jaspers, 2009), who suggest that using more than one method is more effective in identifying usability problems than by using a single method alone. However, as suggested by (Vredenburg, 2002), based on his study with 100 usability professionals, user testing and questionnaire methods are among most often used in practice due to the reliability of these methods.

Also, in the reviewed literature, no method was found to be significantly more effective than other, the opinion previously popularised by a number of prestigious usability experts (Jacobsen, 1999; Jeffries *et al*., 1991; Karat *et al*., 1992; Nielsen & Mack, 1994). Looking at advantages and disadvantages of each of the methods, the think aloud and questionnaire were seen as the most appropriate to obtain an in-depth understanding of the processes taking place when the software is in use.

In the reviewed studies, usability was commonly measured using efficiency, effectiveness and satisfaction as indicators; however, few studies extended these into using facets such as learnability or ease of use if the context of the study required that. A range of outcome measures

was collected in these studies, with time to completion and error rate being the most common of outcomes.

The findings showed that incorporating usability testing in the software development resulted in significant improvements. This however came at a cost as usability testing is time and resources consuming. As discussed above, all methods have distinct pros and cons, and ultimately the choice of the method is driven by practical concerns and study objectives.

Looking specifically at how the usability data was collected and analysed in the reviewed studies, there are few findings. Expert-based evaluations usually concluded with a list of usability problems, sometimes categorised by severity. Analysis of data from user tests was more laborious, with several stages separating data collection from final conclusions. In overall, user tests generated a rich set of audio and video data that in some cases were captured using specialised software. In some of the studies eye tracking equipment was used. The majority of evaluations were conducted in usability laboratories or users' place of work, but examples of remote testing were also present.

In the case of interviews and focus groups, authors followed a fairly standard approach to data analysis. The data was transcribed and analysed using a simple content analysis method. Very often descriptive statistics were also used to describe tendencies.

The most variation in approach to data analysis was found in think aloud studies. It ranged from informal analyses, based on the researchers' impressions gained from conducting the test, to a formal analysis. The choice of the method was not always justified. In the majority of the studies, verbalizations were transcribed verbatim and matched with users' actions or screens, if a separate camera was used to capture movement on the screen. Some authors used coding schemes,

however, the method to finalise them varied. Some used the usability categories derived from the literature (Ossebaard *et al*., 2012), usability problems detected by completing a walkthrough (Kushniruk & Patel, 1998) or codes describing what steps have to be undertaken to successfully complete the task (Benbunan-Fich, 2001).

Although Ericsson and Simon are the most often cited when using the think aloud method, divergence from their instructions for conducting such studies were observed in all of the reviewed literature. For example, the majority of the studies did not specify if and how participants were interrupted. Ericsson suggested that in case when a participant remains silent, a researcher would wait a predefined period of time, usually between 15 - 60 seconds, before he would remind the participant to talk using the simple sentence – 'Keep talking'. Any other direct prompt is not allowed. Divergence from theory in think aloud is a well-recognised problem in usability field, discussed in more detail by Boren (2000).

Furthermore, the literature suggested that the formal consideration of sample sizes for the studies using questionnaire or Think Aloud methods were rarely undertaken. They were based on pragmatic considerations such as the available budgets and the timescales of the projects.

4.10 Impact of the review findings on the design of the study

The findings of this review had several implications for the design of this PhD study. Firstly, learnability, efficiency, and user satisfaction were chosen as the major indicators of usability of the proposed model interface. Learnability was deemed important, as each user would only be provided with a Manual on how to use the tool. At no point a live support would be available to the users if the problems should arise. Therefore, there was a great emphasis on the fact that the tool should be intuitive and straightforward to use for both expert and non-expert users.

The model was created with the aim of supporting cost-effectiveness evaluations of various telehealth systems by the NHS and commercial organisations. The use of the model would be voluntary; therefore, if its quality was poor there was a risk that it would be abandoned. The success of the model and fulfilment of study objectives depended on its adoption by the telehealth practitioners. For obvious reasons, the current study was only able to engage with a limited number of professionals, mainly from the study sites. The tool has been promoted during various workshops and conferences, however, up until the release date the interest of the wider telehealth community in using this tool was unknown.

Secondly, for the evaluation of the proposed model interface a range of methods including cognitive walkthrough, verbal protocol analysis (think aloud), questionnaire and analysis of logs were deemed applicable. At the outset of the project, it was planned that software developers would conduct cognitive walkthrough as part of the software development process, and then share the main findings with the rest of the team. Therefore, this part of the evaluation is not discussed in this thesis.

User testing was seen as a practical and useful method for this study, particularly, that gathering data on end-user behaviour, their thoughts, feelings, working patterns and preferences was seen as crucial for the study. Verbal protocol analysis was seen as a method which could be successfully applied in this research despite the fact that currently there is no consensus on the best practise for conducting such a study and several challenges pertaining to data analysis had to be addressed.

An alternative method that could deliver that type of data is field observation, but it was seen as impossible to conduct for practical reasons. As the use of the model will be discretionary and voluntary, the team would not know when, and if ever, the model will be used in real life decision-

174

making. Also, if the model is used, it could be used at any location across the UK. Another issue of the importance was the timing of the model development, dictated by the MALT project milestones and deliverables and the work schedule of this PhD.

Another proxy selected for this study was user satisfaction. User satisfaction was deemed important, as both the research and theory suggest that people who are not satisfied with the use of a certain application would abandon it. Questionnaires as a method to collect data on user satisfaction were seen as a relatively cheap and easy way to reach future users of the model. Additionally, data on logs was to be analysed as a valuable source of data on end-users' actual behaviour.

4.11 Limitations

The presented review evaluated usability methods employed in 54 research studies. The list of included studies is not regarded as exhaustive, as the main purpose of the study was to gain an overview of methods used in the field, not to gather data from all conducted up to date research.

One of the limitations of this review is the fact that no quality assessment was employed in the process. Also, the inclusion and exclusion of the studies as performed by only one person is prone to evaluator bias.

However, the techniques identified are consistent with the seminal text books identified by specialists from the User Centered Healthcare Design stream of the South Yorkshire CLAHRC. Furthermore, this research builds on the reviews within those textbooks by providing a structured description of the context for each type of test, sample sizes, strengths and limitations (as shown in Tables 4 and 5). This provides a robust justification for the methods adopted in this study in the absence of clear recommendations for good practice.

5. Research methodology

Methodology is the foundation of a research study. It guides the methods used to collect and analyse the findings, and the perspective from which these findings are interpreted. This chapter accounts for the selection of 'mixed methods' as the most appropriate methodology with which to answer the research questions. I describe the factors informing the choice of methodology, the relevant research traditions, and the theoretical underpinnings of the methods used in this research.

5.1 My epistemological position

Denzin and Lincoln (2003) suggest that the researcher and her questions must impact the design of the research itself. The research process could therefore be broken down into five phases:

1) Locating the researcher within the history and tradition of research, including acknowledgements of the self in relation to the present study

2) Establishing the researcher's theoretical paradigm to guide the research

3) Selecting the research design that connects the theoretical paradigm to the methods of data collection

4) Selecting methods for data collection and analysis

5) Establishing methods to interpret and present findings

This chapter presents phases 1-3; whilst phases 4-5 are presented in Chapter 6, 8 and 9, respectively. I gained research experience from my work on randomised controlled trials and smaller-scale qualitative projects. I have always been more inclined towards the qualitative approach, as this offers a unique opportunity to explore the experiences of individuals.

My interests in telehealth started whilst working for a Polish medical consortium. I observed practical issues associated with efforts to establish telehealth in the region and to convince local health authorities and medical professionals to implement it. Prior to my experience in the MALT project, I was involved in diabetes research, where I helped to evaluate services for diabetic adolescents and learned about NHS structure and operation. Coming to the MALT project, and through background reading, I explored issues related to telehealth provision. In this way, I began to understand the approaches used to assess telehealth, and I improved my knowledge and skills in economic modelling.

These experiences have both directly and indirectly impacted this research in the following ways:

- I intend to focus on individuals' experiences, hence the research is predominantly qualitative.
- As this PhD project was conducted as part of the MALT study, the choice of methods was, to some extent, dictated by the aims and objectives of the MALT project, and has undergone several iterations during the project.

5.2 Research paradigm

It is important to state explicitly the philosophical underpinnings of this study, as these are related to the nature of knowledge and my understanding of the social world. The vital questions of 'what counts as a valid knowledge' and 'how it can be gained' also influence the choice of research methods (Hughes & Sharrock, 1997).

In broad terms, there are two main approaches to a scientific understanding of the world: positivism and interpretivism. Positivists seek to understand the cause and effect relationship in the researched phenomenon, and view the world as a collection of observable events and facts that can be measured (Gorman, 2004). Quantitative methodologies are commonly associated with

the positivist paradigm, as they hold that the social world is unaffected by the researcher or research process (Denzin & Lincoln, 1994; May, 2001).

In contrast, interpretivism posits that the social sciences deal with entirely different matters from those of the natural sciences, and must therefore use different forms of explanation (Weber, 1964). Interpretivism attempts to 'understand', rather than 'explain' the social world from the points of view of the actors involved. It holds that the researcher affects the social world, and thus focuses on the explanation of meaning rather than the cause of social phenomena, utilising qualitative methodologies (Denzin, 2003).

The epistemological stance adopted for this study, critical realism, is situated between these two main positions. An understanding of the world that is 'independent of human consciousness', and emphasises 'explaining that reality', brings it close to positivism. On the other hand, it adopts the interpretivist recognition that the social world can only be understood by exploring 'feelings' and 'thinking' of the people involved (Denzin & Lincoln, 1994).

In attempting to understand the nature of the information that I need my research to provide (Ritchie & Lewis, 1994), I adopted the critical realist view that reality is multi-layered (May, 2001; Sayer, 2000). I therefore committed myself to seeking to explain the mechanisms that underlie the social phenomenon, and to understand how the 'different layers' affect people's actions (May, 2001, p. 12). In other words, what is the 'causal logic' of the phenomenon being observed?

Finally, an important aspect of the critical realist approach is the use of theoretical frameworks to guide and interpret the relationships in this multi-layered reality (Sayer, 2000). They are employed as 'a device to reveal the structured reality beneath the surface' (May, 2001, p. 12),

rather than to test hypotheses. The benefit of using the frameworks in this way is the ability to demonstrate 'how the phenomenon is related to the whole' (Sayer, 2000, p. 25). As described by Sabatier (2007, p. 4), theoretical frameworks are 'helpful in organising one's thinking around the issues that are pertinent to a particular subject and those that could be safely ignored'. Action research frameworks served such a purpose in this study, and are described in the next section.

5.3 The theoretical framework

I sought different approaches to investigate how telehealth stakeholders used the MALT model in real life decision-making, and to reflect upon that which affects the adoption of models in the NHS. The dissemination of the MALT model was seen as a form of 'intervention'; not in the controlled, experimental sense, but in the sense that it was used to define the nature of the change required for a wider adoption of economic models and simulations. It was also designed to complement the numerous other interventions delivered in the wider MALT study (as discussed in Chapter 2). As such, the action research framework seemed appropriate for this research.

'Action research' is an umbrella term for approaches that aim to '[do] research with rather than on people' (Bradbury & Reason, 2003). It is succinctly summarised as follows:

> Action research is a participatory process concerned with developing practical knowing in the pursuit of worthwhile human purposes. It seeks to bring together action and reflection, theory and practice, in participation with others, in the pursuit of practical solutions to issues of pressing concern to people, and more generally, the flourishing of individual persons and their communities (Reason & Bradbury, 2001, p. 4)

There are a wide range of definitions of 'action research', as it is used in a number of different disciplines, including operations management (Coughlan & Coghlan, 2002), social sciences, and health services research (Hart & Bond, 1995). The knowledge generated by action research is practical and context-specific (Coghlan & Casey, 2001), and it is used to implement change or

innovation in a complex system or in problematic working practices (Coughlan & Coghlan, 2002). Each action research cycle includes diagnosis, planning, action, and reflection as illustrated in Figure 28 below.

Figure 28. Action research cycles
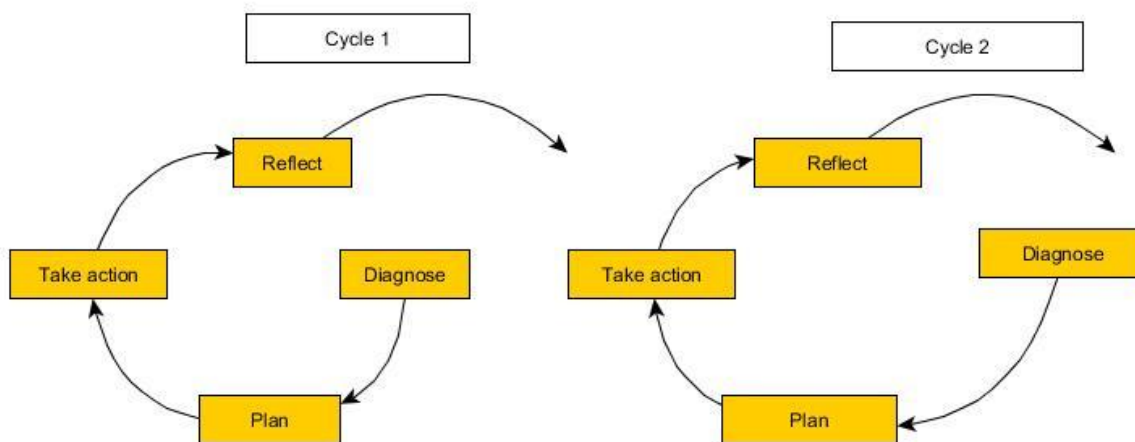
In the first two stages, the researcher assesses the need for change, and based on these observations, identifies the nature of the required intervention. Once the intervention is implemented, its impact is assessed and reflected upon. Because of the iterative character of action research, the fourth stage of the research cycle forms the first stage of a new cycle, so that the need for change and the definition of an intervention are refined over the course of cyclical processes.

5.4 The initial research design and design evolution

As detailed in Chapter 1, this research is predominantly focused on the question of how to facilitate the use of economic models for evaluations of telehealth interventions. In designing the study, I selected the paradigm and approach that I felt would most effectively answer my research

question. I made the decision that the study should use predominantly qualitative methods, as my aim was to understand the needs and experiences of study participants rather than to test hypotheses or theories. A qualitative approach was considered appropriate to develop an understanding of stakeholders' requirements for the model, and to explore the experiences of MALT model users. However, quantitative methods were deemed appropriate to evaluate some aspects of model use and user experience that required numerical data.

Initially, the study was designed to be conducted in three distinct stages, with qualitative and quantitative elements serving as both complementary parts of the whole study, and as standalone studies in their own right. The first part of the study was to inform the design of the economic model. The second part of the study was designed with both qualitative and quantitative elements to test the usability of the model's interface design. The plan was to synthesise the findings from both qualitative and quantitative elements to inform the model's improvements. The third part of the study was then planned to utilise mainly quantitative methods to test end-user satisfaction. As such, the study was designed as mixed methods, with dominant qualitative elements.

This design was modified at two stages during the research process. The first change occurred when the preliminary analysis of the quantitative data collected during usability testing found the results were not meaningful enough to support the aims of the study. The second, more significant change occurred when an analysis of the results of the satisfaction survey revealed a poor response rate. Given that the survey was designed to collect data on MALT model use from 'real' users, the absence of a significant number of responses affected the whole third stage of the research. The decision was taken to conduct additional qualitative interviews with registered users and to abandon the idea of testing EUCS instrument suitability for the health economics field.

Both changes meant that the overall design was changed from a mixed methods study into a predominantly qualitative study. These changes occurred over time, something one might have anticipated in an action research project. Reason and Bradbury (2001) suggested that action research is an emergent process, and as such there is no need to provide a fully detailed research plan at the outset of the project, nor to adhere rigorously to it. Instead, the researcher should be flexible enough to embrace change and act upon it as he goes along.

The new, emergent design is illustrated in Figure 29 below. The dissemination of the MALT model in the 'end-user mode' was seen as the main action research intervention. Each of the research phases (e.g. qualitative interviews, usability evaluation, end-user satisfaction evaluation) were seen as separate action research cycles. They comprised diagnosis, planning, acting, and reflection phases. Reflection was the most critical of the phases and had an impact on consecutive stages and other research activities.

Figure 29: The emergent design of the study
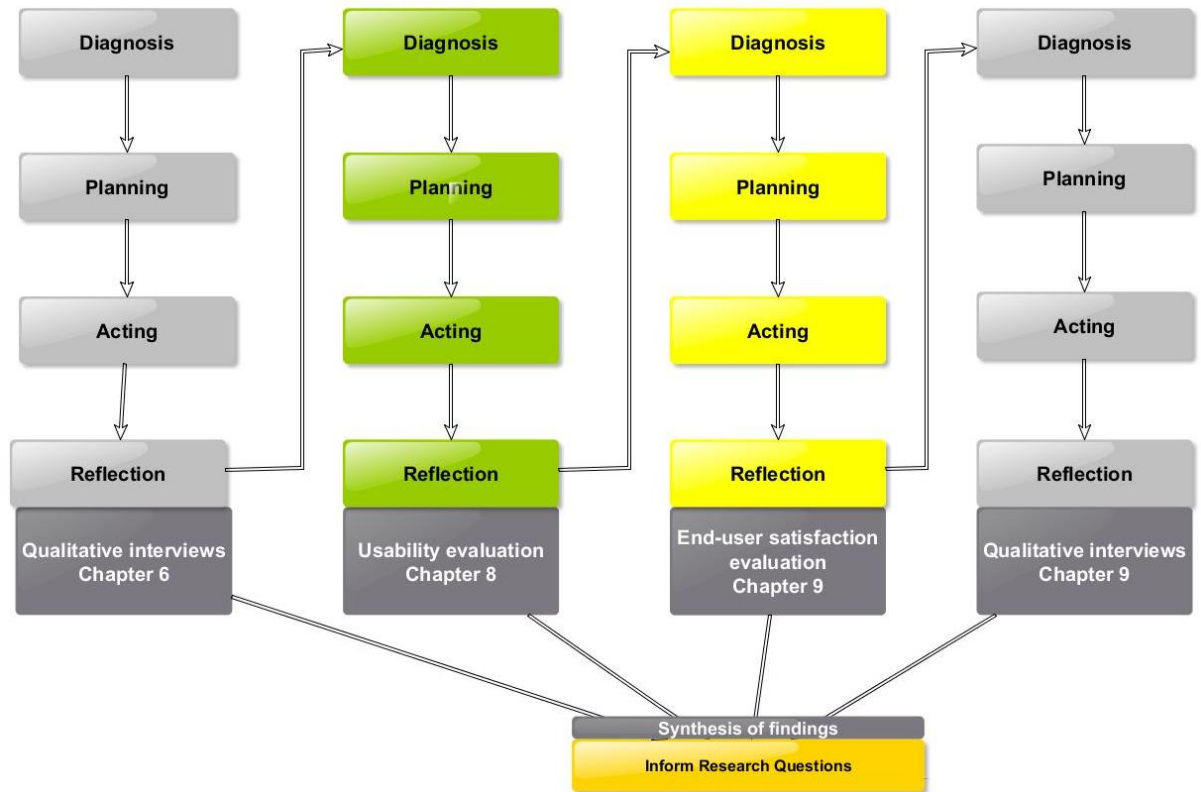
The collaborative and practical character of action research has always been dominant, as emphasised by Denscombe (2002). The benefits of using the action research framework include the fact that context-specific knowledge was generated through collaboration with a wide range of stakeholders. Another distinctive benefit is that my understanding of the barriers to adoption of modelling in the NHS progressed during each of the action research cycles. By the end of the first cycle, my understanding was that the standard approach to dissemination of models might not be appropriate for telehealth evaluations, and that a 'new' approach was required. Conducting the second cycle meant refining my understanding of the value and requirements for usability testing. The third cycle was critical to my understanding of the barriers to modelling use. After completing all four action research cycles, I was able to make recommendations for modellers in

terms of overcoming barriers to modelling and simulation adoption in the NHS and disseminating

models for telehealth evaluations.

6. Qualitative interviews to elicit stakeholders' requirements for the MALT model

This chapter reports the methods and findings of the first section of the case study, which sought to elicit the telehealth stakeholders' requirements for the proposed model. The interviews formed a part of the conceptual modelling described in Chapter 2.

The chapter begins with a description of the study methods and sampling strategy. Issues around the quality of qualitative research and the role of the researcher are discussed. Following that, Section 6.5 presents the main findings of the qualitative interviews and the implications of them for the model development. The chapter concludes with a discussion of the strengths and limitations of this research.

6.1 Rationale

Qualitative research is often used to improve knowledge and understanding of how individuals perceive the world (Maxwell, 1996) and this project used qualitative interviews to elicit stakeholders' requirements for the MALT decision model. Specifically, the approach addressed the main research question:

1) What are the stakeholders' preferences for proposed model inputs, outputs, and scenarios?

The data generated in this research directly informed the development of the proposed model.

6.1.1 Method Selection

Two approaches to collecting qualitative data were considered when designing this research: a focus groups and qualitative interviews.

The focus group method sees the perceptions of, and thoughts and opinions on a specific topic collected from a group of people (Bryman, 2008). Barbour and Schostak (2005, p. 43) describe the focus group as 'a social process through which participants co-produce an account of themselves and their ideas which is specific to that place and time'. The character of focus groups means that they are well-suited for capturing interaction between participants, and group dynamics (Stewart & Shamdasani, 2014). This aspect is an advantage when applied to specific types of research, e.g. in organisational learning (Finch & Lewis, 2003). Focus groups are also appropriate for studying the preferences of large groups of people, e.g. patients, clinicians or software end-users, and as discussed in Chapter 4, have been successfully employed in several usability evaluations (Ebenezer, 2003; Lee & Lin, 2014; Leslie *et al*., 2006).

One of the constraints of the focus group is that it requires a skilled moderator who can intervene to facilitate discussion, whilst remaining 'the observer', giving each member of a group an equal chance of expressing their opinions (Stewart & Shamdasani, 2014). Achieving the right balance between an 'active' and 'passive' role is particularly challenging (Henderson, 1995; Stewart & Shamdasani, 2014). The moderator skills are also important as there is risk that people with stronger characters, or those who are more articulate, can dominate those less capable of expressing their views. The influences of intrapersonal factors on group dynamics are widely recognised as limitations of a focus group (Barbour & Schostak, 2005; Henderson, 1995; Stewart & Shamdasani, 2014). Focus groups also generate a wealth of data that can pose challenges for transcription and data analysis later on, e.g. multiple people speaking simultaneously.

Another practical implication is that a focus group requires all participants to be physically present in one location in a certain time (Stewart & Shamdasani, 2014). This aspect I found especially difficult due to the number of participants invited to take part and the fact that they were located in different regions of the country.

Because of these constraints, I made the decision that an alternative primary data collection method (qualitative interviews) would be more suitable for this study. The interview method is effective for gaining insight into peoples' values, opinions, and beliefs (Bryman, 2008) and researchers can choose between in depth, unstructured, and semi-structured approaches.

In depth interviews 'combine structure with flexibility […] by permitting topics to be covered by the order most suited by the interviewee' (Ritchie & Lewis, 1994, p. 141). As such, the process is perceived as a type of a normal conversation, with the researcher steering it to achieve a certain purpose (Morris, 2015). In depth interviewing requires a lot of preparation and skills (Kvale, 2008; Patton, 2002; Serry & Liamputtong, 2013).

In contrast, semi-structured interviews use a pre-specified set of questions with a specific wording and order, but allow researchers to deviate from them to follow the trajectories of the conversation (Ritchie & Lewis, 1994). In this way, a two-way conversation is created during the interview, where respondents can discuss any issues that were not taken into account and express their views freely (Silverman, 2000). This aspect was especially valued in this research, as the participants' modelling experiences and levels of familiarity with the decision tools were not revealed until the interview. After the initial introductory questions, I expected to deviate from the detailed questions and rephrase them during the interview. I also intended to give each participant a chance to express their opinions, no matter how different they seemed. The most often recognised disadvantage of the semi-structured approach is that each interview must be carefully planned, as they are frequently time-consuming and intensive (Kvale, 2008).

6.2 Study design

6.2.1 Setting

This qualitative study was set in the Yorkshire and Humber region of the UK. The qualitative interviews were conducted between August 2012 and January 2013, with stakeholders in the MALT project. The interviews were conducted in three of the four MALT study sites and with representatives of the telehealth industry nationwide (manufacturers of the equipment and telehealth service providers).

6.2.2 Sampling strategy

Interviewees were recruited within the MALT study sites using several sampling strategies. Firstly, a purposive sampling strategy was used to recruit senior NHS managers and Clinical Commissioning Group (CCG) staff. They were initially approached by the manager of the MALT project, informed about the purpose of the project, and asked if they would be interested in taking part. Following this, I contacted each person by telephone and arranged a suitable date for the interview. The identified participants included senior management staff involved in decision-making, and people in charge of telehealth programmes or working to establish telehealth in the region. They were recruited due to their specific experience.

Secondly, a snowballing sampling strategy was adopted for the recruitment of telehealth providers (Bryman, 2008). The members of the MALT industry panel (senior managers and directors of major UK telehealth companies) were approached in the first instance. Additionally, representatives of the companies providing telehealth-related services, such as triage or training (so-called 'middleware' companies) and other staff members, were identified through these

interviews and approached for interview. Due to the nationwide locations of the interviewees, the interviews were conducted over the telephone.

There is an ongoing debate as to whether telephone interviews are a valuable data collection method (Legard *et al*., 2003; Rubin & Rubin, 2005). This mode has the obvious drawback that the interaction with the interviewee is limited, and non-verbal communication and cues are lost to the researcher (Gillham, 2005). As this can impact understanding of the discussed issues, the telephone mode is generally not recommended appropriate for in-depth interviewing (Legard *et al*., 2003; Morris, 2015). Conducting interviews over the telephone also means that the researcher and the interviewee must maintain concentration, with studies suggesting that these interviews tend to be shorter due to the participants' fatigue (Gillham, 2005).

An obvious advantage of telephone interviews is the convenience of the researcher and participant, due to the avoidance of the necessity of travelling (Irvine, 2011). This also means saving time, and the possibility of conducting several interviews in one day. Telephone interviews are recommended in the case of people resistant or unable to take part in face-to-face interviews (Irvine, 2011). Due to their increasing respondents' perception of anonymity (Irvine, 2011), they have been successfully used in many studies on sensitive topics (Drabble *et al*., 2016; Johnson *et al*., 1989; Stephens, 2007).

In designing this research, all of these constraints were considered. The potential risk to the successful completion of the interviews and time constraints ultimately led to the decision to conduct telephone interviews with industry members.

The overall study target sample size was a minimum of 20, covering a representative from each stakeholder group in each of the sites, as outlined in Chapter 2. The sample size was determined

by practical issues, as engagement with the representatives from some of the identified groups was problematic. Busy workloads and professional commitments were the main reasons given by those who were approached and refused to take part in the study. My recruitment decisions were impacted by MALT study timescales and other pragmatic concerns, such as time, cost, and staff workload. Of the 35 people approached, four refused to take part.

6.2.3 Topic guide

A semi-structured interview topic guide (Appendix A.4) was developed to facilitate discussion with the interviewees and ensure consistency of topic coverage across all interviews. The topic guide was devised following two literature reviews: the review on cost-effectiveness of telehealth in COPD, and CHF interventions and consultations with key members of the research team and stakeholders. Additionally, in the process of finalising the topic guide, I attended several project meetings, where I discussed each question with the research team, and separately with my PhD supervisors.

The final version of the interview topic guide included two introductory questions about participants' experiences of running or working in a telehealth services, and their thoughts and beliefs about telehealth cost-effectiveness. More detailed questions regarding requirements for the economic and financial models followed, focusing on the types of costs telehealth systems are likely to generate and the categories of benefits they would expect. Participants were also asked to reflect on the types of analyses they would like to perform to guide their decisions. The interview was concluded by asking about participants' experiences of financial planning tools and the potential application of such in their organisations. Several prompts were designed to facilitate the discussion and to encourage the respondents to enlarge on their viewpoints, as opposed to a 'question and answer' approach.

6.2.4 Pilot interviews

The topic guide was tested during three pilot interviews conducted in May 2012 with senior NHS and Primary Care Trust (PCT) staff, and a senior representative of the telehealth industry. The pilot revealed the need to make a number of changes to the topic guide.

Of key importance was the feedback from the interviewees that cost-effectiveness and financial models were understood differently by people outside academia, and that future interviewees would need a better explanation of what is meant by these terms. Importantly, the pilot showed I needed to provide an explanation in plain language prior to the interview, avoiding specific economic jargon.

Because of the number of changes required to the topic guide, data from the pilot interviews was not incorporated into the main analysis. The interview topic guide was amended, incorporating changes suggested by the participants to the pilot, and the amendment to the ethical approval was granted.

6.2.5 Interview organisation

The face-to-face interviews were arranged in a place and time convenient for participants, which in each case was their usual place of work. Following Ritchie and Lewis's (1994) recommendations, the interviews were scheduled to last up to an hour: long enough to cover each topic without the risk of the participants losing interest.

However, due to participants' work commitments and other issues, the actual length of the interviews varied considerably, the shortest lasting 26 minutes and the longest 73 minutes. Five interviews were conducted jointly with another researcher from the MALT team. This was

necessary because due to the participants' busy schedules and limited time available. It was therefore agreed that LM would cover general questions about telehealth, and ask one pertaining to business models, and I would use the remaining time to discuss cost-effectiveness issues and model requirements (items 3-10 from the Interview topic guide Appendix A.4).

The telephone interviews were arranged on a day and time convenient for the participants. The shortest interview lasted 20 minutes and the longest 58 minutes.

6.2.6 Data analysis method

Framework analysis was selected as a data analysis method (Ritchie & Lewis, 1994; Ritchie & Spencer, 1994). The method originated in policy research and has since been used in various health-related research settings (Ward *et al*., 2013). The advantage of the framework method is that it focuses on describing and interpreting what is happening in a particular setting, rather than developing theories (Srivastava & Thomson, 2009). Researchers suggest that the method is especially valuable for research focused on resolving issues identified a priori that have a limited timeframe (Ritchie & Lewis, 1994; Ritchie & Spencer, 1994).

The framework approach was seen as particularly useful in this study for highlighting similarities and differences between stakeholder groups. The way it synthesises findings was found useful for identifying patterns and generating a list of requirements to be passed onto modellers. A large amount of qualitative data was generated from the interviews, and the framework allowed me to summarise key issues in a systematic and transparent way (Ritchie & Spencer, 1994). The process of classifying and organising the data into themes and sub-themes was informed in the first instance by the interview topic guide, and consisted of five interrelated stages as illustrated in Figure 30.

Figure 30. Stages of the qualitative data analysis.
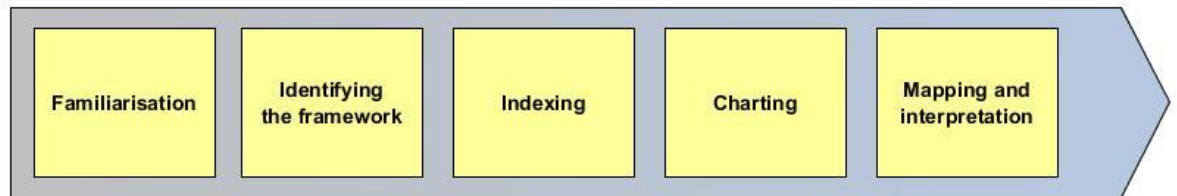
1) Familiarisation

The first step in the analysis includes familiarisation with the data. The aim of this stage is to identify initial themes and concepts and to gain an overview of the data set (Ritchie & Spencer, 1994).

2) Identifying the framework

This phase involves applying codes to parts of the transcripts in order to detect reoccurring patterns. These patterns can be classified into themes that later will form a part of a framework. The process is dynamic in nature.

3) Indexing

In this stage the framework is systematically applied to the data set to code each passage of the transcripts. Any new emerging topic is categorised and added to the framework.

4) Charting

The aim of charting is to summarise and present the key findings. It is advised that charting is completed without losing the context and content of the data. This is ideally achieved by shifting the data from its original transcript and categorising according to themes or cases.

5) Mapping and interpretation

In this stage, thematic charts and visual models of the main themes are used as a tool to compare and contrast perceptions and accounts, search for patterns and seek an explanation of these within the data.

6.3 Data collection

Overall, between August 2012 and January 2013, I conducted 19 face-to-face and ten telephone interviews. The number of participants in each of the stakeholder groups and their specific roles are presented in Table 9.

Table 9: Participants of the qualitative interviews

DoH	CCG	Providers			Industry		Local council
		Acute Care	**Community Care**	**Primary Care**	**Managed service providers**	**Middleware companies**	
Programme Manager	*Commissioning Lead* *Commissioning Manager* *Head of Transformation*	*Consultant*	*Programme Manager* *Operational Manager*	*GP*	*Director of Strategy* *Operational Manager* *Managing Director*	*General Manager*	*Project Manager* *Lead for Technology*

DoH	CCG	Providers			Industry		Local council
		Acute Care	**Community Care**	**Primar y Care**	**Managed service providers**	**Middlewar e companies**	
	Telehealth co-ordinator						
Number of participants							
1	*5*	*1*	*6*	*3*	*6*	*3*	*4*

Data management

Over 13 hours of interview data were collected in the study. The Nvivo (QRS International., 2012) software package was used in the process of data management. To preserve the confidentiality of participants, throughout this chapter and for the remainder of the thesis, the following general terms are used in place of interviewees' first names: NHS staff, NHS provider, NHS manager, member of industry, telehealth provider, GP.

Recording and transcribing the data

The interviews were recorded using a digital voice recorder. Using recordings during interviews allowed me to focus more deeply on listening to participants, and I also took field notes. Later on, the recordings facilitated transcription and a more detailed analysis, as I was able to go back to the data and listen to participants' own words and explore nuances in their accounts. Details of the processes of storing and dealing with the recorded data are explained in Section 6.7.

Several researchers assert that transcribing is an integral part of data analysis (Duranti, 2006; Tessier, 2012). This is based on the assumption that once the interview is completed, any process that involves data becomes the part of the analysis (Tessier, 2012). Having agreed with this assumption, I made the decision to undertake verbatim transcription of the interview data myself. I believed that, in doing so, transcribing the recordings would ease the process of familiarisation with the data and facilitate the identification of themes that would form the basis of the framework.

All of the recordings were transcribed within a week of completing each interview. All the data were anonymised to ensure data protection, with participants' names replaced by a code indicating

the date the interview took place and the study site (e.g. AA200613). Any identifiable information (such as the name of the Primary Care Trust) was removed. After completing the transcription, the accuracy was checked by listening again to the recording.

The data were coded using Nvivo software (QRS International., 2012), and the main themes grouped into tree nodes and free nodes. Several graphical project maps were created to illustrate the relationship between themes and to explore the link between the data sets. The advantage of the software was that, once indexed, codes could be grouped together and retrieved easily, thus minimising the effort needed to manage such a vast data set. The software also allowed me to maintain a direct link between my interpretations and the original interview data by the use of memos and queries.

6.4 Stages of data analysis

Familiarisation

The process began when I had transcribed the interviews and written up the field notes. The transcripts were checked for accuracy, and if needed, field notes were referred to for additional information. I also asked an independent researcher to check the transcripts for accuracy.

Identifying the framework

At this stage, codes were used to label similar material across transcripts (e.g. 'barrier to risk sharing'). The groups of codes on the same topic (e.g. 'telehealth provision models') were then created. The initial codes derived at the beginning of the project were informed by the topic guide and refined throughout the process of coding. By cross-referencing multiple transcripts, these codes were refined. For example, after completing the initial four interviews, I identified 42 codes.

The number of codes grew to 72 with subsequent interviews, and eventually allowed me to identify all of the themes within the data that formed a 'provisional' framework. The coding framework was reviewed by my PhD supervisors, and it was decided that abstract codes were also needed to reflect the data as the whole. A number of generic codes were thus generated for the data where respondents explained the process of telehealth implementation in their organisations, their thoughts on the issue of cost-effectiveness of telehealth, and the general requirements for a financial planning model.

This led to the 'final' framework, which consisted of a total of 85 codes grouped into the 12 themes listed below to reflect the collected data. Each of the main themes included two further levels of coding. The full framework is presented in Appendix A.5.

- Rationale for telehealth implementation

- Indicators of telehealth effectiveness/efficiency

- Reliability of available data

- Identified reasons why the technology did not generate savings

- Categories of cost

- Methodological challenges in estimating costs

- Indicators of severity of COPD disease

- Perceived benefits of using telehealth

- Perceived application of the proposed model

- Requirements for the tool

- Analyses stakeholders would like to perform

- Approach to financial planning tools

Indexing

The framework was systematically applied to the data to code each passage of all transcripts. If a new topic emerged, it was categorised and added to the framework. The process took place entirely electronically, using Nvivo 11 (QRS International., 2012). A matrix of codes was created to represent all themes and sub-themes.

Charting

While charting is not essential to present the results of a qualitative inquiry, I felt that it was a useful way of visualising the data and highlighting differences and similarities in the study findings (Ritchie & Spencer, 1994).

Charts were constructed around the first and second level of coding, following the method recommended by (Ritchie & Spencer, 1994). My approach was to demonstrate the differences in stakeholder requirements for the model inputs, outputs and scenarios, as opposed to presenting accounts of each individual person.

I followed Ritchie and Spencer's (1994) recommendations to review the charts, and then create visual models (Picture 1–3) and research notes to explain the main concepts and associations, and to provide explanations of the findings.

6.5 Data saturation

In qualitative research, data saturation is achieved when no new data emerges, the themes are well developed, and the relationships between themes are established and validated (Strauss & Corbin, 1998). After analysing the data from the telephone interviews and those conducted on the two study sites, it became clear that the saturation point had been reached as themes started to reappear, and no new themes could be identified. At this point, 25 interviews had been conducted, but I continued with the final four interviews. This confirms the findings of Guest *et al*. (2006), who suggested that in qualitative research saturation occurs at 6-30 participants.

6.6 Ensuring quality in qualitative research

The recommendations of Guba (1994) were followed to ensure the methodological quality of this study. This was achieved by fulfilling the requirements for credibility, transferability, dependability, and confirmability.

The credibility of qualitative findings refers to the correspondence between the informant's perception of a social construct and the researcher's portrayal of their viewpoint (Mertens, 2005). In this study I used constant comparison to refine existing themes and develop and incorporate new topics and subtopics emerging from the data while finalising the thematic framework.

To allow for the findings to be generalised beyond the MALT project, I provided a detailed description of the methods used and the research settings. This is considered by the literature to be a requirement for transferability (Guba, 1994).

The dependability of the qualitative findings can be defined as the level of agreement between researchers on interpretation of the data (Lewis, 2009).To enhance the dependability of the

findings, a series of discussions with my PhD supervisors were held during supervision meetings to agree on the coding, the final version of the thematic framework, and interpretation of the findings.

Finally, the confirmability of qualitative findings relates to the extent to which the findings are data- specific, and the extent to which they are influenced by the researcher (Guba, 1994). This last criterion was addressed by reflecting on ways in which the findings may be influenced by the researchers' assumptions and experience (Section 6.7).

6.7 Role of the researcher

Traditionally, positivists viewed the validity of research as depending on the objectivity of the researcher (Guba, 1994). Within qualitative research, however, there is agreement that the role of the researcher is to interpret reality, and that each of these interpretations is equally valid (Guba, 1994). Therefore, instead of focusing on eliminating the impact I, as the researcher, may have on the research findings, I used the process of reflexivity to understand the impact (Holliday, 2007).

A researcher might impact a study during data collection, and later on during data analysis. I might influence my interviewees with the way in which I formulate the questions or in which we interact during the interview. During the process, I had to be aware of my own responses and encourage myself to focus on the perspective of the participants. That aspect of qualitative interviewing I found especially challenging, though it became easier with practice. For example, I quickly learned to remain silent and refrain from nodding when interviewees mentioned something I would regard as a good answer to a question. I also sought to make sure that they had finished their point before I began to speak. I was also mindful of the way in which the data was analysed and interpreted, bringing to the study my knowledge and beliefs gained from previous experiences (Lewis, 2009). My previous experience of working in the NHS as a data analyst, gave

me insights into NHS procedures and the practical application of data analysis, allowing me to form an opinion on operational issues and develop an explanation of why problems related to data exist.

I also acknowledge that I brought to the study my knowledge of cost-effectiveness modelling, gained during my Masters and PhD studies, and with this my preconceptions of what the standard model includes in terms of inputs, outputs and scenarios. My training in health economics also meant that during data analysis I approached the data from a particular angle.

I remained conscious of these influences throughout the research process. As previously stated, my ontological position is that there is no absolute truth and no knowledge we can be certain of. As a researcher, I accept that my assumptions and explanations about the phenomena under study could be wrong, and that my impact on the data could be the source of bias (Guba, 1994). Therefore, I should acknowledge that I am presenting one of many possible interpretations. To acknowledge my active part in this research project, I describe myself in the first person rather than as 'the researcher'.

6.8 Recruitment and ethical issues

The study received ethical approval from the UK National Research Ethics Committee. Access to individual sites was permitted under NHS governance arrangements. To protect the anonymity of the research sites, copies of the approval letters are not presented in this thesis, but are available on request from the MALT project manager. The interview procedures incorporated several measures to ensure that ethical criteria were met.

Firstly, participants were informed about the study by the MALT project manager, who passed their contact details to me if they expressed an interest to participate. For the interviews conducted

over the telephone, the participants were provided with the information sheet by email prior to the agreed date of the interview. The information sheet detailed the purpose of the study, the interview process, and how the data would be used. Participants were also provided with the contact details for my lead PhD supervisor in case of any doubts. A copy of the information sheet is presented in Appendix A.6.1. On the day of the interview, I read through the consent form statements (Appendix A.6.2) and provided each participant with an electronic copy for signature.

I followed a similar procedure for the face-to-face interviews. Potential participants were first contacted by the MALT project manager. If they expressed interest in taking part in the research, they were provided with the information sheet (Appendix A.6.1) for more about the study and to decide if they wanted to be involved. I followed up a couple of days later to agree the details of the interview. On the day, participants were asked to read the information sheet again, and were given the opportunity to ask questions. They were then shown a consent form (Appendix A.6.2) and asked to complete three copies: one to keep, one for the PhD study documentations, and one for MALT study records. They were informed that participation in the study was voluntary, and that they were free to withdraw at any point. All gave consent to record the interviews.

Secondly, the confidentiality requirements were fulfilled by removing all identifiable data, such as interviewees' names, contact details, and the name of the PCT Trust and place of work. The participants' identities were known only to members of the MALT research team. The data were securely stored in a locked filing cabinet on the premises of the University of Sheffield.

The interview was designed to minimise the risk of distress for participants, and to make full use of their time. The initial contact with potential participants was made by the MALT manager. This was to minimise repetitive calls on the same matter and to facilitate multiple interviews with

different members of the MALT research team. In the event of a lack of response, people were contacted only once to obtain confirmation that they did not wish to participate.

To minimise the burden for the interview participants, and to make effective use of their time, the MALT project manager decided that I should conduct my interviews parallel to the interviews of the business modelling workstream researcher. That would ensure that participants were approached and went through the interview procedure only once. To meet this requirement, we met regularly before each interview and planned the conduct of the interviews for each participant. This meant that the interviews were conducted together with the other MALT researcher, taking into consideration overlapping questions and time constraints. I kept recordings from these joint interviews and included that data in this study.

Additionally, there were safety procedures employed to ensure data protection. The data was stored on an encrypted disc in a safe cabinet on university premises. After transcription, the interview audio recordings were destroyed and the remaining data was archived for a period of three years following the PhD completion.

Finally, safety precautions were undertaken to ensure the safety of the researcher whilst conducting interviews on premises outside the University of Sheffield. This included providing details of the interview to a contact at the University of Sheffield, and sending text messages on the day, both prior to and after completion of the interview.

6.9 Findings

The findings from the interviews are presented in the section below, categorised under the following key headings:

- Rationale for telehealth implementation

- Indicators of telehealth effectiveness/efficiency

- Reliability of available data

- Identified reasons for why the technology is not generating savings

- Categories of cost

- Methodological challenges in estimating costs

- Indicators of severity of COPD disease

- Perceived benefits of using telehealth

- Purpose of using the model

- Requirements for the tool

- What analyses stakeholders would like to perform

- Approach to financial planning tools

The interviews generated an abundance of data related to telehealth implementation and service models, which was reflected in the themes generated. Some of these were not directly relevant to the development of the model, but provided very useful data for the wider MALT project.

6.9.1 Rationale for telehealth implementation

Participants' explanations of who was involved in decisions on telehealth implementation, and how the decisions to use telehealth within services were made, were discussed at the beginning of the interview.

Many of the informants pointed to the dominant role of the Department of Health policies and incentives for NHS organisations in guiding decisions to adopt telehealth technology. They explained that the Department of Health perceives telehealth as a significant means of saving limited resources and dealing with rising demand for healthcare. They also acknowledged the QIPP Agenda (Quality, Innovation, Productivity and Prevention) has a direct impact on the decision to provide telehealth services in their organisations. QIPP is a broad policy designed to drive quality improvements in the NHS at the same time as making efficiency savings. However, as suggested by the programme manager in one of the sites, very often technology was purchased because of this national incentive, but services on the ground were not adapted to benefit fully from telehealth. He stated:

> *I think that is how it started in* [coded]. *I think some of the developments have been driven by national directive.* […] *However, it seems to me that whoever has some spare money is buying telehealth because the Department of Health is promoting it heavily. It doesn't mean they are doing things properly.* (Programme Manager, Site A)

All of the GPs interviewed felt that the policy and policy-led initiatives were important to undertake if they lead to improvements in patient care. Whilst they felt that including telehealth in the QoF (a payment framework that incentivises organisations to meet quality-related goals) was likely to improve telehealth adoption rates across the country, the majority of the organisations were not ready to deliver these services appropriately. QoF is a scheme for GP practices that rewards the undertaking of activities to improve the quality of patient care. This means that to achieve better scores and reimbursement, GPs are incentivised to offer telehealth as part of their services. As one of the GPs mentioned:

Well, if I said to you, as a practice, we seem to take on most new initiatives -- we took it on as a pilot, you know, one of my colleagues had heard about it, and you know, it was being offered, and we thought we would give it a go, short of not knowing much about it, and see what happens. So, that's how we got into it, and I guess we set it up and we haven't withdrawn from it because we just continue. […] I think all the other practices have fallen away at the moment, but I do see resurgence and more people being interested in using it in the future, because I think from 1 April it's going to be part of the quality and outcome framework. (GP, Site C)

Some of the GPs expressed that this would be forcing some of the practices to rush the implementation of telehealth in order to meet the target, rather than to prepare a proper plan to enable patients to benefit fully from the technology.

The industry staff also considered the Department of Health policies to be facilitators of wider adoption of telehealth technology in the UK. They pointed to the fact that in order to sustain the growth and popularisation of telehealth, more national initiatives were needed. Many industry members mentioned the need for a change in contracting, where the industry would be paid by the efficiencies achieved in the healthcare system thanks to use of their technology.

Other examples of implementation practices included accounts from one of the study sites that telehealth technology had been used as a result of a research study be conducted and the equipment being left in the organisation after the trial had been completed. Interviewees from this site perceived this as an obstacle, explaining that the full benefits of the technology had never been realised as the service continued to operate on a trial basis. For example, the inclusion and exclusion criteria for patients had not been revised. The staff shared their concerns that the

technology would become obsolete in few years' time and that the service would need to undergo major change.

There was also an example of an approach where the decision to implement telehealth was made by members of the commissioning board after reviewing the evidence from the literature. This encouraged staff to view telehealth as a long-term investment.

6.9.2 Indicators of telehealth effectiveness/efficiency

The NHS managers agreed that telehealth played an important role in provision of services for patients with long-term conditions, and expressed the view that telehealth should be seen as a part of community services and not a standalone intervention. This meant that it was often difficult for them to delineate the outcomes of using the telehealth service as it was not delivered in isolation. There was an agreement amongst NHS managers that the lack of clarity about the impact of telehealth service made it difficult to support future service developments. NHS staff from two study sites identified a need to evaluate the effectiveness of the current services, as their benefits were informally recognised rather than systematically evaluated. There was a recognition that telehealth led to a lower number of hospital admissions, but all managers agreed that the evaluation was not comprehensive.

Interviewees provided examples of the indicators that they considered important for assessing if a telehealth system was beneficial for patients and service delivery. Among the NHS staff interviewed, perception of the efficiency and effectiveness of telehealth were described in relation to directly observable use of resources, such as hospital bed occupancy, Accident and Emergency (A&E) visits, and quality of life benefits for patients. NHS managers felt that the best evidence of telehealth success was provided by an observable reduction in hospital admissions. This view supported commissioners' and industry members' views.

I think the key aspect that was a driver for the PCT was cost. We all know that increasing hospital admissions was a big issue, but we also know it's about the patient as a whole in terms of what's best for the patient in terms of improving health. So, it's not only purely about trying to reduce costs, but it's also trying to find out what's good for the patient in terms of their quality of life. We attempted to look at the cost-effectiveness aspect of it from that point of view. There were of course secondary outcomes things around reducing length of stay, patient satisfaction. Carers and users as well: they tend to be side-lined in the process of implementation. So, those are kind of the secondary outcomes measures.
(NHS manager, Site B)

It was clear that views about telehealth effectiveness were surrounded by concerns about the quality of the evidence used to support the claim that 'telehealth works'. This uncertainty was seen as a major barrier to wider technology adoption.

One commissioning lead felt that there was a need to challenge existing perceptions of telehealth effectiveness. He stated that the most appropriate indicator of effectiveness would be evidence of changes in the pattern of care, such as a reduction in emergency care and an increase in planned care. He argued that the number of admissions alone would not reflect the impact of telehealth, as the number of people with COPD is not likely to change dramatically following the introduction of technology. By influencing self-management and medication compliance in the population, in the longer run telehealth would eventually decrease the incidence of emergency care and hospital admissions.

If the system works fine you should see the changes in suboptimal care – people going to see doctors when planned, not unplanned. So, huge reductions in use of A&E.
(Commissioning Lead, Site A)

This view was echoed in the accounts of industry members. For example, a senior member of staff from a telehealth company stated that he would look primarily at a reduction in the length of hospital stay as an indicator that telehealth works for a certain population.

> *We generally look at reduction in unplanned admissions, because you have to figure out a way to screen out the elective admissions, bed days, length of stay. You should see big changes there if TH is working effectively. What you should see, even if you have been admitted, you should see significantly fewer days in hospital.* (Director, telehealth company)

This view was reflected in the accounts of NHS providers and commissioners as well.

6.9.3 Reliability of available data

Concerns about the quality of the data and robustness of the evidence were the key issues raised by many NHS and industry informants. As illustrated by one member of the industry, these concerns were identified as a potential barrier to wider adoption of technology. For example, the managing director of a telehealth company stated:

> *Because you can get certain information out, but generally you have holes all over the place and the information you might find might be outdated, you have to update it. It is very difficult to ascertain some information, for example on customer cost. And with such an unclear picture, it is difficult to convince people that telehealth works.* (Managing Director, telehealth company)

Whilst these concerns were also expressed by NHS managers, there was a sense of the burden the data-related issues were causing to the industry. In part because of limited access to data and

confidentiality issues, the industry has to rely entirely on NHS staff to generate evidence of the effectiveness of the technology. For example, the director of one telehealth company recalled:

> *The analysis of the data is always in the realm of the NHS, because the NHS has the data. Commercial companies don't have access. We have to rely on NHS staff to be competent to do it, and very often they don't have time to do it.* (Director, telehealth company)

> *It's fairly easy to get hold of the epidemiology side of things — so number of admissions and that sort of information. What is more difficult is understanding what that specific Trust or customer, whoever that be, is spending, what the current practice of care is, because it varies so much from area to area,* [and] *the costs which they have changed dramatically.* (Telehealth provider)

The majority of the industry respondents considered data-related issues to be a major barrier to risk-sharing agreements, with the supplier paid according to the efficiencies achieved by using his technology for a particular population.

For NHS providers, it was the issue of using the data from other parts of the system and dealing with poor reporting and mistakes in coding. It was very clear that the majority of NHS staff perceived some of the evaluation to be flawed, especially where based on an analysis of admission rates. For example, one NHS provider recalled:

> *The work we are doing in* [coded] *and* [coded]*, when it comes to evaluating it, we needed data from the Acute Trust to enable us to measure impact on admissions. But data from AT based on coded data was so poor, it was unusable. We could not align the admission -- and in* [coded] *we were doing COPD. We could not align an admission specifically*

with the symptoms of COPD. So, it is very difficult to say that system has saved that many admissions because you cannot align the admissions specifically with the telehealth service. And then of course the problem is that if you put this as a result, it is all open to questions. You might have saved an admission but how do you know that they were saved because of telehealth? And you don't, because the data is not good enough. (NHS provider, Site B)

The robustness of the existing research evidence about the effectiveness of telehealth was referred to as another significant concern. Members of the industry felt strongly that the secondary care data were more reliable and robust than the primary care and community care data. There was a perception among these staff that as primary and secondary care computer systems are not integrated, the process of retrieving data is very laborious and time-consuming. The main drawback considered by the director of a telehealth company was the fact that the process would require access to multiple databases and several rounds of coding in order to remove identifiable information.

While discussing the reliability of the data, we touched on the topic of the recent Whole System Demonstrator study (Henderson *et al*., 2013), details of which are discussed in Chapter 1. The perceptions which the telehealth industry and NHS staff held about the results of the trial were consistently critical. Members of the industry especially expressed disapproval of the fact that the costs of usual care were not taken into account in the analysis of trial data, and felt that the results were not useful for the telehealth industry.

The picture below illustrates the main themes and nodes in this category. They were depicted using the Huberman and Miles (1994) method of creating 'networks' as a way to display the data and 'tell the story'. The main themes are depicted as green boxes, with the nodes as blue circles.

Picture 1. Nodes and themes in reliability of data category
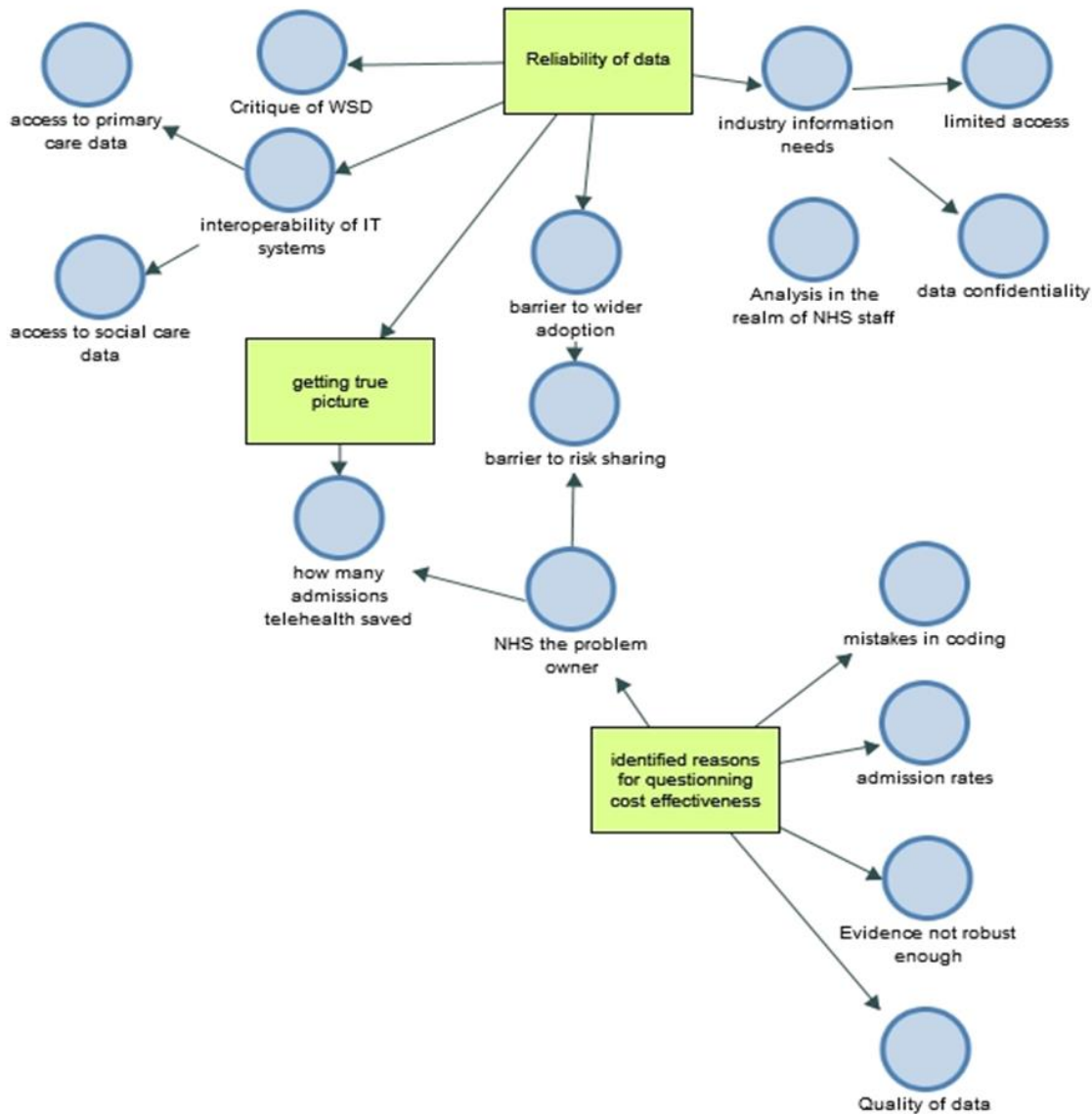
The reliability of the available data was discussed in relation to the research evidence from publications and NHS data. The main reasons for the data being perceived as unreliable was its poor quality, due to mistakes in coding and a lack of robustness. There was a common perception that the lack of interoperability of IT systems, especially between secondary and primary care,

214

was to blame. The reliability of data was found to be one of the factors influencing the industry's information needs, as the limited access and data confidentiality forced them to rely on NHS staff to perform the analyses. The data-related issues affected trust in the evidence on the cost-effectiveness of telehealth and efforts to establish 'the true picture' of telehealth effectiveness.

6.9.4 Identified reasons why telehealth is not generating savings

When discussing the reasons for questioning the cost-effectiveness of telehealth, the lack of organisational changes within NHS organisations was commonly perceived by industry members as one explanation for technology not generating the expected benefits.

A recurring theme among the members of industry was that telehealth should not be viewed as a solution that would automatically create cash savings in the system. Technology was seen as a vehicle to achieve improvements in quality of life for patients and to make savings within the healthcare system. All members of industry felt that NHS organisations had to adapt the design of their pathways of care and staff workloads to the use of technology, and if the process were underestimated the system would ultimately generate losses, instead of benefits. As explained by the managing director of one telehealth company:

> *One of the key problems is if you do not integrate it into care pathways, what you end up with is the cost for extra services: monitoring costs, you have to put together a hub to respond to parameter measurements. If you do not transform the model of delivering care, what you are doing is adding more work and it will cost more. That's why it continues to fail: because nobody looks at that from a viewpoint of the bigger system, and the productivity changes.* (Managing director, telehealth company)

Another interviewee stated:

> *My sense is that people are beginning to recognise that telehealth in itself does not create a lot of values: it is the service and how you effectively change the way you operate that, as many substantial changes take a long time. Therefore, it is coming round to people realising that this is just one tool in a whole change programme.* (Telehealth provider, Site A)

Similar to industry, a commissioning lead viewed the effectiveness of telehealth as achievable, on the condition that the NHS change the way it operates to accommodate new means of delivering care, stating:

> *For the effectiveness of telehealth, it is more important what the provider does than what the equipment does. The equipment can only produce the value if the provider changes the way they do things.* (Commissioning lead, Site C)

An awareness of organisational issues as vital elements for creating an effective telehealth system was also noted among NHS managerial staff. However, they acknowledged that the change was seen as a perpetual challenge to NHS service provision and delivery, with some management staff describing it as source of unsettling experiences, especially for nursing staff whose work had been reorganised several times since the introduction of telehealth. These perceptions influenced how managers viewed the future of telehealth in their services.

A member of the industry identified a need to consider double running costs. He argued that telehealth would only generate real savings if, after achieving efficiencies and moving care into the community, hospital facilities and wards with low occupancy were closed to avoid the costs

of running them. He perceived this as rather long process, as closing facilities would have to be undertaken in stages to match changing demand.

Some industry staff reflected on the lack of scaling-up of telehealth services and the slow speed of deployment as further impeding reasons for the lack of telehealth related savings.

Further reasons for questioning cost-effectiveness were given as the time needed for technology to generate savings, and also that they could appear in different sections of the healthcare and social care system. As illustrated in the accounts of the operating manager of one of the telehealth companies:

> *People are trying to work out how you know that the technology delivered what you expected in terms of savings, but these costs are spread around the whole healthcare system or — even more complex sometimes — the social care system. And you will know that it takes years before you could see how much you saved.* (Operating manager, Site A)

6.9.5 Categories of costs

Different categories of costs were a recurring topic within the interviews, with the majority of participants reporting examples of costs that occurred during telehealth projects. These costs formed two broad themes: implementation costs and operational costs.

Implementation costs

Most interviewees reflected on the costs created during the implementation phase, and expressed the opinion that they were very important to consider in the evaluation of telehealth services. The

rationale for including implementation costs in the evaluations of telehealth to understand the drawbacks was described as an essential part of 'getting a true picture'. The majority of NHS staff and industry members identified common types of these costs, such as the cost of the purchase of the equipment, the installation, training for staff and patients, setting up the machines, setting up the facilities for additional staff, and similar.

As one NHS manager stated:

> *I suppose that there are two elements. The setup element (how much does it cost to set it up as a system) and how much does it cost to maintain it. I suppose within that (set up) there is the initial training, the initial purchase of the equipment, initial setting up the system for monitoring, etc. I guess from looking at it from that point of view, within the maintenance there is time spent actually putting the units out. You get cost around time spent training the patients, the cost of actually monitoring from the alerts, from the admin to hopefully clinical review of that, and then the potentially the intervention that takes place on the back of that.* (Programme manager, Site B)

Both NHS managerial staff and industry members also felt strongly that depending on the service specification, implementation cost could be broader. For example, the risk of detecting 'false alarms' in the early days of operating the system meant that it might be necessary to provide additional training for both staff and patients, or to re-calibrate the system to accommodate individual variations in care needs. The industry members also argued that implementation costs could be perceived as an investment, as long as they occurred when developing a relationship with customers.

Operational costs

Overall, participants had similar perceptions of the operational costs, and reported various examples of day-to-day costs of running a particular service.

For example, one NHS provider highlighted the cost of providing the monitoring, technical and clinical triage, and service administration. He felt very strongly that the costs of responding to patient alerts – either by the matron, or the nurse contacting the patient over the telephone or visiting them at home – were very important to consider. Having a limited number of staff meant that sometimes agency workers would have to be called in to fill these roles.

All participants pointed to the fact that over the long-term they would be liable for maintenance costs. NHS managerial staff provided examples of these costs, including the replacement of faulty equipment or batteries, and deinstallation and cleaning of the equipment when the particular patient had finished his/her telehealth monitoring. NHS staff found the problems with equipment frustrating, and explained how technical difficulties would create additional costs if telehealth provider engineers had to be called in to deal with the situation. Several managers considered good technical support to be a vital element of any contract with the industry they would consider.

More complex technical issues, such as the equipment not working properly and difficulties with uploading the data due to poor connectivity, were seen as major barriers to staff acceptance.

A commissioner lead stated that equipment costs were of major importance to those making purchasing decisions, and there would need to be transparency of the costs of purchase in comparison with rental costs.

6.9.6 Challenges in estimating cost

The difficulties with estimating implementation costs were common concerns for many NHS managers. For instance, one NHS provider described how telehealth had become part of the community care services in their locality as a result of it being promoted by enthusiastic individuals. Therefore, access to retrospective data for that service was limited to the cost of the equipment, and assessment of the true implementation costs was impossible.

The commissioners and members of industry felt that the true costs for providers were not transparent enough for them to assess the affordability of any planned venture.

> *I think the model needs to be clear that the cost question is different in a provider context than in a commissioner context. We're all about the cost to the public purse. So, the cost for us is still the same as the national price list. (We call it tariffs.) And we had a lot of block contracts, and block contracts change only slowly. So, we have national price adjustment to the average tariffs in the contract. For a given level of activity, cost is the same, but for the fixed level of activity the provider costs can vary enormously. If we were doing that modelling with a view to informing the business case, we would still model the costs in the provider side, or there would be elements of that. We want some notion of whether this is affordable to the provider.* (Commissioner lead, Site A)

Another NHS manager felt that in case of telehealth services that were embedded into the existing infrastructure and involved shared costs with other services, separating implementation costs from the overall service costs would be difficult. This would require using the advanced economics/accounting methods that usually require daily recording.

All telehealth industry members appeared to be aware of the significance of implementation costs and wanted to assess them correctly, as this was perceived to be vital in preparing a working business case and relationship with clients.

The challenges of estimating true costs were linked to the fact that, commonly, the costs of telehealth are spread across the whole of the healthcare and social care systems. One NHS manager was very clear that as telehealth improved patients' ability to self-manage and they were not admitted to hospital as frequently, the implications of this and the potential savings would be attributed to the secondary care system; whereas in reality, these patients might need to be visited more often by community matrons or their GP, which would create a greater burden for primary care. These difficulties raised a question of how accurate the MALT model could be in assessing NHS provider costs.

Furthermore, another NHS provider suggested that as is the case in other types of technology, the introduction of telehealth might prompt people to use health services more often. He assumed that this would not necessarily have a negative effect on the healthcare system, as the costs of care, diagnosis, and prevention for COPD were lower than the costs of care for severe stages of the disease. He perceived telehealth as beneficial if its use would prevent greater spending on severe stages of long-term diseases, even if people overused the opportunities telehealth gave them. The relationships between the main cost-related themes are illustrated in Picture 2 below.

Picture 2: The main themes in the cost category
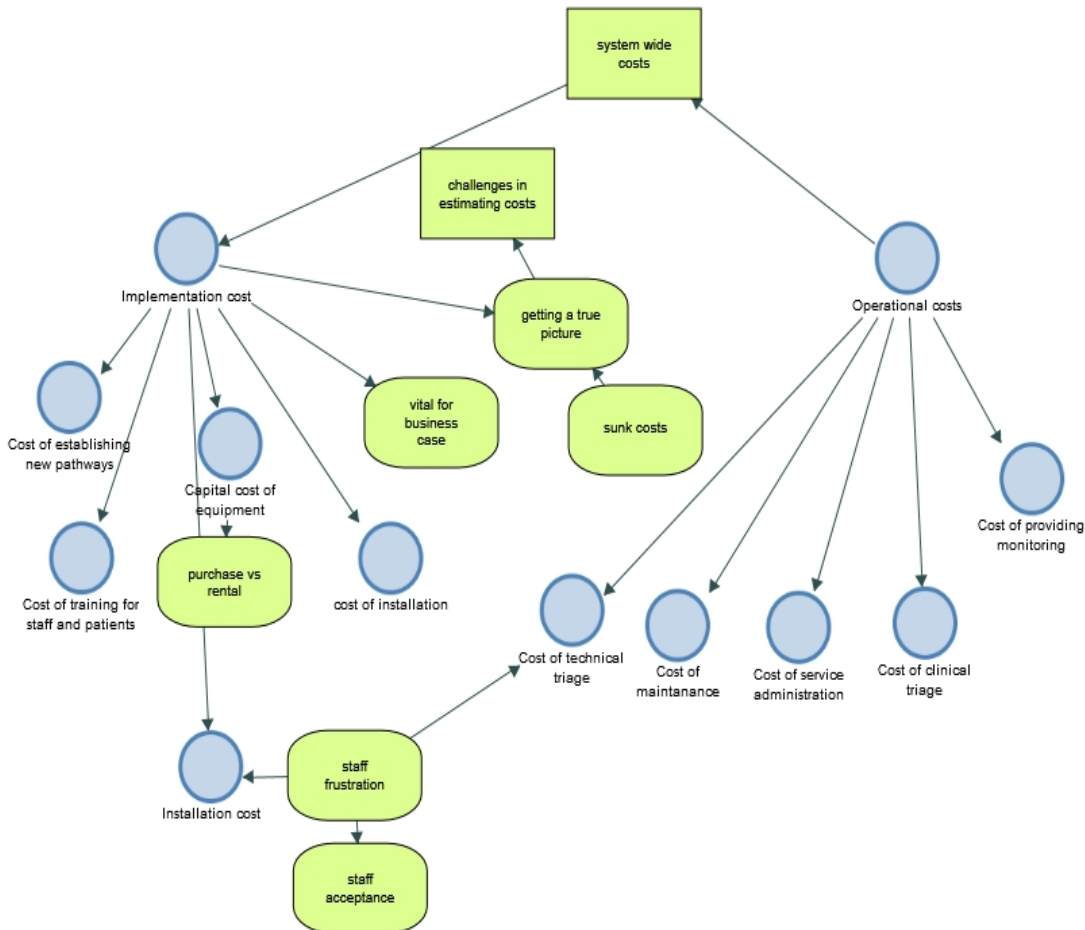
The two main costs categories were identified in the interviews as operational and implementation costs. Interviewees shared the opinion that these costs were spread over the whole healthcare and social care systems, and this later formed part of the discussion on the need to include 'system-wide costs' in the proposed model. The implementation costs were seen as vital to establish the 'true picture' of telehealth and basic elements of any working business case. However, this could only be achieved if the challenges to estimating these costs were overcome. The interviewees were aware that telehealth cost evaluation was associated with several challenges.

Among specific categories of implementation costs, the cost of equipment was seen as vital to assessing whether purchase or leasing were better value for money for the particular service. Later, that theme formed part of one of the major model features. Costs related to telehealth equipment installation were a common source of frustration for NHS staff, impacting their acceptance of telehealth. This is due to the fact that NHS management staff appeared to be testing different provision solutions in order to minimise these costs: either by employing new members of staff or sharing these roles with local authority staff. Staff had similar views of the cost of technical triage, which could change dramatically if there were a need to employ agency workers.

6.9.7 Indicators of the severity of COPD

Across the sample, discussions about the progression and severity of COPD drew on the following themes: COPD exacerbations, COPD-related hospitalisations, and COPD-related A&E visits.

All participants were aware of the progression of COPD, and identified the stages of disease severity. A worsening of the symptoms – commonly known as exacerbation – was considered a major disease event, as it is both dangerous for the patient and costly for the system.

Further explanations referred to COPD-related hospitalisations, as they were seen to have an effect on patients' overall health. As stated by one of the NHS providers:

> *I will always look at how many times a patient has been admitted to the hospital. If you see lots of hospitalisations within a year, you know it's a very serious stage and lots should be done to prevent consequent ones.* (NHS provider, Site A)

All NHS respondents reported that they would pay attention to incidents of A&E visits as an indicator that a patient has to be closely monitored or that current levels of care have not been

efficient in the particular case. For example, a NHS manager stated that for very anxious patients who did not adapt well to changes in their care, any worsening of symptoms might provoke a visit to A&E.

6.9.8 Perceived benefits of using telehealth

All participants expressed the opinion that there were potential benefits of telehealth in terms of health service quality and efficiency, and others related to improvements in patients' quality of life.

For example, one programme manager explained how using a telehealth system in her practice facilitated a faster response to patients' needs and prevented hospitalisations.

> *There were a lot of cases where a patient would alert, and I would call him, and the patient would say, "I don't know why that has happened, I feel fine". And I would say, "Right, well, we will have a look tomorrow". And actually, it occurred that the fact that I reacted so quickly saved him from going to the hospital.* (Programme manager, Site C)

A recurring theme for all NHS and industry staff was that telehealth enabled improvements in health services in the way it impacts access to care for patients, it speeds up the response time in case of exacerbations, and makes up-to-date patient data available in case of deteriorating conditions.

The ability of the technology to provide evidence for patient progress and justification of treatment decisions was also raised as a benefit of the service by both NHS providers and members of the industry. For example, the NHS manager described how she discovered negligence in a patient care:

We have had a serious incident where a matron wasn't looking after a patient, and thanks to telehealth we knew that something was wrong. (Programme manager, Site C)

The director of one telehealth company described how data from the telehealth system were used as evidence that a patient in a care home was being looked after properly.

The critical factor when you talk about the data from the telehealth system is that you get data on what staff is doing. That gives a care staff 'guarding umbrella' when they are falsely accused of doing or not doing something. So, evidence works not only for residents but also for staff. (Director, telehealth company)

Many of the NHS and industry informants pointed to the need to make cost savings in primary and secondary care when explaining the perceived benefits of using telehealth. Examples like a reduction in length of hospital stay, reduction in the number of A&E attendances, and the usage of community and primary care, were prominent in both NHS providers and industry members' accounts. For example, a member of industry stated:

The key point is to look from the utilisation side – reduction in hospitalisations – as a key benefit. You also have an avoidance of secondary care resources as a key saving. You should see a shift from using secondary care to more primary care, which is a big cost saving. (Director, telehealth company)

Reflections on the wide range of quality of life benefits for patients were present in all stakeholders' accounts. Examples included avoided hospitalisations, improved medication compliance and reduction in anxiety. For many NHS managers, the most important benefit of telehealth was the fact that it empowers patients to manage their disease better.

Actually, if we can educate people, if we can find them earlier, we can diagnose them earlier. We can give them the skills to be able to self-manage and identify when things go wrong and what to do when things go wrong. (Programme manager, Site C)

And similarly, the programme manager from one local authority stated:

We want to look for an improvement in self-management, an improvement in patient confidence in managing their disease, recognition of exacerbation and then reduction in emergency admissions, and then reduced length of stay, because we could do early supported discharge if they were being monitored using telehealth. (Programme manager, Site B)

6.9.9 Intangible benefits

Many of the NHS and industry members identified intangible benefits for staff, patients, and their carers, and argued that they were equally important when considering the benefits of using telehealth. The majority of the managerial staff provided anecdotal evidence about the cost and time savings achieved by patients and carers by not having to travel to healthcare institutions. For example, the senior commissioning manager noted:

If someone is unwell and they do not have to travel to the hospital, you see lots of savings, and sometimes they can be significant. The same thing for the carers as well. These are important factors. (Commissioning manager, Site C)

All commissioners demonstrated understanding of the potential benefits of telehealth, especially its impact on staff workload. They generally supported the idea that telehealth would enable care to be provided in the community, which would be associated with significant savings for the NHS.

For some NHS managers, the fact that NHS staff could save time by not having to travel to see a patient was suggested as one of the major benefits of using technology. Among those managers who perceived a positive impact of telehealth on staff workload, there was a general feeling that telehealth might also help with caseload management and thus change the current ways of working with patients.

One NHS manager explained the implications of proper service set-up, stating:

> *We know some services where the patient is linked to the equipment and they have a technical alert. That alert goes straight to community nursing teams. We know that about 40% of all alerts are technical alerts: the patient hasn't used the system properly or is not plugged in properly, or whatever. Of course what we do is validate the technical alerts, and we use the very low-cost resource to do that. So, then you save your high-cost clinicians a lot of time, and therefore the expense of doing things they don't need to do.*
> (Operational manager, Site B)

The view that telehealth might lead to an increase in workload if the service provision and equipment are not properly calibrated was also reflected by the members of industry. Overall, there were strong feelings that an appropriate level of coverage for technology providers should prevent NHS staff and patients from experiencing such difficulties. The main themes from the benefits category are illustrated in Picture 3 below.

Picture 3: Themes and nodes in the telehealth benefits category
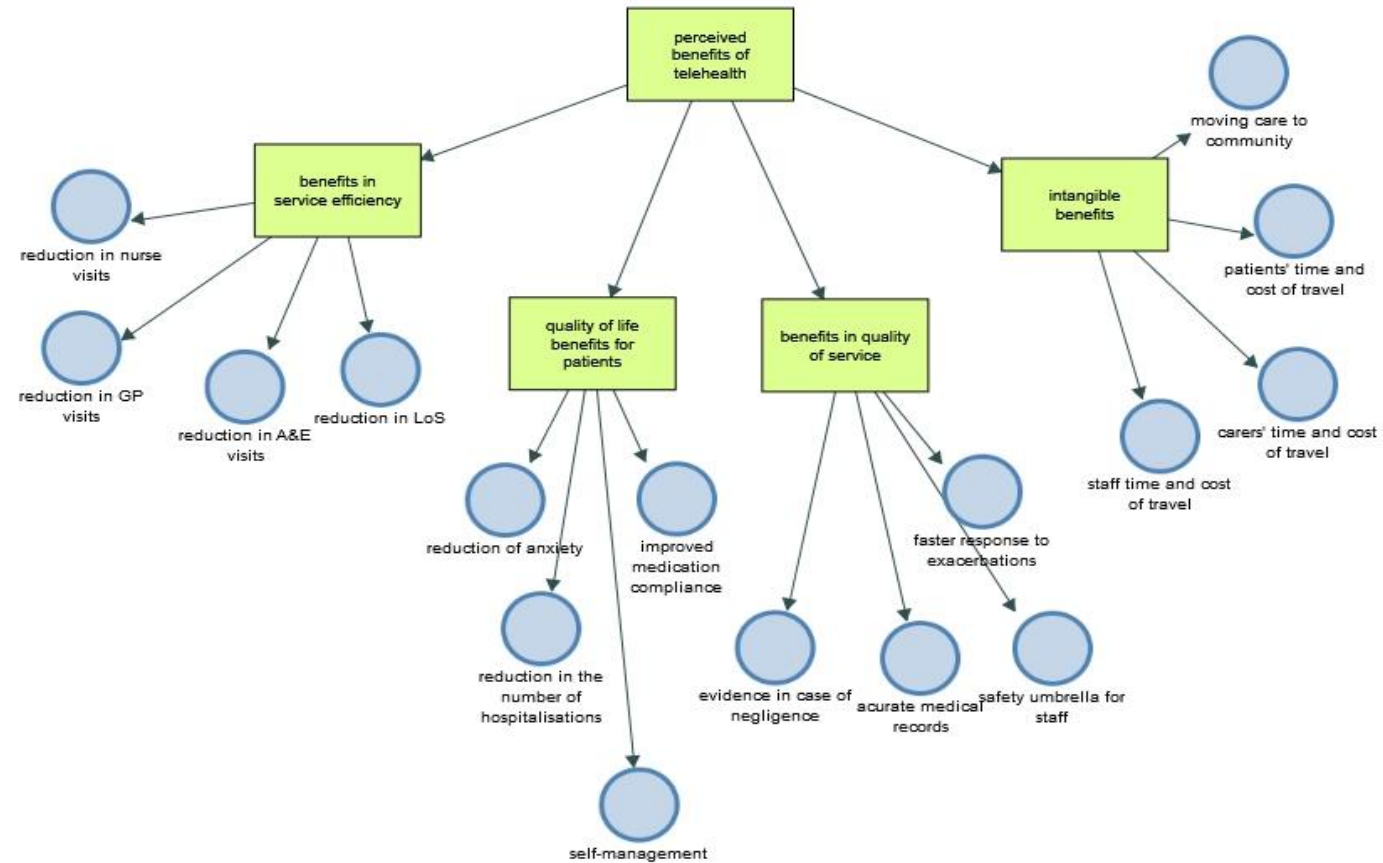
The perceived benefits of telehealth were closely related to four other themes: the benefits in service efficiency, quality of life benefits for patients, benefits in quality of service, and intangible benefits.

6.9.10 Perceived use of a financial model

NHS staff perceived two applications for the proposed model. Managers concluded that the model could help with the allocation of resources and achieving efficiencies: a view echoed by the commissioning managers. There was an awareness that, by using the tool, NHS managers could allocate staff and equipment more efficiently, and thus provide care to a greater number of patients. For example, one NHS manager stated:

> *So, it's about: how do we show that by using the technology we can save resources? Because if we can do things slightly differently, we are dealing with more patients, but actually from the same financial envelope. Because I don't think we're going to release massive savings. I probably shouldn't say this either but I can't see us releasing massive savings. It's about trying to manage the existing activity within what we have got.*

> (Programme manager, Site A)

Further, there was a consensus that the MALT model could assist in providing the evidence that technology generates savings by improving utilisation of resources, as illustrated in the NHS provider account below:

> *We are constantly being told that our services do not have sufficient capacity because of the increase in demand. I would like to be able to prove that what they do not expect is*

that we can manage the increase in demand by improving our efficiency by using technology. (Programme manager, Site B)

The majority of industry members described previous attempts to estimate telehealth-related costs when working on a business case, and seeing the potential to use the model to evaluate costs, especially from providers' perspective. For example, the director of a telehealth company stated:

I suppose I would like to get a true picture of the effectiveness of telehealth. Actually, are the assumptions we're making about the cost of [TH] *borne out of reality? It would allow me to understand the totality of the cost.* (Director, telehealth company)

The majority of the informants pointed to patients' satisfaction and benefits as factors influencing their decision to provide telehealth. However, they stated that the evidence on the cost-effectiveness of technology was equally important for them, and that they would use the model to support the claim that telehealth was cost-effective. For example, a GP stated:

I am still not entirely sure how it improves patient care or if it's cost-effective. I think patients like it because they feel that they are being looked after more regularly. So, although we are comfortable using the system, I am still not clear and would like to have the evidence that it is cost-effective. (GP, Site C)

The same point was raised by a member of industry:

We need to be able to show that there is an absolute correlation between reduced hospitalisations and the telehealth system […] *What I need to show is that those specific*

admissions were reduced because the symptoms were dealt differently. (Telehealth
provider)

6.9.11 Requirements for the tool

At the outset of the project, we assumed that the MALT model's future users would be mainly
managerial staff involved in decision-making and those who would like to use it actively to
perform calculations and assess scenarios, perhaps finance department staff. This was confirmed
by our respondents, who suggested that within the NHS the finance and managerial staff are the
most likely to be interested in using the proposed model.

One of the limitations of this elicitation was that the respondents shared their views on the
requirements for the model without seeing it. Therefore, we did not expect to discuss the technical
details of mathematical modelling or user-friendliness issues, such as the layout or presentation
of the data within the spreadsheets. This formed the basis of the usability evaluation described in
Chapter 8.

However, several key requirements for the tool were identified in the interview data. Many NHS
and industry informants pointed to the importance of the flexibility of the model, especially to
enable the change of multiple service specifications.

The commissioning lead raised a question of model transparency in regards to data sources and
model assumptions. He stated:

> *I would like it clearly and transparently referenced, which is sometimes where NICE falls*
> *down, because they will compound the number of steps into one and it is not transparent*
> *how that reference has resulted in this particular effect.* (Commissioning lead, Site A)

Another requirement was related to the software platform used for the model. One of the NHS staff pointed to the use of commonly available Excel software. This has the advantage of being free and routinely used within the NHS organisations, as opposed to bespoke modelling programmes. She stated:

> *I would prefer Excel because we use Excel. It would take time and effort to learn how to use other programmes.* (Programme manager, Site B)

6.9.12 Examples of analyses provided by stakeholders

Many of the NHS managers and telehealth providers identified the value of performing analyses to generate evidence of telehealth effectiveness. For example, one NHS provider stated:

> *So, thinking about scenarios, I want to look at return on investment. Within the money I spend, are my nurses able to see more patients? Am I improving the quality of care? Am I reducing the number of times patients go to hospital? Do I have fewer GP visits? All that kind of thing. Because every GP visit has a cost, every nurse visit has a cost.* (Programme manager, Site C)

The return on investment was discussed with all interviewees. Whilst some staff reported that the analysis might not be useful to them, there was a general consensus that establishing how much money the introduction of technology brought to a service was crucial to ensure funding for the future. This was especially evident in the NHS managers accounts, as they expressed concerns that at some point in the future telehealth services would need to undergo improvements, whether by upgrading the equipment or introducing the service to patients with other long-term conditions, and this would require thorough investigations to support such investments. A good example of this was provided by one of the NHS managers, who stated that he would like to test how the

effectiveness of telehealth varied if technology was provided for different periods of time, e.g. three months, six months, the patient's lifetime. He also stated that he would like to test the difference it made if the telehealth services were provided for five days a week, versus seven.

All NHS managers observed that current telehealth services had limited scope and served a limited number of patients. They recognised that there was a need to increase the number of patients using technology and the number of hubs available. This raised the question of the number of patients needed for the system to 'break even'. There was a consensus among NHS managers that this was a complex question and depended on patient selection, disease severity, and individual abilities in mastering self-management.

It was therefore unsurprising that another scenario most often requested by NHS managers and commissioners was a comparison of technology provision for different target populations, e.g. people newly discharged from hospital, in comparison to patients recently diagnosed with a chronic condition.

Although some NHS staff tended not to question the future of the services they ran, they expressed interest in evaluating different procurement scenarios as they all understood that purchase was always more expensive than rental, and that other options for commissioning technology might be available. For example, the option of a managed service where equipment was frequently upgraded as part of the package.

Several scenarios were provided by commissioners, and the evaluation of risk-sharing agreements was most often cited. Risk-sharing agreements are a solution used in Europe for a variety of medical interventions. They are based on the assumption that the industry or technology provider should be paid according to the efficiencies achieved due to their technology.

Another scenario suggested by commissioners was the evaluation of cost-sharing for NHS providers. They argued that secondary care providers should cover some of the costs of telehealth systems, as they benefitted from reduced admissions and reduced average length of stay.

6.9.13 Approach to financial planning

The majority of NHS providers stated that they had not previously used any financial models in their practice. On the other hand, these tools were reported to be commonly used by the industry to monitor cash flow on a weekly basis, though they did not regard these as being economic or financial models. The only group which seemed to have some experience of using more advanced analyses were those from Clinical Commissioning Groups.

Three of our study respondents had specific preferences for the types of financial analyses they required, and their comments offer some insight into approaches to financial planning among stakeholders.

For example, one commissioning lead explained how he used Net Present Value (NPV) analysis to assess the affordability of a new venture. This type of analysis is usually required in all NHS business cases as it allows a prediction of the present value of future expenditures, and demonstrates the funding required for the project. NPV calculates the long-term return of investment in a programme by deducting the total costs from the total returns of the investment. A positive NPV means that the investment should be made, as future cash flows will cover current investment. A negative NPV means that the investment should not be made. However, as we reflected upon this with other NHS providers, it become apparent that in the NHS it is a standard practice to evaluate cash flows within a year; thus the NPV approach is useful but not required for the majority of the projects. NPV is recommended by the Green Book (HM Treasury, 2003) as standard practice when performing option appraisals.

Another individual mentioned Return on investment (ROI) as vital for assessing the affordability of new ventures. Conversely, the majority of NHS secondary care providers stated that as they did not manage the surplus of funds, ROI would not be useful for them in the MALT model. All commissioners also rejected related measure – Social Return on Investment, which estimates the value of wider social benefits, including the value of lost productivity. The measure demonstrates the percentage of return which society gains from the project over a given number of years. Commissioners perceived it as useful, but only in the case of public health programme prioritisation.

6.10 Discussion

To explore stakeholder requirements for a proposed model, a series of qualitative interviews was undertaken. The aim was to ensure that the envisaged tool met the requirements of its future users. This section outlines the main findings from these interviews.

6.10.1 Findings

The participants of the interviews formed part of a complex stakeholder matrix of different roles and different organisations. I explored the thoughts, opinions and beliefs of people who sold telehealth equipment and components of the service, people who bought the equipment and provided healthcare services, and people who handled the procurement of technology. It became evident that, to some extent, perceived differences in stakeholder beliefs could be explained by the group to which they belonged. The detailed information on service specification for each of the sites is discussed in Chapter 2.

For example, very early on, it became apparent that the ambitions and goals of industry differed significantly from those of NHS and Clinical Commissioning Groups. The interview findings

revealed how much effort and importance had been given by members of industry to research and development (R&D). Telehealth continued to be viewed as a rapidly changing environment, and telehealth providers emphasised the importance of following market demand and responding to patient needs. They also pointed to the fact that it is a very competitive field, and that in order to satisfy their clients they must continuously improve their solutions. For example, there is ongoing work to reduce the costs of telehealth systems by designing 'fit for purpose' equipment that could be delivered by post and installed by patients themselves.

Members of the industry indicated that they were using specialised tools to assess the profitability of these new R&D ventures. This was a key message to the MALT modeller team, as the inclusion of industry specific costs in cash flow analysis was considered important for ensuring industry 'buy-in'.

One of the key findings from the qualitative interviews was that NHS providers expected the model to demonstrate how to use resources more efficiently. NHS providers in their approach to telehealth were activity focused, and would use the model mainly to understand the impact of technology on admissions and re-admissions. Looking for an interpretation, I found two possible reasons. Firstly, for NHS providers, income was the Payment by Result (PbR) tariff paid for a certain level of activity. Therefore, each saved admission meant a financial loss for the providers. This is a general weakness of the PbR scheme, which is most appropriate for elective care and less suited to services where less activity is desirable. The scheme gives an incentive for providers to increase activity when their prices exceed costs, or to ultimately stop proving these services. The problem could be solved by introducing a 'Year of Care' tariff for telehealth, which has been trialled in a few areas of the UK. Under this tariff, NHS providers would be paid to care for a patient with a long-term condition for a year, rather than to receive payment each time the patient was admitted to hospital.

Secondly, the hospital bears the cost of any re-admission within 30 days of discharge. This rule was implemented to ensure that there was a financial motivation in the system to prevent patients from being discharged too early. If telehealth prevents patients from being re-hospitalised, NHS providers could achieve significant savings.

This is the main reason why NHS providers must find a way to minimise their expenditures by achieving efficiencies through new ways of working. However, the interviews revealed that they did not necessarily know how to achieve effectiveness using telehealth technology. It was mainly members of the industry who provided examples of the changes that healthcare organisations have to undertake in terms of infrastructure and performance management to benefit fully from technology.

When we discussed plans for the future for these services, none of the NHS managers demonstrated an understanding of the need for efficiencies of scale. This created an impression that telehealth was an addition to the staff workload, and not a new mode of delivery of care. Moreover, when reviewing specifications of the services in each site for the purposes of model development, only one site demonstrated a business model in which staff and resources were effectively used. In both Sites A and B, the caseloads of each of the matrons included only a few patients on telehealth, and there were no incentives for change of performance.

On the other hand, the telehealth industry staff demonstrated a great understanding of the benefits and drawbacks of telehealth technology provision, and provided explanations of how to ensure that telehealth created efficiencies. Importantly, they identified the challenge of decommissioning the existing services, as they argued that in the long run dual running of the services (telehealth and usual care) would generate economic losses. They suggested that to benefit fully from telehealth, at some stage of maturity the decision on disinvestment in the

services would have to be made. It was acknowledged that from the perspective of the MALT model and mathematical modelling, this might be a very complex issue, and that the model might not be able to fulfil this requirement. The modelling team agreed that the model could provide estimates for service changes, but ultimately the decision-makers would have to make the decision on issues as complex as disinvestment.

6.10.2 Findings for categories of cost

The interview findings suggest that model inputs in relation to costs should adopt the NHS providers' perspective, and consider system-wide costs and the costs of the care pathway, rather than just the telehealth system components. From the interviews, it was evident that members of the industry would use the model to analyse costs and benefits only from the NHS provider perspective, as this was useful for building a business case and a relationship with customers.

In terms of categories of cost per se, it appeared that for NHS providers and commissioners, implementation costs were less important for service evaluations, especially in the case of existing services. This was the approach suggested by the Green Book (HM Treasury, 2003), where the cost of goods and services already incurred are treated as irrevocable and ignored in evaluations. They are often referred to as 'sunk costs'.

Conversely, members of the telehealth industry presented a different approach, arguing that all costs should be included in the MALT model as this would help to build a viable business case when discussing options with their clients.

All stakeholders agreed that the operational costs of the model should detail all service specifications, and that this could be very complex. All respondents indicated that they would benefit from the model if it detailed how costs changed over time. This was especially important

in the case of telehealth, as equipment became obsolete very quickly, and various types of commercial agreements with telehealth providers existed in the market. For example, NHS providers must consider the trade-off between purchasing and leasing the equipment in terms of long-term costs. As explained by the interviewees, usually the lease of the equipment allows replacing and updating telehealth hubs without additional financial investment. NHS providers also have to consider different service provision models, including the range of fully managed services.

6.10.3 Findings for benefits

Interview findings revealed that stakeholders perceived telehealth benefits in similar ways. This is illustrated in Table 36 in Appendix A.7. Specifically, telehealth benefits are categorised as benefits in health service efficiency, quality of service benefits, quality of life (QoL) benefits for patients, and intangible benefits.

All participants agreed that telehealth enables more efficient use of healthcare resources. The most often mentioned examples of such were the ability for care to be provided in the community, to deal with more patients under the same financial envelope, and for clinicians to have bigger caseloads.

It was believed that telehealth could create intangible benefits, especially for patients and carers, in terms of savings in the cost of transportation to/from a hospital, cost of parking fees, and similar. The fact that staff did not have to travel for home visits was also seen as an important benefit. The interviews revealed that stakeholders preferred these costs to be included in the model. However, they were also aware of methodological challenges to such an approach. The majority of respondents revealed that they did not have access to such data. They also expected

difficulties in estimating telehealth costs, as this would require additional resources such as the input of an administrator.

The gold standard for cost-effectiveness modelling, as recommended by NICE, is use of QALYs as a unit of effectiveness. This enables decision-makers to compare interventions whose effects on health are different. However, the QALY might be not sensitive enough to detect the small changes in health outcomes which telehealth is likely to produce. The benefits of reduced mortality and morbidity would also be difficult to calculate.

The findings revealed that commissioning staff preferred the model to use QALYs, as that would allow them to compare different interventions. NHS providers, however, seemed satisfied with a less complex measure, one which would have practical use, such as 'cost per patient'.

My concern is that telehealth benefits will not contribute to any monetary savings until NHS capacity is reduced to match lower demand. True savings could only be achieved if decommissioning were introduced to hospitals by closing wards and redeploying staff. This is, however, a lengthy process. Another sign of resource savings would be the reduction of waiting times; as other patients would be treated sooner. In cases where health demand is increasing over time, these savings allow hospital capacity to remain at a lower level. The MALT model could help to explore the point at which the roll-out of telehealth programmes starts to reduce hospital capacity and to accurately scale-up telehealth services.

6.10.4 Findings for model scenarios

This is the category of the model features in which the majority of differences in stakeholder requirements were observed. NHS providers expressed an interest in evaluating scenarios, with different durations of deployment, scales of deployment, and target populations. CCG staff

expressed an interest in using the model to evaluate risk-sharing and co-payment scenarios, whereas members of industry suggested that they would use the results of simulations as a basis for evaluating alternative business solutions, for example new products.

My understanding is that some of these queries, especially those concerned with patient selection, could be solved by using risk stratification tools. By using various risk measures, risk tools can predict who is likely to be hospitalised in the future, and therefore who would benefit from telehealth programmes. Of three study sites, only one used a risk stratification tool to screen patients suitable for telehealth. That, in my opinion, explains why so many examples of scenarios concerned with the patient selection were given. The examples of observed differences in participants' accounts are illustrated in Table 10 below.

Table 10: Stakeholders' preferences for the MALT model scenarios

Telehealth Providers	NHS Providers	Clinical Commissioning Groups
Private insurance schemes, Using low cost solutions, Providers' activity: savings in utilisation of staff time and resources	Duration and scale of deployment, target populations *'Nationally all the debate at the moment is trying to deploy at scale and what the barriers are for doing that. We would like to work around increasing scale and scope, coming to point where is that saturation point for our population. (Managing director, Site C)*	Risk sharing and co-payment: *'We want to sort of pass some of the costs onto the service providers because, actually, as they are being more efficient they get some of the benefits.' (Commissioning lead, Site B)*

6.11 Triangulation

Triangulation, in qualitative research, is a methodological process that contributes to the validation of findings, but also increases the depth of understanding (Huberman & Miles, 1994). It is used in three cases: when the study uses multiple data sources; when multiple observers or researchers are involved in studying the same object; and when multiple perspectives are used to interpret the data (Denzin & Lincoln, 1994).

The recommendations of Farmer and colleagues (2006) were used to develop a triangulation strategy for this research. Firstly, the findings from each data set (NHS managers, commissioners, industry representatives, and local authority staff) were analysed to address the research question:

What are the stakeholders' preferences for model inputs, outputs and scenarios?

A number of the themes identified directly informed this research question. The themes in each category were then assessed to determine their meaning and the degree of convergence across the data sets. The content and meaning of the themes were then compared to identify the areas of overlap and divergence.

For example, I compared the four datasets with respect to the meaning and interpretation of the theme 'indicators of telehealth effectiveness'. Whilst the accounts of the NHS managers frequently included the 'reduction of hospital admissions' code, the commissioning leads' and industry members' accounts most often referred to 'change in the pattern of care'. Their views differed on the use of a reduction in hospital admissions as an indicator of telehealth effectiveness, but there was agreement on the meaning and prominence of that indicator per se. The stages of the triangulation process are listed in Table 11 below.

Table 11. The stages of the triangulation process using Farmer (2006, p. 383)

The stages of the triangulation process	Specific activities undertaken
1. Sorting	NHS managers, commissioners, industry members, and local authority staff accounts: sorted as cases in Nvivo 11 (QRS International., 2012).
2. Identifying themes from each data source (case) to address the main research question	Comparison of coding to determine the degree of overlap and the meaning of codes in each of the data sets.
3. Coding	Identification of the meaning and the prominence of codes in each of the data sets. Assessment of whether there is an agreement, partial agreement, silence or dissonance in the coding.
4. Assessment	Review of the entire coding scheme in regards to the cases of silence or dissonance in the findings. Comparison of the key differences in coverage.

Based on this exercise, I was able to point to the themes where an agreement, partial agreement,

silence, or dissonance were noted, as illustrated in Table 12 below.

Table 12. Triangulation results

The main theme and sub-themes where the differences occurred	Triangulation results	Meaning
Rationale for telehealth implementation	Agreement	The meaning of the themes in this category was the same in case of the telehealth industry, GPs, and NHS managers. The prominence of themes was similar in each of the data set
Indicators of telehealth effectiveness/ efficiency	Agreement	As above
Reliability of available data	Agreement	As above

The main theme and sub-themes where the differences occurred	Triangulation results	Meaning
Identified reasons why the technology is not generating savings	Agreement	Except the two sub-themes listed, the prominence and the coverage of the themes were similar in all data sets
Double running costs	Silence in the NHS staff, commissioners, and local authority accounts	
Scale and speed of deployment	Silence in the NHS staff, commissioners, and local authority staff accounts	
Categories of costs	Agreement	The meaning, the coverage, and prominence of coding were similar in all data sets
Methodological challenges in estimating costs	Agreement	As above
Indicators of severity of the COPD disease	Agreement	As above
Perceived benefits of using telehealth	Agreement	As above

The main theme and sub-themes where the differences occurred	Triangulation results	Meaning
Purpose of using the model Strategic planning of allocation of resources To inform business case	Silence in industry and local authority staff accounts Silence in NHS, commissioners, and local authority accounts	Except the two sub-themes the prominence and the coverage of the themes were similar in all data sets
Requirements for the tool Transparent	Silence in the industry accounts	Except one sub-theme, the prominence and the coverage of the themes were similar
What analyses stakeholders would like to perform Risk sharing	Silence in NHS and industry accounts	Except one sub-theme, the prominence and the coverage of the themes were similar
Approach to financial planning tools	Silence in the NHS staff and local authority staff accounts	The majority of the sub-themes in this category were only prominent in the industry and the commissioners' accounts

Analysis of the prominence, the coverage and the meaning of the majority of the themes revealed partial or full agreement between the data sets on the majority of the coding. There were no instances of disagreement, but some of the themes were only present in particular data sets and absent in the others. Having in mind that the aim of the interviews was to elicit the requirements of different groups of stakeholders, these differences were expected. Triangulation helped to deepen my understanding of the nuances of the data and to enable a conclusion as to whether the findings corroborated each other.

6.12 Implications of the interview findings for the MALT model development

The findings of this qualitative study clearly demonstrate the complexity of the issues that modellers encounter when developing models for telehealth. The findings reveal that the proposed model has two important challenges to meet:

a. Demonstrate in a numeric way how costs and benefits flow through the system, given service specifications, and what long-term costs and benefits could be;

b. Estimate potential results from a variety of scenarios, given user-specified assumptions.

The model should provide evidence of how several elements of the healthcare and social care systems benefit from telehealth. This should give NHS managers much-needed evidence and incentivise them to implement restructuring changes within their organisations.

The interviews revealed that stakeholders have different ambitions and goals, and will use the model for different purposes. The telehealth industry is concerned with gathering evidence for the effectiveness of its products, and will use the model to investigate the costs and benefits from the NHS providers' perspective. For the NHS providers, evidence of productivity changes within the system is essential, and they will use the model to investigate how to manage existing capacity

more effectively. Commissioning managers are concerned with the allocation of public money and will seek evidence that by contracting telehealth services they can 'do more for less'.

As highlighted in the review of NHS business cases in Chapter 2, the approach generally undertaken by NHS managers is to assess the affordability of new interventions, as opposed to evaluating how to maximise the health of the population within a certain budget, as is the case in cost-effectiveness analysis. This seems to explain why NHS managers prioritise the need for information on telehealth value for money.

The critical element in the MALT model should therefore involve estimating cost savings within the whole healthcare and social care system, and demonstrating the range of telehealth-related benefits that are not necessary visible in monetary form or cash-releasing. The flow of costs in the supply chain between manufacturer and companies to provide other components of the telehealth service could be omitted, as the industry uses dedicated tools to estimate such cash flows.

Another finding was that the model should consider scale and speed of deployment. It is a common commercial practice that the cost of the equipment decreases with the volume of purchase. The scale of deployment was considered by telehealth providers to be a critical factor when assessing the costs of telehealth solutions, and they indicated that they would prefer to have this included in the model.

In developing the MALT model, a different approach is proposed: all costs are considered, on the assumption that there are certain opportunity costs tied up in these resources. On several occasions during the process of data elicitation for the model, I experienced difficulties in assessing implementation costs in these services, as data were perceived as historical and difficult to

retrieve. For the model development process, this means that less effort should be put into developing sophisticated options for the users to estimate their implementation costs, perhaps replacing them with an aggregate estimate.

I would venture the statement that there is no other intervention in the NHS which represents such complexity in terms of involved actors, including public and private entities, as in the case of telehealth. In light of the interview findings, I argue that the MALT model can serve all of the stakeholders, who will benefit from the evidence on disease progression, structure of costs, and benefits. However, it is possible that to meet specific requirements more closely, three versions of the model, of different degrees of complexity, should be constructed.

6.12.1 General and specific requirements for the model identified from the interviews

Given the number of requirements collected through the qualitative interviews, I have compiled a list of general requirements for model developers, based on an analysis of the findings. The model has to fulfil the requirements for:

- **Transparency in regards to the data sources, assumptions, and calculations:** As previously mentioned, the level of understanding of the principles of cost-effectiveness modelling differs between users. Findings suggest that the sources of the data should be clearly referenced and the main assumptions easily understandable, even for a lay person.

- **Flexibility in allowing users to change parameters:** The model has to incorporate a number of variables to enable users to adapt the template to their particular circumstances and to encourage prospective telehealth planners to think deeply about how all of the elements of the system (technology, people) come together to provide sustainable value.

- **Quick development time:** The time to produce results by evaluating a scenario has to be relatively short. The model has to allow rapid calculations with changing variables.

- **Dynamic and sensitive:** In the case of a new data becoming available, users expect to be able to adjust the main variables.

- **Software platform:** Excel spreadsheets are preferred, as opposed to the modelling-specific software, e.g., TreeAge, DecisionPro.

- **User-friendliness and easy navigation:** The majority of the users mentioned simple tasks they would use the model for, such as producing graphs for commissioning board presentations, or finding evidence for reports. Therefore, it should be uncomplicated and easy to navigate.

The following examples of specific requirements were identified from interviews, and fed back to the MALT modelling team.

1. The members of the NHS stated that they would appreciate a table that demonstrated the total cost of telehealth solutions per patient for every considered option. This was seen as very useful for further discussions with commissioners.

2. The NHS managers stated that a table listing all the costs and total savings due to telehealth would be very useful.

3. It was also mentioned that totals for telehealth-related investments in equipment, training, and staff, would be valuable information.

4. All interviewees stated that a graph or table illustrating optimal scaling-up would be a great option.

6.13 Strengths of this research

A key strength of this qualitative inquiry was the number and range of stakeholders involved in the interviews. The sample frame yielded a high response rate. From the initially approached group of interviewees, only seven participants did not agree to take part in the study – two GPs and five members of the industry. A further two people were recruited through the informants, using a snowballing sampling technique. The positive response from the participants allowed me to meet the sampling strategy criteria. The involvement of a diverse range of stakeholders meant that the data saturation point was achieved very quickly.

The majority of these findings were validated during the MALT stakeholders' workshops and conferences.

6.14 Limitations of this research

One of the major limitations of this analysis came from the involvement of just one mature telehealth service. The telehealth services in the other MALT study sites served just small numbers of COPD patients. Also, in one of the cities, the telehealth services were undergoing major transformation.

A further potential criticism of this research relates to the lack of involvement of COPD patients and carers, and the minimal representation of the Department of Health, GPs, and NHS Acute Trust. In health services research, COPD patients and carers are considered important stakeholders. They were excluded from the interviews because they were not expected to use the MALT model.

Involving representatives from the DoH and the NHS Acute Trust was difficult because of the extremely busy workloads of these types of individuals. There were two occasions where I was offered limited time for an interview and had to follow a shortened version of the topic guide, which meant that the data collected from these participants was considerably limited.

Furthermore, the fact that the interviews were conducted as part of wider research programme meant that their organisations had to be compromised. Specifically, because I conducted my interviews parallel to the qualitative work conducted by a researcher in the other work stream of the project meant that I was given limited time for each interview. I also had to remove overlapping questions to minimise the burden to participants.

6.15 The value of this research

This chapter presented research which produced a specific list of requirements useful for modelling purposes. One can argue that the same results could have been obtained using a less structured interview process and simply listing any requirements given by stakeholders.

However, one of the benefits of the method used was that it provided a detailed understanding of contextual issues that were used to guide model development. More importantly, it also informed other streams within the MALT project, and contributed to MALT projects' overall conclusions.

Therefore, it is not clear whether the additional resources required for conducting such a formal qualitative analysis would be worthwhile in every project. However, it is difficult to draw clear conclusions without a formal comparison of methods. It is likely that the benefits of formal qualitative interviews within conceptual modelling are greatest when it is important to understand a complex decision-making environment, and when the work is embedded in a broader research programme.

7. Capabilities of the MALT model

This Chapter provides an overview of the capabilities of the MALT model. It was developed as part of the MALT study with modellers from School of Health and Related Research. I contributed to the development of the model by assisting at various stages of model development, but the work described here was delivered predominantly by Dr Praveen Thokala and Dr Peter Dodd.

7.1 Rationale

The evaluation of health technologies typically uses cost-effectiveness models to assess the incremental cost-effectiveness ratio (ICER) of current treatment against one or more alternative treatment options (Briggs *et al*., 2006). This approach has been primarily developed to aid reimbursement decisions, which are then operationalised locally through further decision making processes with or without further mathematical modelling (Drummond *et al*., 2008). This focus has led to several methodological traits being adopted, with three having an important role in the evaluation of telehealth considered within this thesis (Drummond *et al*., 2015; Gold, 1996).

Firstly, the costs of implementation, if included at all, are incorporated as an annual increment to the unit cost of the new service through a process called annuitisation (Drummond *et al*., 2015; Gold, 1996). Secondly, all resources are valued using long-run costs, which implies that there are no resource constraints. Thirdly, detailed operational costings are often given little attention. These traits are problematic for several reasons:

a) Implementation costs are non-trivial thereby having the potential to have a significant impact on budgets, and as such, the feasibility of implementation (Gold, 1996, pp. 199 - 203). This issue was considered to be relevant for telehealth due to the need for

significant service redesign, together with the need for the 'dual running' of old and new services for a period of time.

b) Implementation is a complex process (Durlak & DuPre, 2008; Rycroft-Malone *et al*., 2004), with several factors influencing the time profile of costs. This issue was considered significant due to different patients are provided with the telemonitoring equipment in different time, also the length of the intervention varies between individuals, as it largely depends on the severity of the disease and individual needs. These factors have an impact on the time profile to implementation.

c) Resource constraints are present within the service providers (Weinstein, 1990). In larger organisations, there can be greater flexibility relating to the provision of care, whereas smaller community care services have significant fixed or semi-fixed costs (in the short-run).

d) Detailed operational costings are important for financial planning. Due to financial regulations in the public sector that limit an organisation's ability to run up losses, it is prudent to ensure that income must meet or exceed expenditure in all years. Consequently, a cost-effective option that does not meet this criterion may not be viable.

Therefore, within the team the decision was undertaken that the MALT model should still conduct a standard cost effectiveness analysis (CEA), but supplement it with additional, operational level details such as financial flows and their time profiles.

7.2 Cost-effectiveness model

The model was first programmed in Microsoft Excel, when the team considered disseminating the model as a file. Once the decision was undertaken that it should be web-based, programming continued in R.

Patient population

The model simulates the health outcomes and resource use over the lifetime of the cohort. Disease activity is simplified to focus on hospital admissions. Thus the individual's state at any point in time is zero, if they have not been admitted to hospital over the past year; one if they had been admitted once; two if they have been admitted twice, and three if they had been admitted three or more times (Dodd *et al*., 2014).

The natural history model (Figure 31), is based on the assumption that the individuals enter the cohort into state one at their first admission. If during the year they do not get admitted again they would be categorised as zero and transition down the chain. However, if they get admitted to hospital repeatedly over the course of a year their severity state increases to state two and three. They leave the model upon death and it is assumed that all COPD patients die at age 100 years if death had not occurred at an earlier age. The model uses a Markov chain with a monthly cycle length.

Figure 31: The MALT model structure 1 (Dodd *et al*., 2014, p. 8)
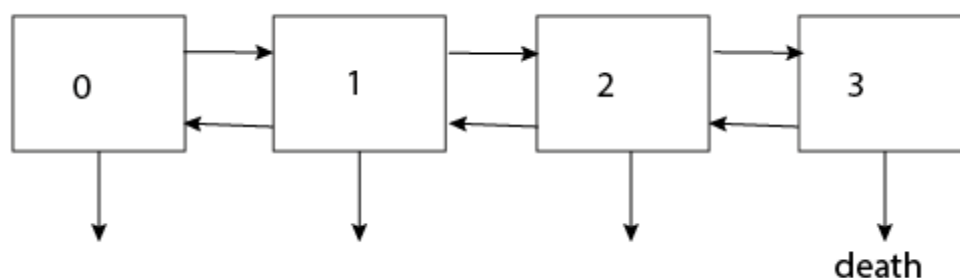
In addition, it is assumed that at any given time the cohort is either receiving the usual care, telemonitoring or are in the post-telemonitoring phase as presented in Figure 31. Each month,

available sets of equipment are allocated to new patients, so they transition from usual care to telemonitoring. Also, after the intervention is finished they move to the post-telemonitoring phase. It is assumed that some patients might receive another referral to a telemonitoring service.

There are a number of inputs that end-users can specify: the length of time between individuals completing their telemonitoring and the unit being ready for redeployment; and the total number of units existing at any point of time.

Figure 32:The MALT model structure 2, (Dodd *et al*., 2014, p. 8)
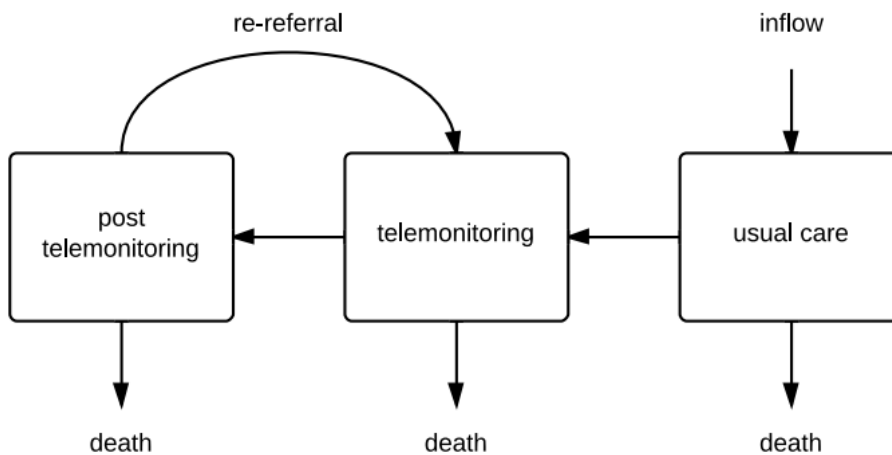
The Hospital Episode Statistics (HES) data on COPD-related hospital admissions was obtained from NHS HES database. Data was obtained for individuals who were admitted to hospital between March 2005 and March 2010, with the primary diagnosis of COPD or related conditions (ICD-10 code J 44) (WHO, 2015). In total, 42 million records, covering the data for each of the 73 PCTs in England and Wales were obtained, and linked to the mortality data from the Office for National Statistics (ONS). With such a database the modellers were able to group individuals

into states based on their hospital use and mortality. The details of the approach are discussed in (Thokala, 2018).

The effect of telehealth on disease progression and mortality was embedded into the model by applying hazard ratios to the transition probabilities. The estimates of hazard ratios were derived from the meta-analysis by Pandor *et al*. (2012) as shown in Table 13 below.

Table 13: Hazard ratios used in the model

Variable	**Hazard ratio**	**Distribution**	**(μ,σ)**
All-cause mortality	0.76 [CrI: 0.49-1.18]	Log-normal	(-0.274,0.170)
All-cause hospitalisations	0.75 [CrI: 0.49-1.1]	Log-normal	(-0.288,0.151)
HF-related hospitalisations	0.95 [CrI: 0.7-1.34]	Log-normal	(-0.051,0.159)

The hazard ratios for mortality and condition-specific hospitalisations were applied to the instantaneous rates of the events derived from HES data. That created 'new' transition matrices representing monthly progression between disease states for those on telemonitoring.

Also, due to the lack of data on potential reductions of ED visits, GP visits or nurse home visits in patients on telemonitoring, we applied data derived from one of the study sites that suggested the individual incidence rate as illustrated in Table 14 below.

Table 14: ED visits, GP visits and nurse home visits incidence rates

Variable	Hazard ratio	Distribution	(μ,σ)
ED visits	0.92	Log-normal	(-0.74,1.14)
GP visits	0.44	Log-normal	(-1.365,1.03)
Nurse home visits	1.14	Log-normal	(0.08,0.32)

Costs

Cohort members accumulated costs and health outcomes in each cycle until death, based on each individual's severity state.

The mean costs for usual care were derived from Pandor *et al* (2012). In addition, some estimates were based on discussions with experts in the MALT study sites. The healthcare costs included the costs of COPD-related hospitalisations, A&E visits, ambulance call-outs, GP visits (both practice and home), and nurse home visits.

The costs of the telehealth intervention were estimated from the interviews with service managers conducted as part of the MALT study. These interviews were conducted by me but are not presented in this thesis. The estimate of an overall cost of telemonitoring per patient per month was embedded into the model. Additionally, we included device cost, the cost of the installation and removal, and the cost of hosting and monitoring derived from the interviews in the study sites. The summary of cost estimates in both categories are shown in Table 15 below.

Table 15. Cost estimates used in the model

Cost category	Cost
Cost of A&E visits	£130
Cost of a COPD-related hospitalisation	£2514
Cost of hospitalisation due to other reasons	£1530
GP visit in a Practice	£46
GP home visit	£104
Nurse home visit	£38
Telemonitoring per patient per month	£175

The estimates were used to develop default cost values that end-users could apply in their calculations if they did not have access to similar data for their own services. Future costs and benefits were discounted at a rate of 3.5% per annum in line with NICE guidance.

Utilities associated with health states

It was assumed that there is no impact on quality of life directly related to telehealth. Whilst it was suggested that greater autonomy and empowerment related to the use of telehealth may have a direct effect on quality of life, there was no robust evidence to show this. The utility estimates

for each health state were derived from a survey conducted on over 300 patients in one of the MALT study sites and are presented in Table 16 below.

Table 16: Health states and utilities in the model

Health state	Mean utility
No hospitalisation	0.589
1 hospitalisation	0.524
2 hospitalisations	0.459
3+ hospitalisations	0.394

Analysis of uncertainty

The uncertainty around key parameter estimates was modelled by the use of Probabilistic Sensitivity Analysis (PSA). Five thousand draws from distributions of treatment effectiveness, health state utility, and disease costs were then used as model inputs.

7.3 Additional assumptions

The model produces estimates of the incremental costs and quality-adjusted life-years gained (QALYs), by estimating the size of the patient cohort for each severity group and analysing how these patients use healthcare resources. In order to capture benefits and costs accrued during patient life times beyond the 5 year time horizon, the population described in the model was

followed for a further 10 years without allowing new patients to enter the usual care or telemonitoring states, by which time the majority of patients had died (Dodd *et al.*, 2014). Additionally, the operational component of the model enables different service configurations to be tested. In order to facilitate such options the following assumptions were embedded into the model.

Firstly, it was assumed that only currently deployed units incurred charges each month. This was in line with the findings from the MALT study sites, as in the majority of the NHS services telehealth providers issued a charge per connected patient. Furthermore, it was assumed that each month spare units are deployed primarily to individuals on usual care. In the case of patients who were provided telemonitoring with a limited duration, they would transit to post-telemonitoring state when their telehealth service comes to an end. If the telemonitoring was provided with indefinite duration individuals did not enter this state.

Prioritized referral to telemonitoring was embedded into the model by using relative weightings applied to different severities of individual on usual care (Dodd *et al.*, 2014). For example, severe COPD patients have 2-to-1 odds over mild severity individuals. The default allocation weightings and initial case-mix were informed from the MALT survey data.

In addition, an option to specify if there is a delay between patients completing their telehealth intervention and the units being ready for redeployment was embedded in the model. In the MALT study sites, the time needed for de-installation and cleaning of telehealth units varied from one to three months, therefore it was assumed that all telehealth services require this additional time.

7.4 Reflection on the models' complexity

Whilst the outputs from the model and its interface were developed specifically for end-users using a series of interviews and user testing, the underlying structure was developed without direct input from the end-users. In addition, whilst I contributed to several design issues (e.g. through the conceptual modelling) many of the technical issues were driven by the modellers within the project team. Reflecting on this development process, I consider that a few issues relating to the Markovian nature of the MALT model require further consideration.

Markov modelling, the most common approach used in economic evaluation of healthcare interventions, uses health states to represent disease progression of a cohort on interventions of interest (Briggs & Sculpher, 1998). Time is considered as discrete time periods called 'cycles' (typically, months), and movements from one disease state to another in the subsequent time period are represented as 'transition probabilities', which are typically constant i.e. Markov models are limited in their ability to 'remember' what occurred in previous model cycles (memoryless assumption). Costs and health outcomes are estimated based on the average time spent by the cohort in each state along with the associated costs and outcomes, to provide a summary of the cohort experience.

Markovian structure is simple, but operationalising it in our case study caused some complexities and some of the associated assumptions (both directly relating to the Markov process and its implementation in this study) are potentially unrealistic. These include (a) the need for classification of the patients into discrete health states, (b) the lack of flexibility in accounting for the patient heterogeneity and (c) the 'memoryless' assumption, where the time spent by the patients in the previous health state is not linked to the disease progression. These limitations are

briefly addressed below and the potential for using an alternative approach ('microsimulation' or 'discrete event simulation') is described.

Firstly, given the need for classification of patients into discrete health states, the way the model accounts for different subgroups of patients (e.g. severity levels 0, 1, 2 and 3), different permutations of supplier, contractor and contractual arrangements for each of the nine activities (e.g. installation, removal, monitoring), makes it quite flexible and capable of representing hundreds of scenarios, but is at the same time very complex.

Secondly, the Markov model assumes that everyone in the cohort is the same i.e. the attributes of individuals do not influence their pathway through the model. To incorporate greater patient heterogeneity in a Markov model, a greater number of health states would need to be generated, each associated with their own parameters relating to costs, utility and transition probabilities. Early discussions with NHS managers within the project identified the existence of various risk stratification tools such as the PARR1, PARR2 and Combined Model (The King's Fund, 2006). These three models produce a continuous score form 0-100, thereby allowing much greater degrees of stratification than the four health states we used. However, even with a much larger number of health states, and assuming that they could be parameterised, it could be argued that the resulting model may be insufficiently flexible to account for the full heterogeneity of the population.

Finally, the model being Markov means that an individual's chance of transiting to a particular state next month depends only on their current state. The number of hospitalisations is not as sensitive as the lung function measure (FEV_1), which is usually employed as the key driver of transitions through the model. The model is also memoryless, (there is no count of how long somebody has spent in the state), ignoring the fact that COPD is a progressive disease and people

who have just entered a disease state are less likely to enter a worse state than somebody who has remained in that state for a period of time. The Markov model cannot treat these two scenarios differently when they are grouped in the same Markov state.

As seen in the paragraphs above, although the MALT Markov model and the analyses have been defined ingeniously to address these limitations to some extent, other modelling approaches may be required (and more suitable for addressing the limitations, especially when modelling more complex diseases than COPD). Simulation modelling approaches, especially microsimulation or discrete event simulation (DES), are better suited for addressing these limitations. As discussed in Chapter 1, DES is an individual level modelling approach using entities, attributes, and events. Entities are individual objects (e.g. patients) with attributes (e.g. age, sex), that experience a sequence of events, e.g. hospitalisation or myocardial infarction (Caro & Möller, 2016). The flexibility provided by the DES approach addresses the limitations of Markov models described earlier – (a) DES does not require the need for classification of the patients into discrete health states, (b) DES has the flexibility in accounting for the patient heterogeneity by using their attributes, and (c) DES can track the time spent by the patients in the previous health state and use it to link to the disease progression. It should be noted that DES is being increasingly used for HTA (Caro *et al*., 2010; Caro *et al*., 2015). The advantages and disadvantages of DES for HTA is highlighted elsewhere (Caro & Möller, 2016; Karnon & Afzali, 2014).

7.5 The interface

The model is stored on a University of Sheffield server. Upon registration on the website (www. malt.sheffield.ac.uk) each user is provided with a password protected profile that enables them to access the tool and to store the results of all analyses.

The end-users have the option to choose one of the following versions of the model: Light, Regular, Full and PSA. The Light version of the model requires input of the patient case-mix and telehealth deployment, with the remaining inputs automatically generated using default values. This version was developed to enable quick and simple exploration of service configurations by stakeholders that do not need to explore operational details and financial flows.

The Regular version of the model allows the end-user to customise the inputs in the following categories: the overall number and the severity of patients, the number of devices in use, and the costs of each of the components of the telehealth service. This version was designed to allow users to examine the most important operational issues of telehealth, whilst avoiding the need for the most complex data requirements and interpretation of PSA.

The Full and PSA versions of the model were designed for more advanced users. In addition to the ability to change the main parameters these versions also include more advanced options, such as the ability to change the efficacy estimate for telemonitoring and the rates of resource use. Lastly, the PSA version includes the option to run a probabilistic sensitivity analysis on the model's results.

7.6 Appraisal of a hypothetical service

To demonstrate the capabilities of the MALT model, two scenarios were evaluated for a hypothetical PCT. The choice of the PCT was made at random.

In this scenario, telehealth intervention was compared to no telehealth. The telehealth intervention was deployed for 60 months, with a maximum duration of 9 months. It was assumed that there is a 2-month delay before the next patient can start using the same unit. This is to allow for de-installation and cleaning. Also, 4% of patients each month would abandon the monitoring. The

initial cohort comprised of 200 people in health state 1, 600 people in health state 2; and 150 people in health state 3. Also, 300 units of telehealth equipment were planned to be used.

- Scenario A: Telehealth equipment will be purchased for £45 per set. Med Tel Solutions will deliver the service for £80 per each connected patient per month.

- Scenario B: Telehealth equipment will be on lease for £45 per patient per month. Parts of the service (installation, removal and patients training for the whole cohort) will be outsourced to an external company for £15,000 a year.

Firstly, the model shows how the cohort of patient will progress through time as presented in Table 17 below.

Table 17: The case-mix of patients

Year	Total	Deaths	New patients	Net population change
1.	12,305	294	444	150
2.	14,278	270	444	174
3.	16,354	275	444	169
4.	18,301	291	444	153
5.	20,043	309	444	135

The model assumes steady growth of the number of COPD patients that enter the cohort each year. In that particular region it is expected to have 444 new patients each year. The number of deaths will remain steady around 300 cases per year.

The model predicts that over the course of 5 years, the majority of patients in states one, two and three will have telemonitoring or be on telemonitoring. At the end of year five, the population of COPD patients in this particular PCT will be dominated by the least severe (state 0) cases.

Healthcare related resource use can be estimated based on these population estimates as presented in Table 18 below.

Table 18: Resource use over the course of 5 years

Year	GP visits	Nurse home visits	A&E visits	COPD-related hospitalisations
1.	£7,528	£20,082	£859	£540
2.	£8,504	£22,683	£1,113	£706
3.	£9,247	£24,666	£1,189	£757
4.	£9,852	£26,278	£1,227	£784
5.	£10,355	£27,620	£1,251	£801

The most expensive resources are nurse home visits followed by GP visits. This data can be further explored by generating a graph, detailing aggregated costs in each of the severity groups, as presented in Figure 33 below.

Figure 33: Healthcare related cost estimates in the telemonitoring patients
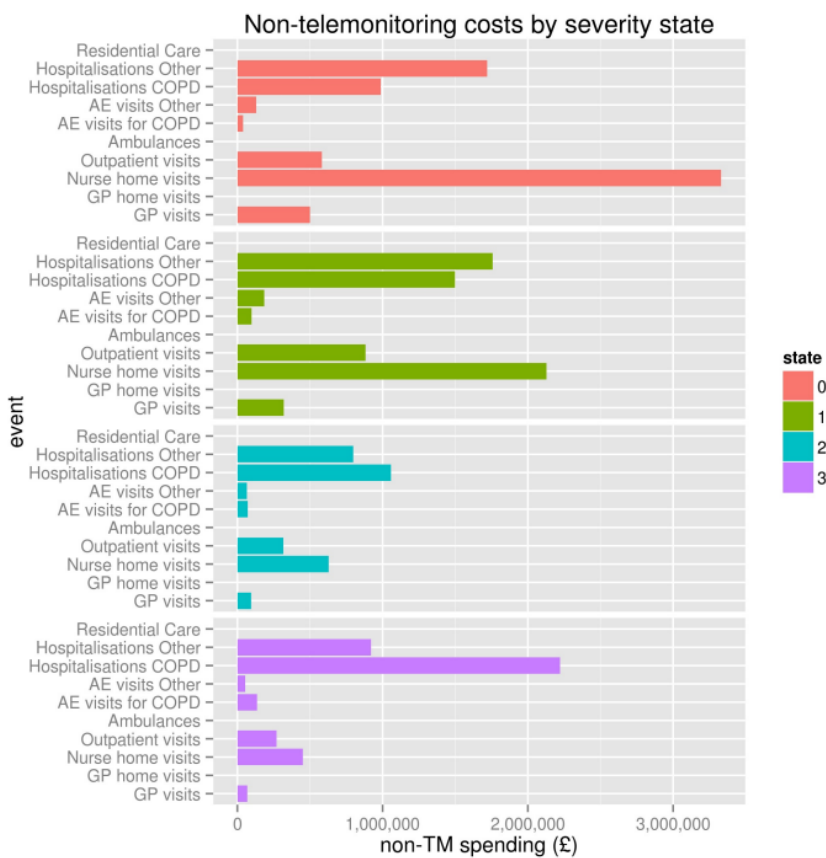
Finally, the model has the capability to present the disaggregated costs for each scenario as presented in Tables 19 and 20 below. In this particular scenario, the results suggest that Scenario B leads to an additional 6 QALYs and an additional cost of £1,584,179. This suggest that the ICER is higher than the threshold of £20 000 per QALYs, however, the model does not calculate the ICER, this is an option that end-users have to perform manually form the data provided.

Table 19: Disaggregated costs and QALYs Scenario A

Category of cost	Year 1	Year 2	Year 3	Year 4	Year 5
Device cost	£0	£0	£0	£0	£0
Hosting/Installation/ Removal	£160,429	£160,431	£160,431	£160,431	£160,431
A&E visits	£313,312	£337,312	£344,652	£352,980	£382,247
Hospitalisation costs	£5,245,473	£5,646,812	£5,767,789	£5,879,615	£6,439,917
Outpatient visits	£1,970,667	£2,034,683	£2,052,423	£2,068,739	£2,126,181
GP visits	£948,664	£972,075	£978,397	£984,590	£1,005,909
QALYs	3567	3658	3681	3707	3791

Table 20: Disaggregated costs and QALYs for Scenario B

Category of cost	Year 1	Year 2	Year 3	Year 4	Year 5
Device cost	£64,171	£64,172	£64,172	£64,172	£64,173
Installation	£90,241	£90,242	£90,242	£90,242	£90,243
Hosting	£48,128	£48,129	£48,129	£48,129	£48,129
Monitoring costs	£12,032	£12,032	£12,032	£12,032	£12,032
Removal	£90,241	£90,242	£90,242	£90,242	£90,243
A&E visits	£311,454	£337,157	£344,901	£352,377	£392,351
Hospitalisation costs	£5,263,013	£5,644,792	£5,765,448	£5,886,715	£6,450,505
Outpatient visits	£1,987,501	£2,036,169	£2,051,638	£2,067,553	£2,132,970
GP visits	£941,382	£971,710	£978,292	£984,565	£1,012,039
QALYs	3540	3656	3681	3706	3815

In this particular hypothetical evaluation, the results indicated that the provision of telehealth service using the external company and an option to purchase the equipment is more favourable over the 5 years in terms of costs and QALYs, than an option to lease the equipment and to outsource the parts of the service. Whilst many other scenarios are viable in the evaluations of telehealth, the provided hypothetical example was only intended to give an indication of capabilities of the model and the findings with respect to cost-effectiveness were only given to signpost to specific features of the model.

7.7 Summary

This Chapter provides technical details of the MALT model and an overview of its capabilities using a hypothetical scenario. It shows the method to define the states in the Markov model and to estimate transition probabilities from the routinely collected hospitalisation data. A hypothetical scenario was used to demonstrate how the model provides estimates that can be used to assess cost-effectiveness of any user-specified scenario.

8. Usability evaluation

This chapter presents an overview of the methods used in the usability testing of the MALT model interface, and the results from three rounds of usability tests conducted between September and December 2014. The usability evaluation was deemed essential to ensure that the complexity of the design of the web interface was managed, and that end-users could work with the model unaided, even if they did not have modelling experience.

8.1 Introduction

Prior to conducting the usability evaluation, it is important to specify the profile of end-users and give examples of the tasks they are likely to conduct with the model (Dumas & Redish, 1999; Nielsen & Mack, 1994).

The major end-users and their needs were specified based on the knowledge acquired during qualitative interviews (as discussed in Chapter 6), MALT study workshops, and events. My understanding of the end-user profile was shared with the wider MALT research team during study steering group meetings, and senior researchers agreed on an exhaustive list of professional job titles (presented in Table 21).

Table 21. End-user job titles and perceived application of the MALT model

Professional job title	Potential application of the MALT model	Computer skills and domain expertise
Commissioning manager (CCG)	To calculate the total cost of telehealth services To explore different options for alternative service models	Unfamiliar with the intricacies of the MALT model itself, but understands its purpose. Uses desktop PC in an office. Familiar with spreadsheets and word processing
Clinical manager of community care organisation	To calculate the workload of staff and resource use in different components of the service To conduct detailed analysis of the impact of telehealth on the service	Unfamiliar with intricacies of the MALT model or health economics
Director of operations of community care organisation	To explore the workload and financial implications of different options To understand longer-term changes to the system To explore interactions with other services	Good understanding of finances and contracting
Director of finance of community care organisation	To explore financial implications of multiple options for the service	Understands how the overall system needs to work

Professional job title	Potential application of the MALT model	Computer skills and domain expertise
	Financial flows and contracting between other providers	
Clinical managers NHS Trust	To calculate the number of hospitalisations under one or two options	Understands how care is provided in the system but not the contracting arrangements. Unfamiliar with the details of the community care telehealth services
Commercial manager private care provider	To have a detailed understanding of income to their organisation To have a detailed understanding of financial flows between all stakeholders to help with client discussions To have a detailed understanding of workload implications for all stakeholders to help with client discussions To explore multiple options in detail	It is assumed that managers will want to be familiar with the intricacies of the MALT model itself and be capable of using all its functionality They will use the model in discussions with sales managers
Sales manager private care	To have a detailed understanding of financial flows between all stakeholders to help with client discussions	It is assumed that some managers will require some understanding of the workings of the model but will not use its entire functionality

Professional job title	Potential application of the MALT model	Computer skills and domain expertise
	To have a detailed understanding of workload implications for all stakeholders to help with client discussions	They will use model in discussions with multiple stakeholders across and within organisations
Analyst / Consultant	To explore the impact of multiple service configurations on overall costs and cost-effectiveness of the system Will want printed reports and will save multiple options Not interested in details of the telehealth service	They are likely to request to be familiar with the intricacies of the MALT model itself and be capable of using all its functionality They will use the model to advise the CCG manager in terms of cost-effectiveness

There was a consensus that the utilisation of the model will vary depending on the individual computer skills, knowledge of telehealth, economic modelling, and the task itself.

8.1.1 Aims and objectives

The aim of the second part of the case study was to test the interface of the MALT model with expert and non-expert users in order to identify and eradicate obvious problems of usage. The tool underwent further development after each of the test rounds, and subsequent versions were tested on new sets of users.

8.2 Methods

Criteria for the choice of the task

The study used two hypothetical scenarios, based on the telehealth procurement contract from one of our study sites and our knowledge of telehealth provision models. The tasks were designed to cover all features of the proposed model. The first and second tasks required inputting numeric data and running the model for each of the given options to decide which company offered a better value-for-money service. The second part of the exercise required evaluation of two options for service improvements. The main criteria for the task were: simplicity, novelty, and the potential to explore problematic areas of the model. The copy of the participants' task sheet and the list of the actions needed to complete these tasks are detailed in Appendix A.8.3, and A.8.4.

8.2.1 The participants and the sample size

Nielsen (1994) proposes that a sample of 5-10 participants in each of the software development cycles is sufficient to detect the majority of the usability problems. The obvious problems of usage and flaws in design, if they exist, would be identified very early on, and none of the additional tests would add anything new to the findings.

The goal was to study two groups of users: PhD students and telehealth and economic modelling experts. A pilot test of two users was conducted to test the tasks and practice the procedure. Initially ten subjects were recruited (five casual users, and five expert users), but after the results of the first round of tests revealed a significant number of problems, it was decided to run an additional round of tests with postgraduate students. The final number of participants was therefore 12 students and five professionals.

The PhD student participants (six females and six males, aged 19-35), were recruited from the School of Health and Related Research postgraduate student population, after the study was promoted on the school mailing list. PhD students were recruited because of their knowledge and NHS work experience. The majority of participants had a background in either health economic modelling or medicine.

A copy of the email used during recruitment is presented in Appendix A.8.2. Participants were informed about the aims of the study and assured that they would be rewarded for their time. A copy of the PhD students' information sheet is presented in Appendix A.8.5 and Professionals information sheet in Appendix A.8.6.

The expert users (three females and two males, aged 30-55), were drawn from the Rehabilitation and Assistive Technologies team, the MALT project team, and the telehealth industry. All participants were experts in telehealth, NHS commissioning, or health systems research, and were chosen due to their experience.

For the purposes of the study, simple demographics of the users were analysed, such as their age, professional background, IT experience, modelling experience, and knowledge of telehealth domain. These were assumed to influence the problems participants encountered when interacting with the tool. All participants were fluent English speakers. The basic characteristics of the participants are presented in Table 22 below.

Table 22: The basic characteristics of the participants of the usability test

	First round of tests (n=7) PhD students	Second round of tests (n=5) PhD students	Third round of tests (n=4) Telehealth and Health systems experts
Gender			
Male	5	1	2
Female	2	4 (2 recordings lost)	2
Professional background			
Medical or health	4	2	3
Health economics	2	1	
Other (engineering, computer science)	1	2	1
IT experience			
>10 years	1	0	0
< 10 years	6	5	4
Prior modelling experience			

	First round of tests (n=7) PhD students	Second round of tests (n=5) PhD students	Third round of tests (n=4) Telehealth and Health systems experts
Yes	3	1	3
No	4	4	1
Prior telehealth experience			
Yes	4	2	4
No	3	3	0

The participants who took part in the usability testing all had higher degrees in medical, health economics, or computer science disciplines. Their experience in economic modelling was varied, and all of the groups included participants with no prior experience in modelling or no prior experience in using decision analytical models.

The experiment took place in a controlled laboratory setting on university premises, and I was responsible for running the experiment. Software developers were on hand to observe the first few tests. Participants had access to a computer with the MALT model platform and writing paper.

8.2.2 The experimentation process

Of the initially recruited sample of people, those who expressed an interest in taking part in this research received the study information sheet (Appendix A.8.5 and A.8.6). On the day of the test, all of the participants were asked to sign a consent form (Appendix A.8.7), and given a brief explanation of the purposes of the project and abilities of the model - though not a formal tutorial on how to use it, as I intended to follow the cognitive processes of first-time users. Participants sat directly in front of the computer. They were directed to read the tasks script (Appendix A.8.3), and told to execute the task using the MALT model. They were informed that they should articulate any thoughts they might have, as if they were alone in the room and talking to themselves, and that I would remain silent during the experiment unless they asked for help. They were given an opportunity to ask questions. Before proceeding to the actual test, they were given a few minutes to explore the site, to register on the website, and to sign up. They were also offered a warm-up exercise to practise thinking aloud. Following each series of tests, participants were asked additional questions concerning the usability of the model and their experience with the tool.

I was present during each session; though social interaction was kept to minimum. I only intervened if participants stopped talking for more than 30 seconds, using a 'keep talking' prompt. I also answered questions, and in some cases offered help if participants were unable to proceed.

Each session was individually run and took between 45 and 60 minutes. A copy of the participants' task sheet (the 'lab sheet') and the list of action needed to solve them are presented in Appendix A.8.3 and A.8.4.

Users were video recorded solving the task, with the camera pointing at an angle so that the screen was also visible. This method was found to be sufficient to capture the features of the interface to

be improved. More sophisticated equipment, such as an eye-tracking device, was not available. During the task, I recorded participants' verbalisations, facial expressions, and body movements.

8.2.3 Data Analysis Methods

Verbal protocol data

After completing each test session, I discussed and compared my reflections with other observers (if they were present) and transcribed the verbalisations. Transcripts were developed in the following way. Firstly, all verbalisations were transcribed verbatim, indicating pauses, and emphasising verbal comments of discontent and similar. The video recordings were then watched again to check for accuracy, to match verbalisations with audio transcripts, and to add more explanation to users' actions. A list of immediate usability problems was prepared to ease further analysis. These stages of data analysis are documented in Figure 34.

Figure 34: Verbal protocol data analysis phases
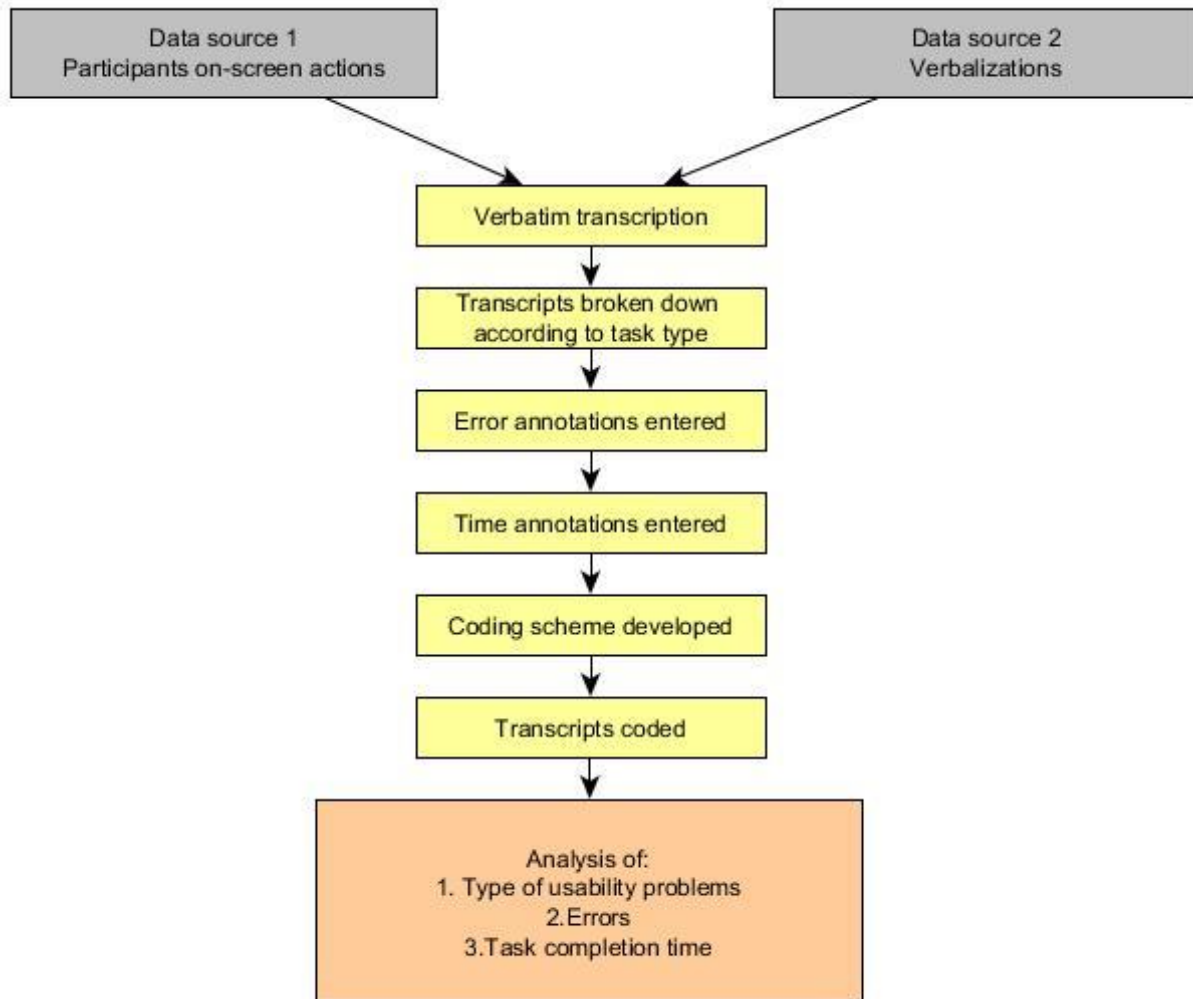
The usability problems detected in each of the test rounds were categorised, using a four-point scale, as unusable, severe, moderate, or irritant (Rubin & Chisnell, 2008). Problems labelled unusable (U) prevented users from completing the task. Problems deemed severe (S) caused confusion for users when performing the task, and were a reason for seeking help. The moderate (M) problems were those which users could intuitively solve themselves; and the last, labelled irritants (I), caused discomfort when running the model and were reasons to complain.

To create a final list of usability problems, I categorised them by their 'criticality': a combination of severity and probability of occurrence (Rubin & Chisnell, 2008). This was to enable the software developers to structure and prioritise the work required to improve the tool before the next test round could begin. To rank the estimated frequency of occurrence, the percentage of total users affected by the problem and its probability of occurrence were estimated. Using this method, very severe problems that made the tool unusable for everyone were the first to be tackled by software developers.

Task performance measures

The data was also analysed quantitatively, focusing on the number of errors made by each participant and their completion time. Errors were calculated and analysed using Rubin's (2008) method. For each completed correct action, a participant received a point. However, to simply calculate the number of errors per person would mean that those who completed no tasks, and thus made no critical errors leading to it being jeopardised, would achieve better results (fewer errors). For example, choosing a wrong version of the model (Light instead of Regular) would mean that service cost information could not be entered, but some participants would run the model anyway, assuming they had completed the task. A point system allowed for the calculation of a percentage of successfully completed tasks. The model would thus run even if the participant chose the wrong option. The list of actions needed to successfully solve both tasks, and the criteria by which points were justified, is provided in Appendix A.8.4.

Completion time

Dumas's (1999) method was used to analyse completion time. A mean completion time was estimated for the whole test group, for each of the tasks, rather than focusing on the individual's completion time. The mean completion time was calculated using all the participants' completion

times and the total number of participants. Time was used as an indicator of underlying usability problems, bearing in mind that three factors have an impact on users' performance: firstly, thinking aloud slows people down; and secondly, an individual's articulacy can influence their completion time. Thirdly, participants' computer and cognitive skills have, to some extent, have impact on how well they perform, although all participants were assumed to have a similar level of computer skills.

8.2.4 Development of the coding scheme

After completion of each test session, the data from verbal protocols were transcribed and coded. The coding of the transcripts was necessary before undertaking a more systematic analysis of the verbal protocols. The coding scheme was developed based on the literature, and devised to identify the processes that occurred during the computer-assisted problem-solving.

Researchers in cognitive psychology and behavioural sciences recognise several stages of problem-solving and computer-assisted problem-solving. The descriptions of the process by Andree (Andree, 1988) and Simon (1977; 1958) are largely similar, and include the following stages:

1) Problem definition (What type of data is required? What will be the output of the program?)

2) Problem analysis (What are the inputs, the desired results and additional requirements or constraints on the solution?)

3) Executing the plan

4) Reviewing the solution

Courtney (2001) outlined the stages that occur during computer assisted decision-making as: (1) problem recognition, (2) problem definition, (3) alternative generation, (4) model development,

(5) alternative analysis, (6) choice, and (7) implementation. According to the author these stages overlap and blend together. Based on these publications the following coding scheme was developed.

1. Understanding of the problem (reading)

2. Generating a solution

3. Data inputting

4. Seeking help

5. Seeking reassurance

6. Unexpected behaviour

7. Making errors

8. Reasons for being confused

9. Reason for being irritated

10. Positive comments

11. Suggestions for improvements

During data analysis, the codes were condensed and collapsed to generate the final coding scheme, as shown above. Each protocol was coded using this coding scheme, and the data analysed using the content analysis method.

8.2.5 Pilot test

The pilot test was conducted prior to the main data collection in order to test and develop the schedule and practise the procedure. The test was conducted on two subjects and followed the format described in the Appendix A.8.1.

In addition, participants were asked whether they had any suggestions as to how to improve the tasks, whether they found any of the scenarios unclear, and whether they had any comments. The researcher took time to reflect on the proceedings of these tests, and discussed them with the research team, comprising economic modellers and software developers. Field notes were compared and used in this process.

The feedback from the pilot test identified only a small number of changes to the schedule. The main changes included more detailed numeric data description, for example specifying if a certain figure was per patient per month or per patient per year. The pilot also revealed that not everybody could complete the tasks within the given timeframe, and one of the tasks was thus removed. Given the minimal changes required to the task schedule, data from the pilot was incorporated into the main analysis.

8.2.6 Ethical approval

The study received ethical approval from the University of Sheffield, School of Health and Related Research Ethics Committee (see Appendix A.8.8 for a copy of certificate).

Several measures were incorporated to ensure the research was conducted in line with guidelines for ethical conduct. Firstly, participants were provided with a study information sheet at least 24 hours prior to the test, detailing the purpose of the study, the experiment process, how the data would be used, and the details of the complaints process. On the day of the interview, participants were first asked to read the information sheet, and given an opportunity to ask questions about the study. They were then asked to complete a consent form that stated that they consented to their sessions being video recorded, and there was an option to consent only to audio recording. They were reminded that participation in the experiment was voluntary and that they were free to withdraw at any time. Student participants received a voucher for £10 for their participation.

Upon enrolment, all participants were given a unique ID number, and these were used throughout the study. Identifiable information was stored securely in a password-protected files, disconnected from the network. Original video recordings were archived and stored securely, and destroyed once transcripts had been made. All of the transcripts were anonymised and stored on a non-shared and secure drive at the University of Sheffield.

8.3 Results

8.3.1 Sample

Over 15 hours of video recordings were created during the study. Verbal protocols were transcribed verbatim. One file had background noise that made transcription difficult and it had to be converted to audio and cleaned using Audacity software. During data collection, recordings from two sessions were lost due to problems with equipment, and analysis of the data for these two participants was based entirely on the field notes. After transcription, the data amounted to 26 pages of text, and this was entered into Nvivo 11 (QRS International., 2012) software for coding.

8.3.2 Usability problems detected in the first round of tests

The first round of tests was conducted on seven student users. A set of 12 usability findings was observed during this phase. This included both the usability problems (such as the problems with navigation) and problems detected through observation of usage.

Overall, participants' first impressions of the MALT model were positive. They found it easy to locate the sections of the model in which the data was intended to be inputted. In particular, they

understood that the main spreadsheet would be used to input the data, and that in order to run the model they had to press the button marked 'Run the model'.

However, the first usability tests revealed serious problems with the design of the interface. Firstly, the specialist language prevented some users from proceeding with the tasks. Some of the terms confused all of the participants and were a reason for seeking help. Secondly, all of the participants were confused by the meaning and purpose of the default values. Moreover, a few people experienced problems with navigation around the tool. In particular, they had difficulty identifying the correct tab, after they had switched on to explore the model's additional features. The majority of the usability problems occurred more than once, and their details are listed in Table 23 below, along with the frequency of occurrence and category of severity.

Table 23: List of usability problems detected in the first round of tests

No.	Type of problem	Details	Number of times it occurred
1.	Does not understand	'Allocation by severity' **(U)**	6
		'Time horizon' **(U)**	5
		'Redeployment time' **(U)**	4
		'TM deployment' **(U)**	7
		'Activity vs. cost' **(U)**	7
		"Variable cost", "Fixed cost" **(U)**	6
		"Technical triage" **(U)**	5
		"A&E" **(I)**	1
2.	Uncertain of the purpose of "Next" and "Previous" buttons **(I)**	-	7
3.	Issues with the default values	Forgets to delete them **(M)**	3
		Forgets to replace them with zeros **(M)**	4
		Confused by them **(I)**	7
		Tries to fill in all blanks with zeros manually **(I)**	6

No.	Type of problem	Details	Number of times it occurred
4.	Attempts to add cost figures together	Training and set up (**M**) Triage cost over 5 years (**M**)	2 2
5.	Unclear whether hospital costs should be deleted or left as defaults (**I**)	-	2
6.	Unclear how to go back to main page (**I**)	-	2
7.	Unclear which version of the model to choose (**M**)	-	5
8.	Selects wrong cost category for	Triage (**S**) Communication (**S**)	5 4
9.	Attempts to add patients as stakeholders (**I**)	-	1
10.	Finds the term "payer" confusing (I)	-	1
11.	Ignores all cost categories on the right of the screen (**S**)	-	1

No.	Type of problem	Details	Number of times it occurred
12.	Selects "log in" instead of "register" **(I)**	-	6

- (U) unusable, (I) irritant, (M)moderate, (S) severe usability problems

8.3.3 Changes introduced after first round of tests

A large number of amendments were made as a result of the first round of usability tests. Using Rubin's classification, problems 1, 2, and 3 were given priority in the development work, as they were both the most frequent and the most critical. As a result of this work:

- ❖ The most confusing terms—such as 'time horizon', 'redeployment time', 'allocation by severity'—were replaced.

- ❖ Abbreviations such as 'TM deployment' were completely eradicated.

- ❖ 'i button' icons were embedded into the interface, to display an explanation of the most confusing telehealth vocabulary.

- ❖ A lengthier explanation was added to the sections where default values appeared, explaining their purpose in more detail. Their use was also made optional. The users could remove them with one click if they did not wish to use them to populate the model.

- ❖ To facilitate the comparison of the results of multiple model runs, an option to enter the name for each of them was provided.

- ❖ A colour scheme was introduced to mark types of model runs (Light/Regular/Full/PSA) and the description of each option was re-written to be more comprehensive and easy to understand.

❖ A colour scheme was introduced to mark types of the model runs (Light/ Regular/Full/PSA) and description of each option was re-written to be more comprehensive and easy to understand.

One of the most discussed issues was whether to change the layout of the page. Based on my observations, I proposed a sequential design, where sections appeared on the screen one at the time and users could move from one section to the next by clicking the 'next' button. Adding a little progress bar at the bottom of the page would allow users to assess their progress through the application. At the time, the project team decided that the layout would not be changed, and that the 'next' and 'previous' buttons would be left on the page. Pictures 4 and 5 below show the first section of the interface before and after the changes had been implemented.

Picture 4: The MALT model interface version 1

Picture 5: The MALT model interface after amendments

Compulsory. In this section, you specify some basic information about the telemonitoring scenario you want to consider. You choose the roll-out and allocation plan for telemonitoring deployment, whether you are interested in using telemonitoring for COPD or heart-failure, and your primary care trust. The trust is needed because the disease progression and resource use are informed by hospital episode statistics specific to a given trust.

Name	* Primary care trust 🛈	* Disease type 🛈	Priority allocation to most ill? 🛈
[ ]	- Select a trust - ▾	Heart failure ▾	⦿ Yes ◯ No

Add an optional name that can be used to identify previously run models

Duration of model run 🛈	Duration of a deployment 🛈	Redeployment Delay 🛈	Dropout rate 🛈
60	months	9	months	3	months	0.00	% per month

↓ Next Section

8.3.4 Usability problems detected in the second round of tests

The subsequent version of the model interface was tested on a further five students. During this round of tests, nine usability problems were detected, as presented in Table 24 below. The severity of the problems was categorised as previously, and they were deemed less severe.

Table 24: List of usability problems detected in the second round of tests

No.	Type of problem	Details	Number times it appeared
1.	Selects 'Log in' instead of 'Register' **(S)**	-	3
2.	Does not understand	'Installation' **(S)**	1
		'Initial case mix' **(U)**	4
		'Redeployment delay' **(U)**	2
		'Variable cost', 'Fixed cost' **(U)**	3
		'Technical triage' **(U)**	1
3.	Uncertain of the purpose of 'Next' and 'Previous' buttons **(I)**	-	2

No.	Type of problem	Details	Number times it appeared
4.	Issues with the default values	Forgets to replace them with zeros (**M**)	2
		Confused by them (**S**)	1
5.	Attempts to	Delete all hospital cost default values as assumed that they are not needed (**U**)	2
		Manually divides installation cost to calculate removal cost per patient (**U**)	1
6.	Unclear how to assign patients to severity groups (**U**)	-	1
7.	Unclear how to go back to main page (**M**)	-	2
8.	Unclear which version of the model to choose (**M**)	-	2
9.	Selects wrong cost category for	Installation (**S**)	1
		Equipment (**S**)	1

- (U) unusable, (I) irritant, (M)moderate, (S) severe usability problems

Fewer people were confused whilst registering on the site, choosing 'Log in' instead of 'Register'. Despite the language being simplified, some still experienced problems with some

of the terms. The two most problematic were 'initial case-mix' and 'variable and fixed costs'. Similarly, default values were still problematic for two of the users, as they were not sure which one they could delete and which should stay in. Two people had problems with navigation and choosing the right version of the model.

The second round of usability tests revealed that the model interface needed further development on aspects of the language used. Similar types of problems appeared in this round, with users assigning costs to wrong categories, forgetting to replace blanks with zeros, and struggling to understand some of the terms used. The frequency of the problems was, however, lower, which may be due to the fact that more explanation and 'i buttons' had been provided in each of the sections of the interface.

8.3.5 Changes introduced after second round of tests

One unanticipated discovery during the second round of tests was that in the section where cost information was to be inputted, a few people had ignored two columns on the right-hand side of the screen because of the impression that they belonged to a different section. The spreadsheet was then converted to include all financial columns on the left-hand side, and stakeholder names on the right. Pictures 6 and 7 on the next page show the interface before and after the changes were implemented.

Picture 6: The cost section of the MALT model interface before changes

ACTIVITY VS COST TYPE ⓘ

	£ FIXED COST (per month)	£ VARIABLE COST (per patient per month)	SUPPLIER	PAYER	£ ONE-OFF COST	£ MONTH
Quick Set			▾	▾		
Device costs	0	32	▾	▾	2500	12
Installation	0	45	▾	▾	2320	12
Training and set-up	300	0	▾	▾	0	0
Communications/hosting	0	24	▾	▾	0	0
Monitoring	0	6	▾	▾	0	0
Technical triage	0	0	▾	▾	0	0
Clinical triage	0	0	▾	▾	0	0
Maintenance/back office/admin	0	0	▾	▾	8500	6
Service review and modification	0	0	▾	▾	5000	24
Removal	0	45	▾	▾	4000	24

↑ Previous   **✔ Run the model**

Picture 7: The cost section of the MALT model after the improvements

Optional. In this section, you can specify the costs of different elements of the telemonitoring service. Costs are fall under pre-specified category headings, but you can change whether they are fixed, or variable, or one-off, and you can specify which organisations in your area pay and receive these tariffs.

These costs are defaulted to the national average

🔄 CLEAR COSTS	£ FIXED COST (per month)	£ VARIABLE COST (per patient per month)	£ ONE-OFF COST	£ MONTH (for one-off cost)	* SUPPLIER	* PAYER
Quick Set						
Device costs	0	32	0	12		
Installation per patient	0	45	0	12		
Training and set-up	0	0	0	0		
Communications/hosting	0	24	0	0		
Monitoring	0	6	0	0		
Technical triage	0	0	0	0		
Clinical triage	0	0	0	0		
Maintenance/back office/admin	0	0	0	6		
Service review and modification	0	0	0	24		
Removal	0	45	0	24		

⬆ Previous Section   ✔ Run the model

An option to have the name of the stakeholder filled in multiple fields by one click was added to expedite completion time. Many people were reminded to go back to fill in these, something that often caused irritation and frustration.

8.3.6 Usability problems detected in the third round of tests

The third round of tests was conducted with four telehealth and health systems experts. Three usability problems were detected as presented in Table 25 below.

Table 25: Usability problems detected in third round of tests

No.	Type of problem	Details	Number times it appeared
1.	Unclear of	How to classify costs as either variable or fixed **(I)**	4
2.	Assumed that	Blank fields are recognised by model as zeros **(I)**	1
3.	Irritated by	Arrows in the fields indicating totals – e.g. the number of hospitalizations **(I)**	4

- (I) irritant usability problems

The sessions with telehealth and health systems experts resulted in a further three usability problems being discovered, as listed below.

❖ The majority of participants had difficulties in assigning costs to certain categories (variable or fixed) and asked for help or reassurance at that point.

❖ One person stated that he usually assumes blank fields to be zeros, and was irritated that the model asked for the missing information to be filled in before the model could be run.

❖ All participants pointed to the arrows next to the 'Total' field as meaningless, and as suggesting that the total value could have been changed by the unit.

However, the problems were categorized as 'irritant', which suggested that the interface had been sufficiently improved. The test provided very useful feedback, as the participants were fairly senior and had the advantage of several years' experience in the telehealth field. They provided a lot of comments about the practical implications of using this tool and suggestions for interface improvements. For example, one of the participants stated:

I need to say that I think it will be difficult to map the real world challenge into the buckets provided [cost table], and so once you have done that categorisation, then you can get nice neat results. As it stands, people will have to be quite motivated to use it. (Telehealth expert, PH3)

I'm guessing in real life someone making decisions would not have these numbers listed out like that — and it will be a challenge to have them translated to correspond to categories in the model. (Telehealth expert, PH3)

And another:

> *I am just thinking whether it would be possible to have something like a test of reasonableness. When you have these totals displayed — they are summing up automatically — to have something to show you how many devices you physically have when you take into account dropout rate. That would let you have a track of dropout rate — if you are a commissioner you want to know 'how much you physically have at this moment of time'. (Telehealth expert, PL3)*

8.3.7 Changes introduced after third round of tests

Only minor changes were implemented after this round, including removal of the option to change the totals, as they are set to be calculated automatically.

8.3.8 Task performance measures

Prominent researchers in the field (Dumas & Redish, 1999; Nielsen, 1994; Rubin, 1994) are supporters of the view that different data generated throughout the usability evaluation should be used to look at how they support each other. For example, long task times and frequent errors might point to the same usability problems. I triangulated the data on the frequency of errors and completion time to explore my earlier findings on the source of interface usability problems.

Errors

An analysis of errors was conducted for 14 of the participants, as two recordings had been lost during data collection. In the analysis process, I used a script detailing the actions needed to successfully complete each task (as shown in Appendix A.8.4). An error was counted if a participant had chosen an action different than the one specified, or had not asked for help or

reassurance. If an error had been noticed and reflected upon, it would not be counted. In the case of asking for help, the correct answer was given. Some people, however, had a habit of checking to reassure themselves after each entry, which thus improved their ratings. Figure 35 below shows the number of errors made by each participant in task one (blue bar) and task two (orange bar).

Figure 35: Number of errors in task one and task two per participant
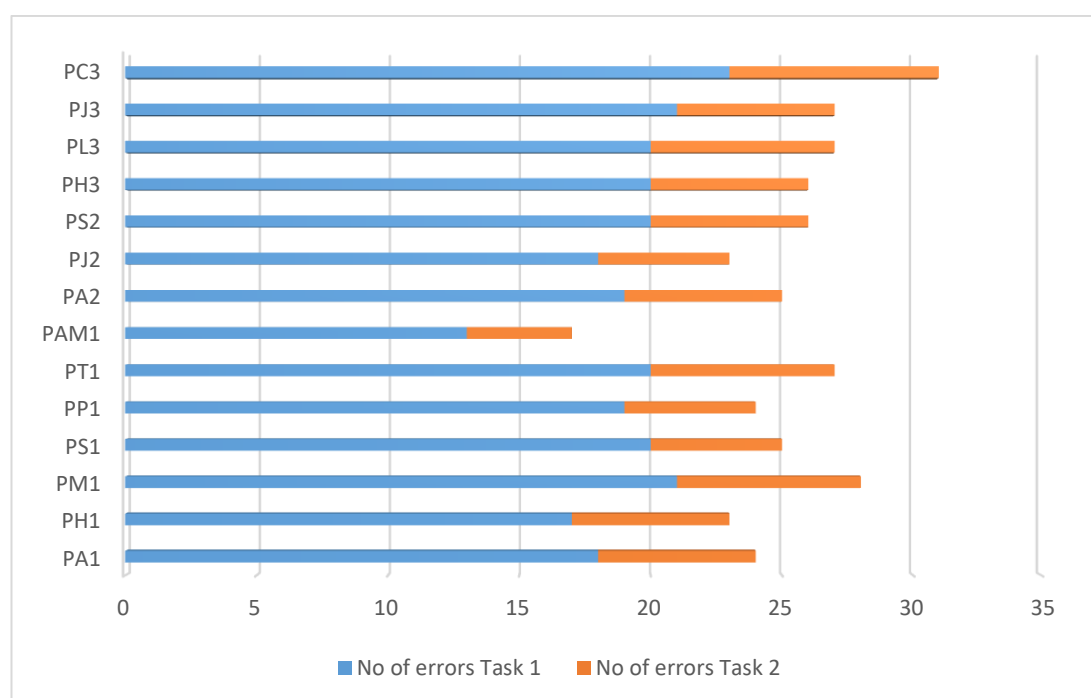
The number of errors made during task one ranged from 13 to 23; whereas in task two the lowest number of errors was four and highest was eight. Participants made fewer errors in task two, which suggests that participants found it easier to execute a task once they had completed the first scenario.

Overall, the completion rate for all three groups of participants was high, ranging from 53% to 98%, as showed in Figure 36 below. The first and the second group of participants (PhD students)

are marked as green and blue bars, the third group (telehealth and health systems experts) are shown as orange bars. The trend line was added to facilitate interpretation.

Figure 36: Percentage of successful task completion based on scoring system
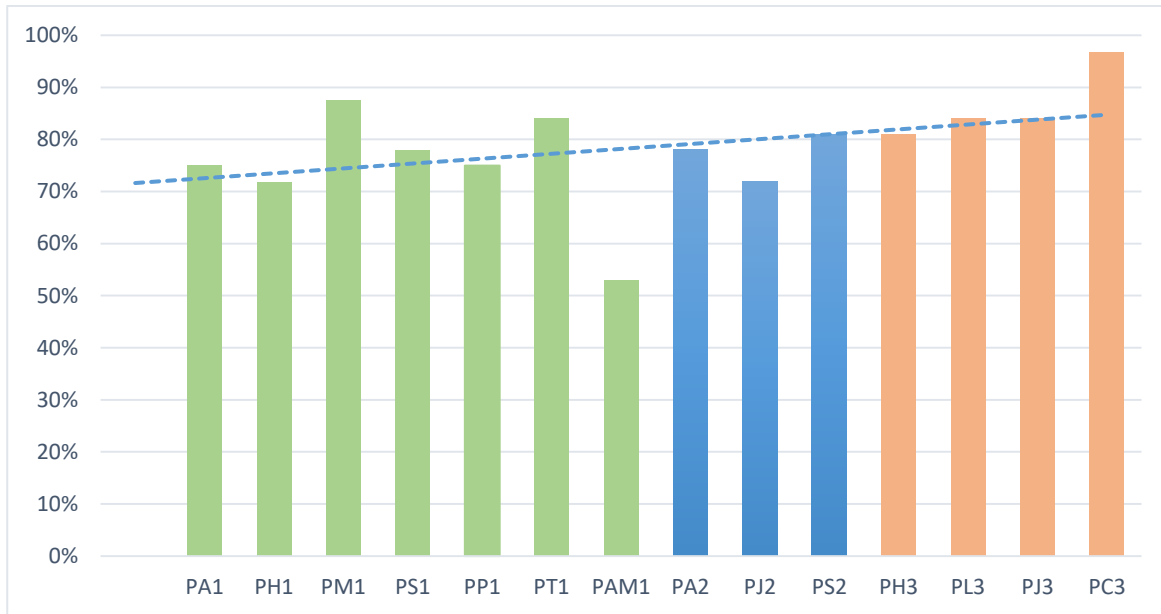
As shown in the graph above, in the first round of tests, participants managed to successfully complete around 70% of both tasks. Similar results were achieved in the second round of tests. In comparison, experts who took part in the third round of tests performed better than in the previous groups, with one participant achieving a 98% completion rate.

Completion Time

A mean completion time for the whole test group was estimated, using the sum of all completion times and the total number of participants. The results are illustrated in Table 26 below.

Table 26: Mean task completion time

	Task 1	*Task 2*
Group 1	32 min	4 min 01 sec
Group 2	11 min 14 sec	3min 26 sec
Group 3	28 min 15 sec	4 min 29 sec
Benchmark	33 min 42sec	5min

The first finding was that, in general, the completion time decreased in the case of task two, despite the task having a similar level of complexity. This was further supported by my observation during the experiment that users learned how to use the tool when completing the first task. This allowed them to improve their performance in the second part of the experiment.

A benchmark was set up during the trial with two participants. All groups performed better than the benchmark. Overall, group two performed better than group one in both tasks. Group three had the worst completion time, but as explained earlier, the longer completion time was due to a much greater amount of verbalisation and more thoughtful completion of the tasks. It was observed that professionals were trying to understand how the tool worked as they went along, rather than simply completing the task in ignorance.

8.4 Discussion

A usability evaluation study was conducted to examine usability problems with the MALT model interface. The evaluation consisted of three separate rounds, and after each the interface underwent improvements. Think Aloud protocols and additional data collected allowed identification of usability problems, and provided explanations of why these had occurred.

Twelve major usability problems were identified in the first round of tests. In the second round of tests, the number of problems was reduced to nine, and in the following round to three. The problems identified in the last round were less severe, and classified as irritant. This indicated that the interface had been improved to the extent that telehealth and NHS stakeholders could use it.

The principal finding of this research is that language is one of the most important aspects of the tool, and eliminating abbreviations and specialist jargon significantly reduced occurrence of these problems in further test rounds. As far as problems with the language appeared in the second round of tests, these were managed by the embedding of icons in key areas of the model to display pop-up windows with additional information. Participants also expressed the need for clearer descriptions of model functions and more informative labels, which were subsequently implemented after the second round.

Default values were also the source of problems. Participants did not find them intuitive to use, and their purpose was often questioned. They were a source of frustration, especially when users forgot to delete them. Two solutions were implemented to resolve this problem. Firstly, a lengthier introduction of the purpose and workings of the default values was provided in the interface. Secondly, a colour scheme was introduced to indicate the quality of the data that fed into these values. Despite these efforts, the problems related to the use of default values appeared even in the third round of tests.

One unanticipated discovery during usability testing was that, during the test, a few participants ignored the right-hand side of the screen. Changing the arrangement of columns and tables eliminated this problem.

One amendment proposed during the test was an alteration to the design of the page, to render it sequential. As a result of discussions with the project team and software developers, it was decided to maintain the current, linear design, as fundamental changes in the design were not easy to implement at that stage of the project. The issue, therefore, remained unresolved.

Analysis of quantitative data on time and errors revealed that the time taken to solve two tasks was lower in the second test group. These findings should be analysed with caution, as participants verbalising their thoughts and commenting on issues impacted their performance speeds. The longer time required for formulating queries could also have been a result of users testing the model's features, as they had never seen the interface before the test.

The frequency of errors served as in indicator of learnability, and I expected that the more user-friendly the software became, the fewer mistakes or errors participants would make. In general, the number of errors did not deviate between the test groups, and all of the participants were able to achieve a success rate of more than 80%. It could be argued, however, that the number of errors depended upon the person's experience and IT skills. These were not assessed by an independent test, but all participants were assumed to have at least basic IT skills, as they all claimed to be familiar with Excel.

Despite the usability issues, participants' comments suggested that the MALT model interface was well received. Generally, users appreciated the aesthetics of the tool and found it easy to use.

8.4.1 Strengths and weaknesses

The strength of this research is in its positive response to recruitment strategy, which enabled the completion of three usability test rounds. By recruiting 16 participants, I exceeded the number of participants usually enrolled in usability evaluation studies. The results of the review in Chapter 4 suggest that, with the exception of one study, the majority of published studies relied on recruitment of 5-15 participants. I also ensured that the data were collected from diverse samples: students, health economics professionals, and telehealth specialists.

To strengthen the quality of this research, a systematic review of the usability evaluation methods has been conducted (Chapter 4) to ensure that the limitations of the published literature were explored and avoided. The instructions for conducting usability evaluations were strictly followed, particularly those pertaining to prompting participants after precise time intervals, using the 'Keep talking' command (Bastien, 2010; Ericsson & Simon, 1984).

The strength of the study also comes from the design and analysis of data being conducted with usability specialists from the user-centred healthcare design group of South Yorkshire, CLAHRC.

The major limitation of this study is that it was conducted as laboratory usability testing, and as such, it ignores the environment and the context of real-life decision-making situations. In real life, MALT model users are expected to make an effort to collect the data and to translate decision-making problems into a set of numbers that can be entered into the model. This is not always a straightforward task. As observed by Brailsford *et al* (2011) during the Scenario Generator study, data pose a major barrier to simulation and modelling use in the NHS. Some of the specific issues mentioned in the study were problems with data access, quality, and required analysis skills. For example, even if the data were available, the format may not be compatible with what is required for the purposes of modelling. This means that further efforts would be required to reformat

available datasets. In addition, some of the data might exist in a paper format, and not electronic. This means that a skilled analyst's input is needed in this first crucial step, and not all organisations can afford this.

The unified theory of acceptance and use of technology (discussed in Chapter 3) proposes that effort expectancy can reduce the intention to continue working with a technology. In other words, modelling and simulation tools may be abandoned if data acquisition proves too problematic.

This study is also by limited by the absence of video-based eye-tracking software. Software of this type would be able to adequately identify the areas of the online spreadsheet which the participants are considering for data entry. However, the software could not be used due to constraints of time and finance. Analysis of the data from eye-tracking software is also arduous and requires extensive additional training. The limits of the project prevented the use of such software.

8.4.2 Comparison with other studies

The current study demonstrated the feasibility of the verbal protocol analysis method for testing the usability of an economic model. This is the first of such studies conducted in the health economics discipline. The study was conducted within the constraints of an ongoing research project, which imposed deadlines and the need to interact with the software developers. However, this situation is not thought to be unusual, and as such, the approach taken is considered realistic and generalisable.

The systematic review in Chapter 4 identified 14 published usability studies that had employed the Think Aloud technique. However, the guidelines as to how such study should be executed, and how the data should be analysed and reported, are ill-defined, and there is observable

divergence in practice in the published literature. The choice of method in the majority of the reviewed literature was based on practical considerations, such as financial resources and project timescales.

Unlike the majority of the studies, the current research followed an approach based on the methodological review. This review was conducted to better understand which usability evaluation method would be the most appropriate for the study. Additionally, the findings from the qualitative interviews were instrumental in developing a user profile and defining users' expectations regarding the proposed tool. Some of the published usability studies employed a questionnaire method for these purposes, a method cheaper and less time-consuming.

Moreover, in the current study an effort was made to recruit an appropriate number of participants and to involve a representative sample of end-users. With the exception of the study by Swaak (2009) and Wrubel (2007), the current study recruited more subjects than any other reviewed usability study. As detailed in Table 4 in Chapter 4, the number of participants in the published usability studies ranged from 5-15. The current study recruited 16 participants. The majority of these studies relied solely on student subjects. In the current study, an effort was made to involve both PhD students and modelling and telehealth professionals.

In addition, some of the studies collected verbal protocols by collecting audio data only, whereas the current study also collected video data. The video data was scrutinised to gather data on verbal protocols and on-screen movements, a method previously applied in Jaspers *et al* (2004). This is one advantage of this study, as the majority of the published usability evaluation studies that adopted VPA do not note the exact type of the technique used in the analysis of the data on verbal protocols, and a combination of coding, sorting, and scoring techniques were the most common.

8.4.3 Critical reflection on the value of the usability testing

Critical reflection requires a description of the situation and an attempt to find meaning using ideas or theories (Fook & Gardner, 2007). Throughout this section, I reflect upon the value of the MALT model usability evaluation, outlining the perspective of other researchers from the wider MALT project team who I consulted on the findings. The reflection process was guided by the list of questions from Fook and Gardner (2007, p. 170), and is written in the first person to provide an exploratory perspective.

The development of a decision tool for policy-makers, commissioners, providers, and the telehealth industry to support evaluations of telehealth interventions was one of the main goals of the MALT project (as discussed in Chapter 2). The results of the usability testing informed further developments of the model's interface.

I consulted a senior researcher from the MALT team. He stated that usability testing was of the highest importance for ensuring that a model was tested prior to its official launch. He noted that it was important to understand its flaws and eradicate them during the final six months of the MALT project. The anecdotal evidence collected during the MALT study events and stakeholders' meetings highlighted the expectation that the model would be 'flawless', and it was suggested that the model would be used by future projects to support scaling up, adoption, spread, and implementation of telehealth.

Another senior researcher stated that the usability testing was an opportunity to further develop and refine the final form and content of the model. Whilst time was a critical factor, and the model's launch could have not been compromised, usability testing prevented the team from dealing with some of the issues after the model's launch. He felt that changing the political

environment and the technology advances had created a momentum for the model to gain the attention of policy-makers, and that the team was focused on fully exploiting that opportunity.

Another valuable insight provided by the MALT research team was a comment on the significance of the usability testing having a pragmatic focus. Whilst the modellers' team was busy collecting data from validation exercises, I was able to observe end-users performing simple tasks, such as logging in, data entering, and analysis, which enriched our understanding of how the model would perform in 'real life'.

Furthermore, the usability testing was seen as 'way of addressing user issues' that needed to be considered to optimise performance and future uptake of the tool. The development of the model and its successful dissemination was seen as supporting the mainstreaming of telehealth technologies in the Yorkshire and the Humber region. Locally, the initiative was joined by a number of research organisations, including the Collaboration for Leadership on Applied Research and Health Care for Yorkshire and Humber (CLAHRC Y&H) and the Academic Health Sciences Network's Telehealth Board. Through collaboration with these research institutions, it was agreed that the MALT model and other toolkits generated through the MALT project were to be incorporated into a comprehensive online training and awareness resource platform for industry, policy-makers, and NHS staff. The plan was that these tools would exist in the wider context of web-based telehealth resources. Therefore, the usability testing gave the MALT model the credentials, ensuring that the tool was sufficiently developed for end-users.

CLAHRC Y&H agreed to adopt the model following the end of the MALT project, and it was planned that CLAHRC would continue to provide hosting and implementation support for this tool. The plan was to ensure the continuing impact from the MALT project, even after the funding period had finished. The fact that the model had undergone usability testing was mentioned when

the project management team initiated the discussions with NHS England, who expressed an interest in adopting the tool.

The final MALT dissemination event was held on 30 September 2014. During the event, which was attended by a varied audience, including academic, policy, practice, and industry, the MALT model was officially launched. Participants were encouraged to explore the abilities of the tool and were asked to complete a short evaluation questionnaire. The people I spoke to praised the model's visual attractiveness, and its layout and format. Further comments from participants were all positive and encouraging.

8.4.4 Further research

To my knowledge, this study is the first to adopt usability evaluation methods during the development of a decision model. Whilst I was able to identify and apply the most appropriate methods possible within the research project, these were limited to testing in a 'controlled' laboratory environment. Further studies on the use of economic models in real life decision-making need to be conducted, as a model that is usable in the lab may not actually be so in practice, due to environmental factors, such as a lack of time.

8.4.5 Conclusions

This chapter presents the results of a usability study conducted to understand how expert and non-expert users perceived the interface of an economic model. The study included the evaluation of the MALT decision analytical model using Verbal Protocol Analysis method. Twelve experienced postgraduate students (NHS experience, or health economics) and four telehealth professionals took part in the experiment. To assess the usability of the tool, I measured learnability and efficiency. The main findings were that, after changes had been implemented to

the design of the interface, the completion time improved. Fewer usability problems were detected in a subsequent version of the model.

9. User satisfaction evaluation

This chapter presents the methods and findings from the third case study, which aimed to assess user satisfaction with the proposed economic model. The result section presents a mixture of findings, from the questionnaire face validity testing, the main questionnaire study, and additional qualitative interviews conducted to explore user satisfaction. Following on from that, strengths and weaknesses of the study are discussed.

9.1 Introduction

User satisfaction is now regarded as an important measure of software and IT systems effectiveness (Ives *et al*., 1983). Ives and colleagues (1983) define 'user satisfaction' as the degree to which users believe that the information system at their disposal fulfils their needs. The concept has been widely debated since its conception. One point in this debate is the extent to which user satisfaction leads to greater use. Goodhue (1995) opposed this idea, suggesting that it has not been consistently demonstrated in previous research, particularly in environments where use is mandatory. Previous studies have also failed to explain why in some situations higher performance still results in user dissatisfaction.

As a result of recognising the importance of user satisfaction as a measure in information systems research, considerable study has been devoted to developing a standard user satisfaction instrument. The first of such, by Bailey and Person (1983), included 39 items. The respondents were asked to choose the word which best described each item on a two-dimensional scale. For example, 'accuracy', understood as the 'correctness of the output information', was evaluated by choosing one from each pair of statements: accurate vs. inaccurate, high vs. low, consistent vs. inconsistent, sufficient vs. insufficient. In 1983, Ives developed a short, 13-question tool based on their questionnaire (Ives *et al*., 1983).

In 1988, Doll and Torkzadeh (1988) developed an instrument for measuring the satisfaction of end-users. They argued that advances in computer technology required a focus on end-users: the people who have direct contact with the application in use. They regarded previously published instruments as inadequate, as they were developed to measure satisfaction with information system functions, rather than satisfaction with a single application.

As noted in the review in Chapter 4, several studies have used such instruments to assess the usability or effectiveness of IT systems and different types of software. Typically, the questionnaires are administered post-task and aim to identify subjects' views and opinions of the system. The review suggests that, among the most often adopted instruments, are EUCS (Doll & Torkzadeh, 1988), discussed further in Section 9.2.1, and QUIS (Harper, 1993).

The QUIS questionnaire has been developed by researchers from the University of Maryland to assess subjective user satisfaction. Over the years, it has been updated to version 7 (Harper, 1993), and it currently consist of six scales that rate the system, four measures of specific interface factors, and a demographic questionnaire. A major criticism of the tool, and the basis of my decision to reject it, is that it is lengthy, which may discourage model users from completing it.

9.1.1 Aims and objectives

The aim of this case study is to assess user satisfaction via the proposed decision analytical model. The study was open to all people who had registered to use the model on the project website. Users received satisfaction questionnaires one month and three months after registering on the site. The collection of satisfaction data was undertaken at two different time points to assess if their retrospective satisfaction had changed. It was assumed that measuring satisfaction in a one-time evaluation might produce misleading results, as the use of the tool requires some learning, and therefore perceived ease-of-use might change over time.

316

9.2 Methods

9.2.1 EUCS instrument

Considering the overall aims of the evaluation, I made the decision to apply the EUCS questionnaire. The choice of EUCS over other instruments was dictated by the fact that I intended to make comparisons between the main groups of model users (the members of the telehealth industry, the NHS staff and commissioners, and telehealth providers). The advantage of EUCS came from its high reliability and internal consistency (Doll *et al.*, 2004). It has been extensively validated and used in evaluations of various IT technologies, including websites (Abdinnour-Helm *et al.*, 2005), hospital information systems (Aggelidis & Chatzoglou, 2012) and decision support systems (McHaney *et al.*, 1999).

The end-user computer satisfaction instrument measures user satisfaction in five domains: content, accuracy, format, timeliness, and ease of use. The users are asked to answer 12 questions, using a five-point Likert scale, with answers ranging from 'almost never' to 'almost always'.

The content category focuses on the usefulness of the results provided by the tool. The accuracy and format categories question the way the outputs are presented. The ease of use focuses on the user-friendliness of the tool. The last variable, timeliness, verifies the speed and the quality of the provided information. The instrument in its original form is presented in Table 27 below.

Table 27: EUCS instrument (Doll & Torkzadeh, 1988)

Content	C1: Does the system provide the precise information you need?
	C2: Does the information content meet your needs?
	C3: Does the system provide reports that seem to be just about what you need?
	C4: Does the system provide sufficient information?
Accuracy	A1: Is the system accurate?
	A2: Are you satisfied with the accuracy of the system?
Format	F1: Do you think the output is presented in a useful format?
	F2: Is the information clear?
Ease of use	E1: Is the system user friendly?
	E2: Is the system easy to use?
Timeliness	T1: Do you get the information you need in time?
	T2: Does the system provide up-to-date information?

The benefit of EUCS is that it easily identifies the areas of user dissatisfaction. In addition, the instrument enables the comparison of user satisfaction with prolonged use and between user groups by comparing an overall EUCS score. The score is derived by summarising the scores

from each category. The higher the score, the higher the end-user satisfaction, with the minimum being 16 and the maximum 60.

9.2.2 Face validity testing of the EUCS questionnaire

Face validity of a questionnaire refers to the conformity of meaning across the sample (Klenke 1992). Face validity of an instrument is best assessed by asking the respondents to confirm their understanding of each item (Nevo, 1985). As such, face validity is regarded as the weakest of validity measures, in comparison to content or construct validity. However, in order to conduct these types of tests, large samples of users are required to conduct confirmatory factor analysis.

The methodological choice for this study was determined in consultation with a specialist in psychometrics (Dr Georgina Jones).

Face validity of the EUCS questionnaire was tested on four users. They were asked how they understood each question and what their likely answer would be. The users were audio recorded and transcripts from these interviews developed for further analysis.

9.2.3 Demographic and satisfaction questionnaire

A demographic questionnaire was developed to collect demographic data on model users. The users were asked to fill this in when they first registered on the website. Completing this questionnaire was obligatory. The questions pertained to users' profession, their experience of using economic models, and the software packages most often used. The list of questions was kept short at the request of the MALT project team, so as not to discourage users from completing registration. A full list of the questions is given in Figure 37.

Figure 37: Demographic questionnaire

Q1: How do you describe yourself?

Academic/ Member of industry/ NHS employee/ Analyst/ Other

Q2: How do you characterise your current professional background?

Medical/finance or economic/ management or related/other

Q3: Have you undertaken any financial or economic modelling before?

If yes Q4: how many years of experience in modelling do you have?

Moreover, users received an additional satisfaction questionnaire that consisted of six open-ended questions (presented in Figure 38). This short questionnaire was sent alongside EUCS one month and three months after registering.

Figure 38: Satisfaction questions

Q1: What was the purpose you used the model for?

Q2: Did you generate any results from the model?

Q3: How did you use the results?

Q4: How useful has the model been to you?

Q5: What changes would most improve the model?

Q6: Do you have any other comment about the model?

8.4.3 Critical reflection on the value of the usability testing

Critical reflection requires a description of the situation and an attempt to find meaning using ideas or theories (Fook & Gardner, 2007). Throughout this section, I reflect upon the value of the MALT model usability evaluation, outlining the perspective of other researchers from the wider MALT project team who I consulted on the findings. The reflection process was guided by the list of questions from Fook and Gardner (2007, p. 170), and is written in the first person to provide an exploratory perspective.

The development of a decision tool for policy-makers, commissioners, providers, and the telehealth industry to support evaluations of telehealth interventions was one of the main goals of the MALT project (as discussed in Chapter 2). The results of the usability testing informed further developments of the model's interface.

I consulted a senior researcher from the MALT team. He stated that usability testing was of the highest importance for ensuring that a model was tested prior to its official launch. He noted that it was important to understand its flaws and eradicate them during the final six months of the MALT project. The anecdotal evidence collected during the MALT study events and stakeholders' meetings highlighted the expectation that the model would be 'flawless', and it was suggested that the model would be used by future projects to support scaling up, adoption, spread, and implementation of telehealth.

Another senior researcher stated that the usability testing was an opportunity to further develop and refine the final form and content of the model. Whilst time was a critical factor, and the model's launch could have not been compromised, usability testing prevented the team from dealing with some of the issues after the model's launch. He felt that changing the political environment and the technology advances had created a momentum for the model to gain the attention of policy-makers, and that the team was focused on fully exploiting that opportunity.

Another valuable insight provided by the MALT research team was a comment on the significance of the usability testing having a pragmatic focus. Whilst the modellers' team was busy collecting data from validation exercises, I was able to observe end-users performing simple tasks, such as logging in, data entering, and analysis, which enriched our understanding of how the model would perform in 'real life'.

Furthermore, the usability testing was seen as 'way of addressing user issues' that needed to be considered to optimise performance and future uptake of the tool. The development of the model and its successful dissemination was seen as supporting the mainstreaming of telehealth technologies in the Yorkshire and the Humber region. Locally, the initiative was joined by a number of research organisations, including the Collaboration for Leadership on Applied

Research and Health Care for Yorkshire and Humber (CLAHRC Y&H) and the Academic Health Sciences Network's Telehealth Board. Through collaboration with these research institutions, it was agreed that the MALT model and other toolkits generated through the MALT project were to be incorporated into a comprehensive online training and awareness resource platform for industry, policy-makers, and NHS staff. The plan was that these tools would exist in the wider context of web-based telehealth resources. Therefore, the usability testing gave the MALT model the credentials, ensuring that the tool was sufficiently developed for end-users.

CLAHRC Y&H agreed to adopt the model following the end of the MALT project, and it was planned that CLAHRC would continue to provide hosting and implementation support for this tool. The plan was to ensure the continuing impact from the MALT project, even after the funding period had finished. The fact that the model had undergone usability testing was mentioned when the project management team initiated the discussions with NHS England, who expressed an interest in adopting the tool.

The final MALT dissemination event was held on 30 September 2014. During the event, which was attended by a varied audience, including academic, policy, practice, and industry, the MALT model was officially launched. Participants were encouraged to explore the abilities of the tool and were asked to complete a short evaluation questionnaire. The people I spoke to praised the model's visual attractiveness, and its layout and format. Further comments from participants were all positive and encouraging.

9.2.4 Analysis of data from logs

The study collected data on users' activity on the model website, such as how many times the tool had been accessed and what type of model runs were executed. I intended to explore if the model were adopted for routine use or only accessed for one-off exploration of its capabilities, and also

what were the types of analyses most often used. These data would be further explored using statistical tests.

9.2.5 Ethical approval

The study received ethical approval from the School of Health and Related Research Ethical Committee (see Appendix A.9.1 for a copy of a certificate). Several measures were incorporated into this research to ensure that it was conducted in line with the guidelines for ethical conduct.

Firstly, model users were required to give informed consent to participate in the study. This was achieved by creating an online consent form. The users were informed that their participation was voluntary and that they were free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, they were informed that they were free to decline a question if they did not wish to answer it. Consent was given by checking a box.

Secondly, the confidentiality of respondents was protected by removing identifiable information and personal details from the collected data. Each participant was identified by a code, which was then used throughout the research. Confidential data was stored in a secure locked cabinet at the University of Sheffield.

9.3 Results

9.3.1 Sample

The sample for this part of the case study consisted of all the users who had registered to use the model on the project website between December 2014 and May 2015, and who had expressed interest in taking part in the study. In that period, 43 users registered. The structure of the sample is illustrated in the Table 28 below.

Table 28: MALT model registered users between December 2014 and May 2015

NHS	Academics	Telehealth Industry	Local Authorities	Consultancy companies	Government agencies
11	9	4	4	8	7

9.3.2 Face validity testing of the EUCS instrument

The qualitative data generated during face validity testing were analysed using simple content analysis. A full content analysis was considered neither necessary nor possible for data on only four participants. The transcripts were broken down into sections corresponding to each of the items. Table 29 below presents the participants' main comments on each of the EUCS questionnaire items.

Table 29: EUCS instrument face validity testing

C1: Does the system provide the precise information you need?	P1: '*You need to specify if it refers to the results of the model or particular sections of the model*' P2: '*It is clear to me*' P3: '*It is fine*' P4: '*I would understand this as referring to the accuracy of the final results*'

C2: Does the information content meet your needs?	P1: '*I don't understand that question*' P2: '*It seems to be a duplicate of item C1*' P3: '*I think I would struggle a bit with this one, it is not clear enough*' P4: '*I would say it depends on your expectations of the tool. The question is bit vague*'
C3: Does the system provide reports that seem to be just about what you need? C4: Does the system provide sufficient information?	P1: '*It is fine, I understand that*' P2: '*I understand what you mean*' P3: '*I understand that you are asking here about the final results*' P4: '*Clear*'
A1: Is the system accurate? A2: Are you satisfied with the accuracy of the system?	P1: '*How [do] we define accuracy of the model? I think all these questions will be difficult to answer*' P2: '*It looks like you are asking the same thing again*' P3: '*It is a bit vague; do we refer here to the validity of results?*' P4: '*Is there a way to assess the accuracy of the results?*'
F1: Do you think the output is	P1: '*Fine*' P2: '*Clear*'

presented in a useful format? F2: Is the information clear?	P3: '*I understand this'* P4: '*I understand what you mean'* P1: '*I would refer here to the style of outputs'* P2*: 'Fine'* P3*: 'I guess you are asking about transparency of results'* P4: '*I understand this item*'
E1: Is the system user friendly? E2: Is the system easy to use?	P1: '*What is the difference between user-friendliness and ease of use? Aren't they the same?* P2: '*I can guess what both mean'* P3: '*I would probably answer the same thing for both of these. Why you need the distinction?'* P4: '*I know what you mean'*
T1:Do you get the information you need in time? T2: Does the system provide up-to-date information?	P1: '*Does it refer to the speed which the answers appear on the screen. For example, if I press 'i button' would that appear quickly enough? Or does it refer to the speed of calculations? It is not very clear to me'* P2: '*How would I know? The data is on the server; I do not have access to it'* P3: '*It is beyond my control'* P4: '*I wouldn't know how to answer that question'*

Changes were made to the questionnaire after the analysis of the face-validity testing. These are as follows:

❖ Item C1 and C2 need to state more clearly if they are applied to the final results or to the results of the scenario analyses. Item C3 was removed, although the majority of participants did not have any concerns about it, as the study did not intend to investigate understanding of the reports. As such it was regarded as not relevant.

❖ In a case of item A2, the respondents needed to know if they were being asked about the accuracy of the predictions of the final model run, or that of the data displayed in the graphs or tables.

❖ Similarly, item F2 needed to specify which 'information' we were referring to.

❖ Both T1 and T2 questions in the 'Timeliness' section were considered problematic. Although I felt that the wording of these two items was not relevant to this study, as we did not expect the users to maintain and update the data that feed into the model, I decided to keep them as a study on validity of EUCS. (McHaney *et al*., 1999) regarded 'Timeliness' as relevant in all types of IT applications.

❖ In addition, the five-point Likert scale was found confusing for some items. For example, interviewees pointed to the fact that in case of a C2 item, the logical answer was 'yes' or 'no'. This could not be changed because interfering with the scale would destroy the overall scale. It was necessary to pay greater attention to the answers in this category.

9.3.3 The results of the demographic questionnaire

As illustrated in Section 9.3.1, the model drew the attention of members of the telehealth industry, academics, NHS staff, local authority personnel, and consultancy companies. An analysis of the demographic questionnaire, collected from 43 users, revealed the data on backgrounds illustrated in Table 30.

Table 30: The background of the MALT model users (n=43)

Professional background of the respondents	Percentage of the sample
Management or related	46.5%
Finance or economics	27.9%
Medical	14%
Public administration	4.7 %
Research	2.3 %
Public services	2.3%
Policy development	2.3%

The main groups of professionals who registered to use the model were managers, analysts, medics, and public servants. All of these categories were concordant with my earlier predictions (Chapter 8, Table 17), in which I predicted that the majority of users would hold managerial or operational positions.

Furthermore, 51% of the sample stated they had worked with the models before. Their modelling experience ranged from 1-10 years, as illustrated in Table 31 below. When asked about how often they used models in their current work, of those who claimed to use such tools, only 18% said they used them most of the time.

Table 31: Economic modelling experience of the registered MALT model users

Years of modelling experience	1-2 years	3-7 years	Over 10 years
Percentage of the respondents with previous modelling experience	35%	35%	30%

9.3.4 Results of the data analysis of the logs

The data on the MALT model collected from logs for the period of December 2014 to May 2015 revealed that 43 people had registered on the website to use the tool. Of these, 34 had registered on the site but not run any analyses. An additional qualitative investigation was undertaken to reveal the reasons for this.

In the observed period of time, nine people actively used the model. An examination of the 'active' users revealed that they were from academia, NHS trusts, consultancy companies, and the telehealth industry. Their years of modelling experience varied from one year to 15 years. There were no people without experience in economic modelling.

As shown in Table 32 below, the majority of 'active' users ran the model once and used the light version of the tool. They were mainly from academia and the NHS. This might suggest that the level of detail required was sufficient for light and medium versions of the tool.

Three users used the full version of the tool and ran the model more than once. They were from consultancy and telehealth companies. None of the users ran the PSA, which may suggest that this option was too advanced for them.

Table 32. The active users' professional background and modelling experience

Number of model runs	Number of active users		
	Light	Medium	Full
One	4	2	0
Two	0	0	3

9.3.5 EUCS questionnaire results

The response rate for the questionnaire was very low. Only two people responded to the EUCS questionnaire sent by email one month and three months post-registering. This data is shown in full in Appendix A.9.2.

Additional qualitative interviews were conducted to explore why users failed to use the model, and of those who did use the model, what were their views on its usefulness for the process of decision-making and their overall satisfaction. This is described in section 9.5.

9.4 Discussion

The first observation on the results obtained from the demographic questionnaire is that the majority of the sample constituted different level managerial staff, from NHS services, local authorities, telehealth companies, government agencies, consultancy companies, and academia.

The number of registered users was low, and anecdotal evidence suggested that an evaluation tool was urgently needed, considering the existing policy push to implement telehealth in the majority of locations in the country. The research team also took a number of approaches to raise awareness of the model and encourage registration. These included conference presentations during the development process, a plenary presentation of the model at an end-of-study event, plus a workshop at the same event, and the development of an attractive website, which was advertised in the study newsletter and on various social media platforms (e.g. Twitter, blogs, LinkedIn).

The majority of the sample had managerial or operational positions. The second largest group of users consisted of people with backgrounds in finance or economy. The model also appeared to draw the attention of clinicians.

The level of economic modelling expertise varied. The results of the questionnaire revealed that 49% of respondents did not have any modelling experience. They were therefore non-experts. Eleven people claimed to use modelling some of the time, and only five stated that they used models as a primary tool in their work. This interesting insight, that the majority of people who registered to use the model were not modellers, can explain to some extent why many registered users did not proceed to run the model.

By any standard, the data generated from analysis of the logs was unsatisfactory. I expected to be able to analyse feedback generated by a larger and more diverse sample. A further discovery was that only three people had attempted to use the full version of the model. Further investigation was needed to explore this issue.

The low response rate of the EUCS questionnaire was unanticipated. It is suspected that the reason for this was that users were being asked to complete the questionnaire one month and then three months after registering, potentially having lost interest by this point, after their initial attempts to use the model.

One could argue that the choice of instrument was to blame, and that a simpler questionnaire should have been chosen. The advantage of employing EUCS in this study was that it produces results which could be compared across samples in order to investigate which aspects of the interface had the greatest impact on user satisfaction, and how they varied across these samples. In addition, one-month and three-month responses could have identified how EUCS scores were affected by longer use. A simpler, bespoke questionnaire that could be completed after a single use would probably elicit a higher response rate, but the aim of the research was to use a validated instrument that fitted with the existing literature on theory and applied studies.

In conclusion, with such a low response rate, an assessment of the EUCS and user satisfaction was not possible. However, the log data suggested that there were considerable barriers to use in terms of usability, relevance, motivation, and time. Further study was therefore undertaken to investigate this.

9.5 Additional qualitative interviews

The poor response rate led to the decision to undertake additional qualitative interviews with the registered users of the model. The main aims of the interviews were to answer the following questions:

- What was the users' subjective satisfaction?

- What is the usefulness of the model in real decision-making?

- Why was the tool not used in some cases?

9.5.1 Methods

Sampling strategy

A purposive sampling strategy was used for the research. Interviewees were recruited from those who had registered to use the model on the project website. The recruitment strategy was heavily impacted by ethical approval constraints, as it allowed only non-NHS people to be involved in the research, as well as by the time constraints of the PhD. Taking these considerations into account, the target sample size was set at nine users. The process is illustrated in Figure 39 below.

Figure 39: Model feedback interview sampling strategy
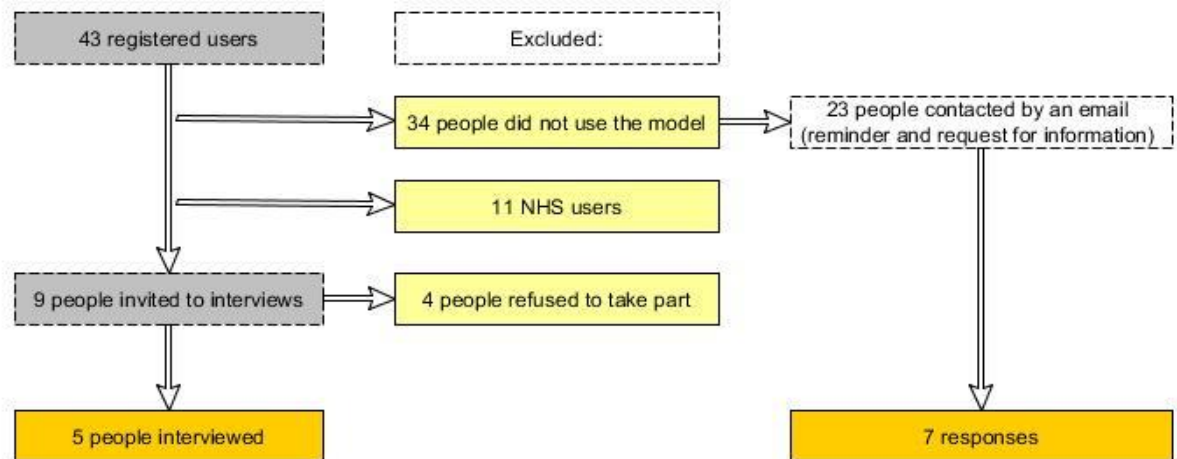
Of the initial sample of 43 registered users, 34 people were excluded because they had registered on the site but failed to use the tool. Of those 34, 11 were excluded because they were members of the NHS. The remaining nine people, who had used the model actively, were approached by me and received an email invitation to take part in the study. Email addresses were the only contact details provided during registration, and were therefore the only method of approaching potential participants. Of these nine, one person refused to take part because he felt that he could not share any constructive feedback because it had been four months since he had used the tool. Three people refused to take part because of other commitments. This left five participants for the study.

Of the 34 excluded, 23 people were sent a reminder about the tool and a request for information as to why they had not used the tool. Seven responded with feedback. The final sample is presented in Table 33 below.

Table 33: Final sample of the model feedback interviews (n=5)

	Consultancy companies	**Telehealth companies**	**Academic**	**Local Authority**
Job titles	Consultant	Programme Manager	Programme Manager	Service Manager Analyst
Modelling experience	Expert	Expert	Expert	No previous experience
Number of participants	1	1	1	2

Topic guide

The topic guide included four basic questions pertaining to experience of using the model. Firstly, I intended to explore how each person had found out about the model, and what had motivated them to use it. Additionally, each person was asked their opinion of the model's features and the options it provided. Lastly, they were asked if they had found the model useful, and if they were satisfied with the tool. A copy of the interview topic guide is presented in Appendix A.9.3.

Data collection

Telephone interviews were conducted between June and September 2015. The interviews were recorded and transcribed verbatim. More than 3.5 hours of audio was collected in the process. The qualitative data was then analysed using Nvivo 10 (QRS International., 2012) software.

9.5.2 Ethical approval

The work received approval from the School of Health and Related Research Ethical Committee. The approval was granted on condition that the research would not involve NHS staff, as their involvement would require separate NHS research governance approvals from each participating organisation. A copy of the certificate is available in Appendix A.9.4.

9.5.3 Findings from the interviews (n=5)

How users found out about the tool

The first qualitative finding revealed how users found out about the tool. The majority came to hear of the tool from colleagues who had participated in MALT events, or through information about the model on social media. In the process of promoting the model, various strategies were used. For example, the model was heavily promoted during MALT study events and conferences, telehealth related events, on social media, on the project website, and by NHS speciality groups. The interviews revealed that some of these strategies were more or less effective than others.

Examples of application of the model in practice

The interviewees revealed that the model had been applied in various decision-making situations. In one instance, it was used in business case development, where it served as a benchmark for developing a business case. In another, the tool was used to evaluate a currently run telehealth project that had undergone major changes. It was also used to support contracting in one of the organisations. The tool was routinely used in one organisation, where ongoing discussions with commissioners and providers were taking place.

One of the participants stated that he had used the model during ongoing discussions with NHS commissioners. In his opinion, having such a tool available allowed him to provide a useful argument against the option favoured by his partners. He particularly appreciated the default values that allowed him to run his scenario very quickly, and to demonstrate to his partners that the option was not worth considering. Another person used the example of negotiations he had carried out with a PCT, and how he had used the tool on his tablet to answer some of the queries. The following quotes refer to the situation:

> *I have taken it to a meeting and run it on my tablet when a certain thing was mentioned. Immediately it got the attention of everybody present in the room. For the rest of the meeting we looked at a number of things and it really helped the conversation. (Consultant)*

The quote implies that the tool was used as some sort of novelty, and the fact that it could provide the results in a relatively short period of time enabled it to be used during a particular meeting. The interviewee also provided examples of the analyses considered during that meeting, such as considering the dynamics of the patient population for the region of interest, the overall

338

expenditure on care for COPD patients, and implementation costs in case of a five-day telehealth system.

Another participant made an interesting point, suggesting that the tool helped him and his team to change the perspective on data analysis. They had been seeking to demonstrate a reduction in hospitalisations and admissions, but after using the tool for a short period of time they realised that demonstrating the changes for a whole cohort of patients was more appropriate.

One of the participants went even further, calling the tool 'a benchmark' and admitting that it was used to learn more about the requirements of commissioners. He said:

> *We use it primarily as a benchmarking tool. It provides intelligence from our point of view on what local commissioners would expect, and we can take it to contractual meetings with the NHS Commissioners to say 'can we look at doing this'. (Local Authority Service Manager)*

Users' opinions of the model's features

The evaluation results indicated that all participants had positive attitudes towards the model's features. However, when explaining the reasons for the limited number of model runs, two participants pointed to system-wide barriers. For example, the manager from the local authority suggested that in her organisation there was nobody who could acquire data from a variety of sources, a process she perceived as very laborious and time-consuming. The lack of data was a source of frustration and led to the tool being abandoned. When I pointed out that the model could have been run using default values, the informant identified further constraints, suggesting that local data varied significantly from the national average. She stated:

So, I registered to have a look and see what kind of things were needed to use the tool and realised that you needed quite a lot of data, demographic data and other things to put in. So I had to go back to partners and sort of say well you need to get a load of data, and at the moment there's nobody to do that. I could use the national average, but if you're preparing the evaluation of your service it is quite important to have your local data. (Local Authority Service Manager)

Another person pointed to the fact that not knowing in advance the level of detail required to populate the model had left her frustrated, especially as there was no option to save the work in progress and come back to it later.

Particularly, I think when you just have been trying to use something for the first time and you are not really familiar with what kind of approach has been taken, quite useful would be that you can have a go at something, or come back, and you can go through all the pages and then come backwards. And so, you can move backwards and forwards without having to fill everything in, and that you can save it as it is even if it is an incomplete one. (Programme Manager)

Users' overall satisfaction

All of the interviewees expressed satisfaction with their experience of the model. When asked to rate their experience, four stated that they were satisfied, and one very satisfied.

Usefulness of the model

During the interviews, participants expressed views on the usefulness of the model. All of the comments implied that the users appreciated the tool and perceived it as useful. For example, one person stated that, based on his experience of the model, he would continue to use it.

> *It has quite useful functionality, and I think personally that the tool has a lot of potential. It definitely helped us in the project and we will continue to use it. (Programme Manager Telehealth company)*

And another:

> *It was very useful from my perspective. It gives you the opportunity to look a little bit more widely. We get very tied up with looking at specialties, but we tend then not to understand what the patients are presenting with, and the tool does allow us to focus on that a bit more and saying, well let's not worry about non-elective thresholds or length of stay or readmissions but actually, what are we doing with these patients, and could we build cohorts patients here that we could look at doing something differently. (Programme Manager Local Authority)*

Other comments

During the interviews, several informants shared their ideas for improving the tool. As mentioned earlier, one mentioned that she had felt frustrated by not being able to populate the tool with local data for her service, which resulted in the tool being abandoned. She was eager to use it, but data acquisition was beyond her control. As previously mentioned, one person explained that an option

to save the work in progress would be particularly useful for people who have to follow a lengthy process to obtain the data.

Another person, who claimed that she was not very skilled with computers, suggested that creating a video tutorial to explain which data were required for the tool would help. She described the process of collecting information and how laborious it was. The activities included database searches and following up on previous projects and reports. Even with all these efforts, she found some of the data unsuitable for populating the model. The following was stated during the interview:

> *Maybe it is just me, I don't know, I have to say I'm a bit slow when it comes to new technology, but in order to get that data I had to search in our database, I followed refs in older projects, and I think getting everything in order took longer than it should. If I could watch something like a video tutorial before I started at least I would know what to look for because some of it was not good at all. (Programme Manager University)*

Describing his experience with the tool, one person suggested that including another long-term condition would be beneficial for the tool's usefulness.

> *Of course that brings in sort of slight frustration when you think, you know, could I look at it for this malignancy? Could I look at this condition? Can I look at diabetes? As with all tools, as soon as you reach the limit, you want a bit more in there. (Consultant)*

This implies that the person was satisfied with the tool: he explored all its options and considered it useful for replicating the approach to other long-term conditions.

Two further examples of improvements were given when we discussed how telehealth had evolved over the past couple of years and the types of devices available on the market. One respondent felt very strongly that stand-alone devices would soon be replaced by standard telephones or mobile phones. He suggested that this would not only reduce costs, but also shorten the procedures when a new patient was enrolled into a telehealth programme. The interviewee was pleased to hear that he could evaluate such a scenario in the current version of MALT model, simply by having the costs of the equipment set as zero, and raised the issue of the implications of market growth and the impact for private insurers. He stated:

> *Private insurance schemes are something you did not consider here, and from my perspective this is the way the market is going. (Consultant)*

Again, I explained that if we assumed that an insurance policy was simply modelled as a fixed monthly payment to an insurer, such calculations could be performed with the current model.

Additional feedback from users who had not used the tool actively (n=7)

Those who responded to the request for feedback provided a valuable insight into why the model had not been used more actively. The most often mentioned reason was problems with obtaining the data. Four people stated that the gathering of the data was so laborious that they had not proceeded with the analyses. The second most common reason was lack of time, with two people stating they were interested in using the model but their jobs had required other tasks to be completed urgently, and as a result they had abandoned the model. One person stated that he had no intention of using the tool at the time of registering on the site, and he had sought to explore the tool's options because someone had mentioned it to him. He stated that he was willing to use it when the opportunity arose, but at this time telehealth was not considered in his organisation.

9.6 Discussion and critical reflection on the findings

The evaluation was designed to use a variety of data to explore different aspects of the MALT model's use and user satisfaction. As discussed in Section 9.4, one of the first unexpected discoveries was that the model had not gained as much attention as expected. More importantly, the data releveled that there were a number of people who had registered on the website but not run the model, and this aspect needed further investigation. Surprisingly, some of the model's features, such as PSA, had not been used at all.

On reflection, I consider the number of registered users to be satisfactory. I consulted a senior researcher in the field to discuss my findings, and he suggested that it was important to realise that the models for end-users are emerging in a limited number of domains, e.g. population health, public health interventions. The majority of decision models are used to generate results and do not have to be accessed by end-users. This echoed the findings of previously mentioned literature (Chapter 1), especially the work of Franco (2011; 2010). That made it difficult to judge whether 43 registered users was a low or high number. The number of NHS users may have been impacted by the fact that, at the time, the number of NHS trusts with ongoing telehealth projects was relatively sparse.

In an attempt to understand why people did not use the model more actively, I considered the barriers that had been highlighted in the review in Chapter 4A, namely data constraints, user skills bases, and barriers within organisations. These studies have shown that one of the major barriers to modelling use is access to and quality of data, and this problem is especially pertinent in the NHS (Brailsford *et al*., 2011; Eldabi *et al*., 2002). Brailsford and colleagues (2011) discussed this issue, based on their experiences of the project to embed the Scenario Generator in NHS trusts. They provided examples of how access to data might impose problems; for example, data may

exist in the system, but in a form unsuitable for modelling. This would mean that additional efforts were required to clean and analyse the data before it could be used in modelling. Similarly, Harper and Pitt 2004 discussed the challenges of using service user record data, and identified the messiness of the data sets and errors as the greatest challenges. Thus, a skilled analyst is required in order to make use of modelling, and not all organisations can afford this. The following observations echo the findings of my qualitative interviews (Chapter 6), where some stakeholders raised questions about the reliability of the existing data for supporting evaluations of the telehealth.

Furthermore, previous literature attests that the training and education in economic evaluation methodology leads to enhanced understanding and use of economic evaluation (van Velden *et al*., 2005). On reflection, it may have been useful to understand more clearly peoples' skills, both in computers and economics. Importantly, as suggested by the unified theory of acceptance and use of technology (Venkatesh *et al*., 2003), discussed in Chapter 3, user skills play an important role in the willingness to use technology. The decision to use may be also impacted by effort expectancy or the influence of others.

It is worth considering that there are various determinants of innovation adoption in an organisational context. One such, suggested by the diffusion of innovation theory (Rogers, 2003), posits that in order for an innovation to succeed, there is a need for a champion who would promote the innovation above others. On reflection, MALT could have been adopted more widely if additional efforts had been made to recruit local champions, especially within NHS organisations. Unfortunately, at the time I was prioritising other methods of promoting the model.

Reflecting upon user satisfaction, the interviews revealed that the majority of users appreciated the model. Overall, responses were positive, with the exception of two people left frustrated about

difficulties related to populating the tool with the data. They still, however, stated that they were satisfied with the model. The majority of users found the interface visually appealing, simple and easy to use, with minor suggestions about the additional options highlighted for consideration.

All interviewees judged the model to be useful and provided plenty of examples of situations when the tool was particularly beneficial. Overall, the comments were very positive and showed that users appreciated the effort that went into developing the model. Several users also enquired about when and if the model would be improved and developed further, and provided useful suggestions for doing so. For example, it was suggested that a function to save the work in progress would be very beneficial. One of the interviewees suggested a video tutorial to inform users about the types of data required by the tool.

Arguably, these issues could not be picked up by usability testing as they related to the way in which the user engaged with the tool in the normal work environment. This is especially important as, during the usability test, all the data were provided in a ready-to-use form. In a normal work environment, obtaining these figures might require substantial effort. It is even likely that some of the data would require pre-analysis, for example to deliver disaggregated costs.

Upon reflection, this suggests that model development for end-users (e.g. decision makers) needs additional dimensions relating to on-the-job testing/evaluation. As suggested earlier by the theory review, variables such as characteristics of the task or individual preferences have all an impact on how individuals perceive usefulness of a tool and what is their subjective satisfaction. One solution would be an iterative lab-based usability testing followed by separate on-the-job evaluation. The latter due to the ethical constraints of engaging NHS staff and other practicalities was not possible to achieve in this research. Importantly, only recently this gap in research have been addressed by the human-computer interaction specialists. Folstad (2007) felt that standard

methods of usability testing were lacking in this respect and proposed an enhanced framework for usability testing that would account for context of use.

Worth noting is the fact that the evaluation used the data available for the first 6 months' post dissemination. By the time this thesis was written (Jan 2018) 86 end-users registered on the site. A further evaluation on a larger sample and over a long period would be helpful for understanding the true potential and usefulness of the MALT tool.

Notably, it is important to understand that the perspective provided by the interviewed users does not necessarily represent that of all model users. Specifically, they do not represent the views of NHS users, as due to ethical constraints these were not involved. This is an important gap which requires further investigation. The literature suggests that in the NHS specifically there are various challenges to the use of modelling that do not exist in other fields that use modelling, such as the industry (Brailsford, 2005). Some of these barriers are now better understood than others, mainly thanks to the research contribution of Brailsford (2005, 2007). More research is needed to develop interventions that would help to improve engagement by NHS staff and to enhance their understanding of modelling benefits.

9.6.1 Quality

Several steps were taken to ensure the quality of this research. Firstly, the interview transcriptions were checked for accuracy by an independent person. During data analysis, we held meetings to discuss emerging findings, their implications for the model, and whether additional amendments were necessary. External scrutiny was sought from Dr Georgina Jones, who reviewed the findings and provided valuable advice. Special attention was given when extracting quotes to ensure that they reflected the views of participants. Lastly, in the process of reflexivity, I contemplated if I, as a researcher, had impacted the data, and sought as far as possible to ensure that the findings

were objective. The addition of respondent validation would have added credibility, but despite efforts, the findings were not validated.

10. Discussion

10.1 Introduction

This study is concerned with the development and evaluation of a decision analytical model for telehealth interventions in COPD. As set out in Chapter 1, the lack of evidence on the cost-effectiveness of telehealth technology is one of the main barriers to its wider adoption (Goodwin, 2010; Sanders *et al.*, 2012). The rationale for this study was based the premise that providing telehealth stakeholders with a web-based decision support tool may facilitate greater uptake of simulation and modelling in telehealth. It was seen as especially useful for decision-makers at the local level of the NHS, who are believed to deal with complex and 'messy' operational problems (Jennett *et al.*, 2003). For them, the potential to appraise various service configuration options may be more important than cost-effectiveness information alone.

The three main objectives of the study were to explore user requirements for the model, evaluate the usability of the model's interface, and explore user satisfaction with the tool. To address these objectives, the study employed three action research cycles. In the first cycle, qualitative interviews with telehealth stakeholders were conducted to elicit perspectives on requirements for the model. Based on these recommendations, the model was developed by a team of modellers from the Health Economics and Decision Science section of School of Health and Related Research. In the second cycle, the usability of the interface of the model was tested using the standard methods of user-centred design and usability practitioners. The third study explored end-user satisfaction and was intended to assess how useful the MALT model was in real-life decision-making.

This final chapter begins with an outline of, and reflection on the principal findings on barriers to modelling and simulation uptake in the NHS and the mainstreaming of telehealth technology.

349

This is followed by a discussion of the study's contribution to knowledge, and a comparison with the existing literature. The chapter concludes with a discussion on the strengths and weaknesses of the research and recommendations for further work.

10.2 Reflection on the principal findings

Critical reflection aims to find the meaning of a situation using ideas or theories (Fook & Gardner, 2007). Reflections on the main findings are discussed below in relation to each of the three research questions.

1) **What was the value of applying conceptual modelling and the usability testing during MALT model development?**

Conceptual modelling

In the process of developing the model, it was of primary importance to collect and understand background information on telehealth service delivery and operations in order to develop an appropriate model structure and viable model boundary. The conceptual modelling framework guided me through several strands of various research activities required to collect and to analyse that data. Within a short period of time, and before the coding work on the MALT model could start, it was necessary to answer some crucial questions.

The literature review enabled me to develop concept maps and causal diagrams that explained the natural history of COPD, the different components of the telehealth interventions, and the interplay between them. For me, the most important benefit was that I could use these diagrams in discussions with the modellers and the researchers from the MALT team. The next important

goal was to understand and depict the currently used COPD care pathways, with and without the use of technology. This was important to build a viable structure for the decision model.

Following this, I reviewed NHS business cases and costing tools to investigate the level of detail in financial analyses used by the NHS. NICE produces costing tools to support implementation of health programmes, but these are developed separately from the economic models. It was important to identify the level of detail that would be providing stakeholders with information they could use to support their business cases.

I then reviewed published economic models for COPD interventions to identify which types of models and model structures were viable and which should be avoided. This was important because structuring the economic model takes time, and once completed, amendments may not be possible without rebuilding the entire model (Kaltenthaler *et al*., 2011; Tappenden, 2012).

Finally, I conducted a series of 29 qualitative interviews (presented in Chapter 6) that explored the stakeholder requirements for the decision analytical model for telehealth. This was an 'amendment' to the conceptual modelling process developed by Squires (2014), who recommended conducting a focus group. There were several benefits to conducting qualitative interviews as part of the framework, and these were discussed in Chapter 2. For example, they enabled me to engage with stakeholders on a level that would not be achievable if a focus group were conducted. Having the opportunity for face-to-face meetings afforded the participants privacy and space for reflection. There were potentially some issues that may have not be articulated well if all the stakeholders had been gathered in one room.

The main findings from the qualitative interviews (Chapter 6) were how the views, goals, and ambitions in regards to telehealth technology differed between the main stakeholders. These

revealed how potentially they would use the MALT model. Moreover, the interviews directly informed the development of the model by providing examples of categories of costs, outcomes, and scenarios that were important to include in the model. Rich contextual data was also collected on the issues surrounding telehealth, some examples of which include operational issues, quality of evidence, and pre-requisites for cost-effectiveness. These all enhanced my understanding of telehealth and the surrounding issues.

On reflection, I believe that the conceptual modelling helped me to develop and to share my understanding of the nuances of the decision problem related to the use of telehealth in COPD. Due to the complexity involved in telehealth modelling, mistakes in depicting the conceptual model are not always obvious, but thanks to the framework I adopted more attention was paid to verification of service models, pathways, business models, and cash flows in the early stages of model development. This also prevented errors that would have been committed if we had started to build a model with an alternative structure without understanding vital components of the intervention. I feel that following the framework strengthened the credibility of the model and made the development process transparent to stakeholders.

One issue worth considering is whether or not incorporating additional qualitative interviews and deviating from the framework structure proposed by Squires (2014), was beneficial for the study. One of the obvious constraints was that conducting qualitative interviews instead of a focus group was time-consuming. I believe that in certain situations both methods can be applied, for example by conducting a focus group and then individual interviews if the group does not provide sufficient representation or coverage of specific issues.

Reflecting on methods that would make the process less time-consuming, I believe that a comprehensive, formal thematic analysis of qualitative data was unnecessary, and in certain

situations could have been abandoned. However, these are just two aspects of the qualitative interviews. In order to understand fully the benefits of conducting qualitative interviews as part of a conceptual modelling framework, further research is needed.

Usability testing

With a central task of developing a tool for end-users, it was necessary to conduct usability testing of the model's online interface in order to eradicate obvious problems of usage. A separate methodological review was conducted to inform the choice of the evaluation method as described in Chapter 4, and the 'Verbal Protocol Analysis' method was deemed the most appropriate (Ericsson & Simon, 1984). During the evaluation, usability problems were identified, and the model was improved through successive rounds of software development. An analysis of verbalisations suggested that language was one of the most important factors for the use of the model.

There are two views on the extent to which the usability testing of the MALT model was beneficial for the study. The modellers were principally concerned the quality of the model and that the results it produces are academically rigorous, thus considered the usability of the tool to be of secondary importance. This was because of the aforementioned issue of the dominance of the expert mode approach to modelling and their inexperience in end-user modelling. In contrast, as discussed in Section 8.9, the wider MALT research team acknowledged the significance of the testing. They were satisfied with the results of the testing and agreed that it was necessary to ensure the user-friendliness of the model.

On reflection, I consider the testing beneficial for the wider MALT study. My view was reinforced by the conversation I had with the CEO of a simulation company, who showed me that usability

testing was being routinely implemented by consultancies to ensure the quality of the tools they develop. My interviewee suggested that the development of a model that people can understand remains a challenge, and requires some form of validation. In her opinion, usability evaluations always lead to better uptake, as they prevent the model being abandoned due to issues related to ease of use. However, usability evaluation is time-consuming, and there is a trade-off between time spent conducting a series of tests and enhanced uptake.

This trade-off was not clear to me at the beginning of the project, and whilst I spent time undertaking testing and rigorously following the testing process, other modelling activities (such as model verification or promotion) were compromised. My understanding now is that the time taken to undertake usability testing should be flexible according to the needs of a particular project. For example, conducting it on a limited number of users in a less strict environment might be sufficient to identify areas of improvement.

Another criticism relates to the time taken to conduct the literature review on the usability methods (Chapter 4B). Arguably, similar results could have been achieved if I resorted to a consultation with a specialist in computer science and the usability field, rather than exploring the voluminous literature of a field outside my own expertise. However, at the time, the advice I received from usability experts and the literature to which they pointed me suggested that there is a lack of a clear statements of good practice. Therefore, the review provided a rationale for using certain methods, and meant it was not necessary to rely solely on experts' opinions. Upon reflection, I believe that I was too ambitious in my attempt to review and dissect the usability literature, given that the methods were not a central feature of the thesis. As a result, I identified a lack of good practice guidelines in this area, and feel that more research in this domain is needed.

Furthermore, as previously mentioned, the study demonstrated that some of the usability issues would only appear in real-life situations, and some practical issues with the model were not picked up by the usability testing and only revealed by model users after dissemination. Therefore, observations in the real world would be recommended as better than lab-based testing alone. Also, based on the results of theory review, I recommend that the work environment, user skills, and complexity of tasks be thoroughly examined when designing such a test in real-life situations.

What were the benefits of disseminating the MALT model in the 'end-user mode'?

In December 2014, the MALT model was disseminated in open access, with the view that upon registration on the project website anybody could use it free of charge. The users were given an opportunity to use the model themselves as a plug-and-play tool. This is the approach I named 'end-user mode' (Chapter 1), as opposed to the practice where users do not work with the models directly, but are provided with modelling results in the form of written reports.

At the outset of the project I assumed that the number of people who had signed in to use the MALT model, their feedback, and the logs of their activity would all be good indicators of whether the 'end-user mode' worked or not. Unexpectedly, by the end of May 2015, the number of registered users was 43. Of those who had registered on the website, the majority did not use the tool actively to any great extent. The number of registered users was lower than expected, but with a changing NHS environment in which the role of telehealth was uncertain, these numbers were deemed satisfactory.

The feedback generated from those who actively used the tool was positive and reflected appreciation of the fact that they could use the model on an ad hoc basis. The users provided examples of the MALT model being applied to telehealth evaluations and during development of

business cases. Importantly, they emphasised how useful it was to be able to run the analyses during meetings.

Upon reflection, I believe that the 'end-user mode' is a successful method of dissemination of decision models in this type of evaluation. It addresses the previously mentioned barriers to modelling use (Chapter 4A), namely the cost of modelling, the problem of a lack of modelling expertise, and a general problem of accessibility of modelling. These are believed to be some of the major barriers to the uptake of modelling (Williams *et al*., 2008). The mode successfully addressed the problem of cost of modelling expertise, as it was free of charge and available to all. Additionally, the MALT model was intuitive to use, so that even non-experts or people with minimal knowledge of health economics could use it. It also enabled people to conduct analyses flexibly, for example during the meetings.

Furthermore, as previously mentioned, the MALT model garnered a lot of interest during MALT project conferences and events, with attendees providing very favourable comments. However, this interest and the positive comments/experience is difficult to reconcile with the low number of registrations and high rates of supervision use. This issue was not fully resolved by the additional interviews undertaken after the release of the model, and evaluation of such an eventuality should be built into future releases of end-user models.

Moreover, seeking to understand the benefits of the 'end-user' mode and its potential for replication in other modelling projects, I interviewed the CEO of a leading simulation company. She emphasised that the 'end-user mode' is the favoured approach to model development and dissemination for commercial use. She suggested that the company is committed to making models straightforward and intuitive to use. This statement reinforced my belief that there is a potential for this approach to be replicated and applied to other areas of modelling.

However, in order to ensure successful dissemination of models in this mode, the above-mentioned company provides various types of support for users. Firstly, they provide video tutorials to help people understand the modelling process step-by-step, and a 24-hour support line where users can seek help from an analyst. Secondly, there are also several training packages available for novices, and an online forum for discussion with other software users. This type of support was lacking in the case of the MALT model. Importantly, even with all of these advances, the company had evidence of people struggling and unable to work with their models, mainly due to the lack of data and/or time to obtain data. The conclusions reached were similar to those from my interviews with MALT model users and suggests that NHS management specifically needs to be made aware of the importance of data-related issues.

2) **What was the user satisfaction level with the MALT model? What was the usefulness of the model?**

The third part of the study sought to explore the usefulness of the model in real-life decision-making. Initially, I used a questionnaire to collect data on user satisfaction. However, when an analysis of the preliminary data revealed a poor response rate to the satisfaction questionnaire and limited use of the model, I decided to conduct additional qualitative interviews with registered users to explore why the model has not been used, along with the subjective satisfaction rates of those who had used it.

The data from the demographic questionnaire provided a valuable insight into the structure of the user sample, and the results of the qualitative interviews indicated a positive experience for active users. It also revealed the variety of projects to which the model was applied.

The satisfaction evaluation was an important component of the project. However, upon reflection, it could have been conducted using different methods. The feedback I received from the OR

practitioners during one of the annual conferences suggested that an alternative method could have been asking registered users to video record their activity and then to rate their satisfaction. Instead, I took a more traditional approach, conducting a satisfaction survey and qualitative interviews. Importantly, one of the weaknesses of this method is that it did not provide data on users' satisfaction after prolonged use, as the survey only delivered the data on one time-point.

I believe that I put too much emphasis on identifying and using a pre-existing and validated instrument for the purpose of a satisfaction evaluation. The decision to adopt the EUCS instrument was dictated by there being currently no instrument that could be used for satisfaction evaluations in health economics. I considered developing a new instrument, but the time constraints of this PhD meant that this could not be done. I also envisaged that applying EUCS in this domain could potentially add to the wider literature related to the instrument.

Importantly, reflecting on the usefulness of the MALT model, there might have been some non-directly observable benefits for model users. Monks (2014) suggested that through experimentation with the model, end-users learned about resource utilisation in hospital wards. This suggests that working with the model helps with to obtain knowledge that could be applied to future projects of a similar character. There is therefore potential for the MALT model to contribute to a better understanding of issues related to the evaluation of telehealth.

3) The role of government policy in the uptake of telehealth

I believe that telehealth is a powerful technology that has a potential to empower people to better manage their condition(s) (Car *et al*., 2012; Polisena J. *et al*., 2009). Specific telehealth applications have been shown to improve self-management of long-term conditions (Car *et al*., 2012; Gorst *et al*., 2014; Taylor *et al*., 2014), reducing the rate of hospital admissions (Darkins *et al*., 2015; Holland, 2013), and improving patients' outcomes (Ekeland *et al*., 2010; Totten *et*

al., 2016). Most importantly, studies suggest high satisfaction of patients using telehealth technology (Kruse *et al*., 2017; Polinski *et al*., 2016).

With all the benefits of telehealth related to patient care, there are still uncertainties around its impact on costs and use of resources. The results of cost-effectiveness studies are inconclusive and generally recommend that more research in the area is needed (Mistry, 2012; Mistry *et al*., 2014; Udsen *et al*., 2014). I believe that telehealth has the potential to be a meaningful tool to address the problems of the growing demand for healthcare and resource pressures, as long as decision makers understand the prerequisites of successful deployment in their local context. As explained by Car (2012) p.1) 'particular interventions may be successful', but that depends on, 'the interplay of 'technology, service designs, clinical input and patient involvement'.

This is the area where I perceive the MALT model to be particularly useful. Any evaluation conducted by the MALT model recognises that different types of technology and different service provision models exist, and in order to find an optimal solution, several scenarios should be considered. A particular option (combination) of technology and service model might be cost-effective for one service (with a particular size population, disease severity, and percentage of uptake) but less effective for others.

Reflecting on the meaning of government policies related to telehealth and ICT in general (as discussed in Chapter 1), there has been a shift from the promotion of technology for administrative and managerial purposes to the application of technology for empowering patients, and, 'to improve the health of population' (DOH, 1998). Importantly, the majority of these policies encourage the employment of new technologies with an emphasis on efficiency and effectiveness. As a result, decision-makers see telehealth as a means of containing the cost of healthcare services, without acknowledging that this needs to be part of wider organisational change

(Klecun-Dabrowska & Cornford, 2000). The literature suggest that following publication of the WSD trial results, Department of Health support for telehealth waned, which had a negative impact on the expansion of telehealth schemes (Clark & Goodwin, 2010; Goodwin, 2009, 2011).

With all of the uncertainties related to future financing of telehealth schemes, and an unclear evidence base, many of those who implemented telehealth abandoned it after realising that it did not deliver the expected benefits. For example, a recent post-implementation review (personal communication) that examined the use of telehealth in care homes in one NHS service in North Yorkshire concluded that the telehealth system did not deliver the desired reductions in non-elective admissions. As a result, the service was decommissioned after 12 months of operation. This is one of many examples of telehealth schemes struggling to embed or being abandoned after a short period of operation that I came across during my work as a research associate (personal communication).

10.3. Contribution to knowledge

The main contribution of this study is its demonstration that the 'end-user mode' of model dissemination is viable for evaluations of complex interventions. The mode is particularly beneficial for interventions where service specifications change frequently and end-users might need to explore various scenarios or options. This is an alternative to practices used in OR (Franco & Rouwette, 2011) and HE, especially in the development of models for the purposes of HTA (Philips *et al*., 2004). As mentioned in Chapter 1, with the exception of 'facilitated modelling', in standard practice, decision models are developed in 'expert mode' to deliver an answer to clients' queries. Some elements of the approach proposed here have been previously applied to case studies of vaccines (Hubben *et al*., 2007), blood transfusion (Van Hulst *et al*., 2009), and hospital decision support system (Brailsford *et al*., 2011), though never formally named and tested in a

case study of COPD telehealth. This study contributes to the extant knowledge by providing an enhanced understanding of the development and use of the decision model for telehealth COPD in this mode and complementing the findings of Hubben (2009; 2007) and Van Hulst *et al* (2009). The current research makes three additional key contributions to knowledge that are discussed below and situated within the current literature.

10.3.1 The uptake of modelling in the NHS

Previous research stressed that decision modelling is frequently overlooked because of various barriers (Chapter 4) and emphasised that more research was needed to find the right approach to communicating economic evidence to stakeholders (van Gool, 2007; Williams *et al*., 2008).

The results from a recent review of simulation and modelling in healthcare suggested that only 5% of the published models had been used in practice (Brailsford *et al*., 2016). The improvement of modelling uptake and use has been a concern of OR researchers, and efforts have been made to resolve the issue. The last 20 years of behavioural OR research was devoted to developing methods for improving modelling uptake and use, with a particular interest in improving decision-making with models, and understanding how individuals work with models (Willemain *et al*., 2003) and how models are developed (Tako & Robinson, 2010).

As discussed in Chapter 4A, the rates of modelling use and embedding within NHS organisations are limited by existing system-wide, organisational, and personal barriers. Brailsford and colleagues (2011) suggest that the primary reason for NHS staff abandoning the Scenario Generator model was due to data-related issues. In addition, the study confirmed that time and effort expectancy played an important role in the decision of whether or not to use the model. Similarly, in this research, the feedback from those who failed to use the MALT model suggested the existence of the same mechanisms.

The current study also replicated the findings from the above-mentioned study (Brailsford *et al*., 2011) related to the type of users. As well as a group of people who used the tool actively, other MALT model users either showed an initial interest but did not use the tool actively, or used it but then abandoned it for a variety of reasons. Brailsford *et al* (2011), categorised these end-users as 'active users', 'given up', and 'not started'. Importantly, they concluded that modelling is adopted within the NHS for routine use by only a limited number of end-users. This was also pertinent for the MALT model.

In an attempt to understand what affected the decision to use modelling and how individuals work with a model, early behavioural OR studies considered the impact of the type of display. For example, Bell and O'Keefe (1995), and Willemain *et al* (1989), experimented with different animation and visualisation options and found that visualisation methods in simulation modelling were associated with more efficient use of a simulation model, though they did not affect the locating of a correct solution. They also found that people who obtained incorrect solutions investigated fewer alternatives than those who obtained correct ones. The importance of visualisation has been confirmed more recently by Akpan *et al* (2014), who demonstrated that 3D animations enhanced user performance more than other types of graphic displays (2D) and that types of display may have an effect on the overall success of a simulation project.

With the increasing role of models serving as support systems, Willemain *et al* (2003), conducted a series of experiments with a spreadsheet-based decision support system to test if humans could notice and overcome errors made by the software. Whilst not the focus of the study, Willemain and colleagues (2003), identified important differences in the way models were used by experts and novices. They pointed to the fact that expertise in modelling did not lead to better performances with the model, and that experts and novice users in general worked with the model in different ways; e.g. novices explored more options. Although this aspect of MALT has not

been formally tested, e.g. by comparing the performance of novices and experts, the findings from the usability evaluation confirmed that language is one of the key barriers for novice users.

Within behavioural OR, the involvement of stakeholders in the model development process is valued (Franco & Rouwette, 2011; Howick *et al*., 2008). Traditionally, stakeholders have been involved in model development mainly during facilitated modelling sessions, where they attend series of workshops to discuss approaches to modelling. Previous studies concluded that this led to improved understanding of modelling problems and facilitated consensus between those involved (Franco & Rouwette, 2011; Franco & Montibeller, 2010). Recently, the development of the PartiSim framework complemented this research, suggesting ways in which stakeholders could be engaged in other stages of a modelling project (Tako & Kotiadis, 2015). This included a series of workshop activities when the model was being developed, as well as informal meetings, field observations, and one-to-one interviews when the study is initiated (Tako & Kotiadis, 2015). This study contributes to an understanding of the benefits and challenges of engaging with stakeholders during the conceptual modelling phase, and reveals some of the challenges that occur after the dissemination of the model. I accept the view of the literature, that improved stakeholder engagement may ultimately lead to models getting 'buy in' and being implemented, and more research is needed in this domain (Taylor *et al*., 2013c).

Moreover, the findings from this study provide some evidence on the use of modelling in real-life decision-making related to telehealth. To my knowledge, it is the first study in the field that attempts to evaluate the satisfaction of end-users and to assess the usefulness of modelling. Importantly, the concept of usefulness is still unclear in this context. In general, the OR literature suggests that model results can be considered useful if they have been applied in real life (Robinson & Pidd, 1998) or contributed to a change in the perception of the decision problem (Eddama & Coast, 2008). The current study provides empirical evidence of what happens with a

363

particular application after its release. The data collected includes at least two examples of results of modelling used in decision-making.

I found this aspect of my thesis to be particularly interesting, as the literature suggests that even if a model is not implemented or routinely used, stakeholders benefit from being involved in the process by acquiring knowledge that could be applied in future projects (Monks *et al*., 2014). The transfer of learning has not been tested empirically in this research, but the findings from the qualitative interviews suggest that there were some indirect benefits from being involved in the MALT model development, for example, the stakeholders realised the importance of data.

Moreover, the current study answered Ormerod's (2013) call for more informative modelling studies, as it reveals how the process of model development evolved or changed direction over time, and how it was influenced by the people involved in the project (Chapter 2). This anecdotal evidence of what happens during OR intervention is especially useful for other OR practitioners, who could use the knowledge to avoid mistakes in other similar projects.

In summary, whilst the current body of evidence in the OR and modelling literature suggests that the barriers to simulation and modelling still exist, resulting in low uptake of models (Eldabi *et al*., 2007), there have been various research efforts to overcome them. This research has been designed to complement the existing research by proposing an 'end-user-mode' for evaluations of telehealth interventions.

10.3.2 Modelling in telehealth

My findings provide further insight into ongoing discussions of barriers to telehealth uptake. For example, when discussing barriers to telehealth mainstreaming, Goodwin (2010) concludes that there is insufficient evidence of the cost-effectiveness of telehealth and that more clinical trials

are needed. My research highlights that the quality and availability of data also plays an important role in the decision to use evaluation tools, and without data, even the most sophisticated models will fail and be abandoned.

The study also provides evidence of unresolved issues relating to the cost-effectiveness of telehealth, by exploring the views of various stakeholders. The main findings from the qualitative interviews (Chapter 6) demonstrate a need for a shift in understanding to what telehealth could achieve and how value could be created. In addition, the qualitative interviews reveal a number of barriers currently preventing the effective adoption of telehealth within the region. Expert advice was provided to participating stakeholders on several issues that are beyond the scope of this thesis.

It also demonstrates that routine hospitalisation data could be used to define the states in the Markov model and to estimate transition probabilities. This approach has been successfully applied in two previous developments, e.g. in the Model for Assessment of Telemedicine Applications (MAST) (Kidholm *et al*., 2012) and in the Long-Term Conditions Year of Care Simulation model (LTCYOC). In this sense, the current study may broaden the applicability of this approach to the area of telehealth.

The current research also provides evidence in support of the claim that modelling could be more widely used to support resource allocation planning in the NHS. Over the years, NHS England has released evaluation tools to support budget planning. One such tool, highlighted previously, was the Long-Term Conditions Year of Care (LTCYOC) simulation model (NHS England, 2014), which was disseminated as part of the Year of Care Commissioning Programme. I discussed the LTCYOC model development process and its main features with the developer, Dr Jamie Day, and representatives from simulation company, Simul8. They agreed that default values to use in

the existing scenarios and costs presented in disaggregated form are features that could facilitate the use of the model by non-experts.

10.3.3 Methodology

From a methodological perspective, this study further improves our understanding of the value of conceptual modelling frameworks. The literature suggests that during model development, modellers spend less time on problem structuring than on any other activities, such as data input, model coding, and model validation (Tako & Robinson, 2010; Willemain, 1995). This study demonstrates the practical implications and benefits of committing time and resources to these early stages of model development.

As an initial step towards addressing the complexity of telehealth modelling, the study adopted Squires's (2014) conceptual modelling framework. Because of the character of the project and many uncertainties related to service designs and provision models, additional qualitative interviews were conducted with stakeholders to elicit their requirements. In addition, because of the end-user character of the model, the conceptual modelling was extended to include issues of model functionality and outputs. These were not originally covered by Squires's (2014) framework.

Another methodological contribution was the exploration of the potential value of end-user satisfaction concept as a metric for research. Through a structured review, the EUCS instrument was identified as a potentially valuable evaluation instrument and an attempt was made to adapt and apply it to this research. Whilst response rates prevented any further evaluation of its value, end-user computer satisfaction remains an important topic for further research.

Lastly, the study identified limits to lab-based usability testing by demonstrating that some of the usability issues appeared only after the dissemination of the model. This points to other methods (e.g. observation) as being of importance to the evaluation of decision support systems.

10.4 Strengths and weaknesses of the current study

A key strength of this research is its interdisciplinary character. The methods and knowledge from the fields of information technology and system research, operations research, management, and the social sciences were successfully applied to respond to the research question.

In particular, the use of the 'Think Aloud' method and the use of video to collect participants' verbalisations and capture their non-verbal expressions was novel for the modelling domain. The use of video for data capturing and analysis is actively exploited in social sciences (Margolis & Pauwels, 2011) and is prevalent in the assessment of usability in the engineering field. However, the usability practitioners with whom I engaged for this study had not yet agreed on standardised methods for analysis of such data, which made the task very challenging.

Significant efforts were made to ensure that the usability test sessions operated strictly in accordance with the good practices identified in my review. The study by Norgaard and Hornback (2006) revealed that the majority of usability test sessions were open to bias if researchers did not follow the guidance provided by Ericson and Simon. For example, researchers are not allowed to communicate with the participants when they fall silent, and can only use the 'keep talking' prompts in predefined time intervals.

More generally, important design decisions were identified throughout my work and checked against other research studies or discussed with specialists in the field. For example, literature

reviews were conducted to strengthen understanding of key concepts such as usability evaluation methods, theories of technology adoption, and barriers to modelling and simulation use in the NHS. Particularly challenging was the review of theory, as no clear guidance exists on how such reviews should be conducted or their results analysed.

A key strength of the qualitative phase of this mixed methods study was the large number of participants recruited. The qualitative interviews were conducted with 29 participants from four NHS trusts. The usability evaluation was conducted with 16 participants over three rounds of tests, and qualitative interviews of the model's usefulness were conducted with a further five participants.

The qualitative data was also collected from different samples. Participants of the first case study were representatives of the telehealth industry, NHS managers, and commissioners. In the second case study (the usability test), students, telehealth professionals, and telehealth consultants were recruited. In the last of the case studies, non-NHS professionals took part.

The main limitation of the study is the low number of NHS staff involved in the satisfaction evaluation and subsequent interviews. It is therefore likely that the study does not represent the full range of views of potential users of the MALT model.

As discussed in Chapter 5, the mixed methods approach had to be abandoned and a new iterative and emergent design put in place. The final design was a result of several cycles of reflection that contributed to answering the main research question.

Another potential weakness is that it was not possible to collect and analyse the data from the EUCS questionnaire. The low response rate meant that it was not possible to explore the determinants of user satisfaction and how it varied between one-off and routine use.

10.4 Recommendations for future research

The study demonstrated an approach to the development and dissemination of a decision model for telehealth. Despite the growing evidence from BOR on the use of models, more research into how models are used in the process of decision-making and the creation of knowledge is needed. This is one of the gaps identified in this research that suggests that direct observation of decision-makers' interactions with models and the development of methods would be beneficial for economic modelling and OR.

Furthermore, the results of this study suggest that the development and validation of a specific instrument that could measure the usefulness of, and user satisfaction with decision models is needed. For the purposes of this evaluation, EUCS was used. As discussed previously, this instrument may not be appropriate for systems that are not used regularly, due to the response categories.

Appendix A.1 The timescales of the MALT project

<u>**PHASE 1**</u>

Table 34: The timescales of the MALT project in phase one

Ref	Title	External / Internal (E/I)	Res pon sibi lity	Due Date
D1	Literature review of existing cost-effectiveness models in COPD/CHF	I	KL/ AB/ SD	June 2011
D2	Conceptual model of COPD and CHF care	I	KL/ AB/ SD	August 2011
D3	Cost-effectiveness model	I	AB/ SD	November 2011
D4	Literature review of the use of short-run costs	I	KL/ AB/ SD	Dec 2011
D5	Discussions with Site 1 and Site 2 service providers	E	KL/ SD	Dec 2012
D6	Discussions with Site 3 service providers	E	KL/ SD	Jan 2013

Ref	Title	External / Internal (E/I)	Res pon sibi lity	Due Date
D7	Discussions with Site 4 service providers	E	PT/ HB	April 2013
D8	Interviews with service providers and other stakeholders relating to modelling requirements	E	KL/ SD	Jan 2013
D9	Development of financial model	I	AB/ PT	August 2013
D10	Development of linked cost-effectiveness and financial model	I	AB/ PT	August 2013

Table 35: The timescales of the MALT project in phase two

PHASE 2: Application of models to case studies
Start date: April 2013 End date: November 2014 Work Package Leader: UoS, Simon Dixon Contributing Partners: UoS, UoL, Epigenesys

Objectives:

- To use the financial model alongside service providers to predict activity and costs

- To disseminate the model to wider public

Description of work

- The linked model from WP 1.1 will be used to produce projections of activity and costs based on the planned implementation of ALTs in COPD and CHF in two sites
- Projected and out-turn figures will be compared to test the validity of the models with discrepancies described
- Models will be amended in line with new service data becoming available and to remove discrepancies
- Development of an online interface on projects website

Summary of Deliverables : Detail the planned external (E) and internal (I) deliverables

Ref	Title	External / Internal (E/I)	Responsibility	Due Date	Comments / Notes
D1	Initial projections of activity and costs based on linked cost-effectiveness and financial model for Site 2 and Site 3 and start of associated data collection	I	AB/PT/SD	January 2014	Internal document
D2	Ongoing feedback from associated sites on their experiences with the linked model	E	PT	March 2014	Internal document

D3	Comparison of predicted and out-turn costs and outcomes, feedback on user experience and final report	I/E	AB/PT/SD	Nov 2014	Final report
Dependencies					
Items which must be available for this Work Package					Responsibility
- Outputs from WP1.1 - Access to service providers - Service data from study data collection, e.g. EQ-5D					- SD - MH - MH, SB, GM
Work Packages dependent on this Work Package					Responsibility
WP3					SD

The MALT project team, economic modellers: Prof Simon Dixon (SD), Prof Alan Brennan (AB),

Dr Praveen Thokala (PT), Dr Hassan Baalbaki (HB), Dr Pete Dodd (PD), Kinga Lowrie

The MALT study group: Dr Chris Armitage, Lauren Beaumont, Lucy Bolton, Dr Liz Brewster, Dr Hassan Baalbaki, Prof Alan Brennan, Dr Lizzie Coates, Chris Clegg, Prof Simon Dixon, Dr Pete Dodd, John Eaglesham, Sarah Gorst, Prof Mark Hawley (Chief Investigator), Dr Helen Hughes, Huw Jones, Kinga Lowrie, Olubukola Otesile, Kathryn MacKellar, Prof Gail Mountain, Dr Duncan Ross, Dr Johanna Taylor, Dr Praveen Thokala and Prof Bridgette Wessels.

Throughout the Phase 1 and Phase 2 of the project we held weekly meetings with modellers,

monthly meetings with MALT study group and quarterly meetings with Technology Strategy

Board.

APPENDIX A.2 NHS Business cases included and excluded from the review

NHS business cases included in the review:

Oxford University Hospitals NHS Trust. Expansion of Newborn Facilities. Full Business Case. http://www.ouh.nhs.uk/about/trust-board/2012/may/documents/TB201235a_ExpansionofNeonatalFacilities2.pdf Accessed on the 29/08/2012

Hull and East Yorkshire Hospitals NHS Trust. Emergency Department Reconfiguration. Full Business, http://www.cht.nhs.uk/fileadmin/site_setup/contentUploads/Publications/Full_Business_Case_-_FINAL_Approved_by_Board_August_2017.pdf Accessed on the 30/08/2012

The Ipswich Hospital NHS Trust. Invasive Cardiology Service Development. Full Business Case. http://www.ipswichhospital.nhs.uk/aboutourhospital/Documents/Publication%20Scheme/IHT%20Cardiology%20Development%20FBC-v1.2-PUBLIC.pdf .Accessed on the 14/08/2012

The Royal Wolverhampton Hospital. Integrated Pathology Services. FBC. http://www.royalwolverhampton.nhs.uk/EasySiteWeb/GatewayLink.aspx?alId=4841.Accessed on the 10/08/2012

The Shewsbury and Telford Hospital. The Future Configuration of Hospital Services. FBC. https://www.sath.nhs.uk/wp-content/uploads/2016/09/Encl7-FCHS-appx1.pdf Accessed on the 11/08/2012

NHS Lothian. Royal Victoria Building. FBC. http://www.nhslothian.scot.nhs.uk/OurOrganisation/PropertyDevelopments/CompletedDevelopments/NewBuilding/News/ReProvisionRVHServicesOutlineBusinessCase.pdf Accessed on the 11/08/2012

NHS business cases excluded:

1. Contractual Transfer (Externalisation) of NHS Haringey and NHS Islington's Community Service APOs ("The Alliance") to The Whittington Hospital NHS Trust. Full Business Case
2. Delivering Integrated Health and Social Care, Full Business Case for the development of community health and social care centre at Yate, South Gloucestershire, August 2009.
3. Developing a NHS Dental Centre, NHS Highlands.
4. Full Business Case for the Redevelopment of Salford Royal Foundation Trust site. October 2007
5. Modernizing Health Services in North Staffordshire - The Redevelopment of the City General and Haywood Hospital Sites. Full Business Case.
6. Musselburgh Primary Care Centre, Full Business Case, November 2009.
7. New Hospitals Programme, Barts and the London NHS Trust, July 2006
8. NHS Forth Valley Healthcare Strategy, New Acute Hospital Full Business Case Addendum.

9. NHS Grampian Health Campus Programme Emergency Care Centre Project – Full Business Case
10. NHS Highland. Sutherland Older Adults Services. Full Business Case. July 2009
11. NHS Wirral. Full Business Case St Catherine's Health Centre. September 2009
12. North Somerset NHS. Clevedon Hospital Redevelopment Full Business Case. March 2012
13. North Walsham & District War Memorial Hospital. Full Business Case. March 2011
14. Redevelopment of South Petherton Community Hospital. Full Business Case, June 2009.
15. Single Stage Full Business Case for Morpeth NHS Centre, Northumberland
16. St Helens and Knowsley Hospitals NHS Trust, Strategic Redevelopment Project (hospitals)
17. The proposed merger of Ealing Hospital NHS Trust and The North West London Hospitals, Full Business Case, May 2012.
18. West London Mental Health, Broadmoor Redevelopment Enabling Works, Full Business Case (FBC)

APPENDIX A.3 Search strategies in the systematic review on Markov model structures

MEDLINE Ovid 1948- present

1. exp chronic obstructive pulmonary disease/
2. exp pulmonary disease chronic obstructive/
3. exp lung disease$, obstructive/ or bronchitis, chronic/
4. or/1-3
5. Economic filter
6. 4 and 5

EMBASE 1980-present

1. exp chronic obstructive pulmonary disease/
2. exp pulmonary disease chronic obstructive/
3. exp lung disease$, obstructive/ or bronchitis, chronic/
4. Economic filter
5. 4 and 5

Science Citation Index: Web of Science. 1990 - present

1. Chronic obstructive pulmonary disease
2. (cost* and (effective* or utilit* or benefit* or minimi*))
3. (economic* or pharmacoeconomic* or pharmaco-economic*)
4. Or/ 2-3
5. 1 and 4

Economic filter

Ovid

1. Exp "cost and cost analysis"/
2. Economics/
3. Exp economics, hospital/
4. Exp economics, medical/
5. Economics,nursing/
6. Exp models,economics/
7. Economics,pharmaceutical/
8. Exp "fees and charges"
9. Exp budgets/
10. Budget$.tw
11. Ec.fs
12. Cost$.ti
13. (cost$ adj2 (effective$ or utilit$ or benefit$ or minimi$)).ab
14. (economic$ or pharmacoeconomic$ or pharmaco-economic$).ti
15. (price$ or pricing$).tw
16. (financial or finance or finances or financed).tw
17. (fee or fees).tw
18. (value adj2 (money or monetary)).tw
19. Quality-adjusted life years/
20. (qaly or qalys).af
21. (quality adjusted life year or quality adjusted life years).af
22. Or/1-21

Embase

1. Socioeconomics/
2. Cost benefit analysis/
3. Cost effectiveness analysis/
4. Cost of illness/
5. Cost control/
6. Economic aspect/
7. Financial management/
8. Health care cost/
9. Health care financing/
10. Health economics/
11. Hospital cost/
12. (fiscal or financial or finance or funding).tw.
13. Cost minimization analysis/
14. (cost adj estimate$).mp.
15. (cost adj variable$).mp.
16. (unit adj cost$).mp.
17. Or/1-16

APPENDIX A.4 Interview topic guide

Overcoming barriers to mainstreaming Assisted Living Technologies

TOPIC GUIDE

Providers and Commissioners
Version 2 13-09-12

Aim of interviews:

▪ To gain an understanding of NHS service providers', Telehealth providers and commissioners 'requirements for economic and financial models for Telehealth commissioned services.

Introduction:

▪ Introduce yourself

▪ Introduce the study:
The MALT (Mainstreaming Assisted Living Technologies) project is undertaken by the University of Sheffield in collaboration with the University of Leeds and an industrial partner the Advanced Digital Institute. The project aims to understand the barriers for mainstreaming assisted living technologies through work developed by three workstreams: economic, business modelling, and patients' and carers acceptance. The findings of the research from these workstreams will inform action research in the telehealth services in three project sites. The economic workstream of the project aim to develop economic and financial model that can be used for variety of purposes. The model will enable evaluating cost and benefits in the existing services. Moreover, it will allow planning changes by assessing several "what if" scenarios. The model will be made available to all participants after the study is complete.

- In order to develop the tool that meets the requirements of its users, we are conducting the interviews with providers and commissioners of telehealth services and representatives of telehealth industry.
- Reassure participants that all information will be kept confidential and that it will be anonymous.

  - Confirm that the interview data will be used only by the MALT researchers

- Explain process of transcription, that the data will be stored securely and destroyed after publication of findings

- Ask for the permission to use a tape recorder:

  - If yes, proceed

  - If no, researcher to record notes manually and explain that the written notes are to be used as an aid for the analysis

- Check, if participant has any questions at all at this stage

- Ask participant to sign consent form to provide written consent for participation

- Thank the person for agreeing to participate

1. **I would like to start by asking few general questions about telehealth in your organisation.**
   - Do you recollect how the decision to introduce telehealth was made?
   Prompt Who was involved in decisions?
             Do you think these were the right people or others should be involved?
             Were there any meetings organised to discuss this?
             Does any written reports exist from that time?

2. **Could you tell me more about your current telehealth system?**
   - In your opinion is it reliable?
   - What data is collected on performance of the system and what does it include?
   - How useful is that data?

- In your opinion, do you have enough data available on performance of (your) telehealth system?

3. **My next question is related to the issue of cost-effectiveness;**
   - Could you explain to me how do you understand the term "cost-effectiveness"?
   - What does "cost-effectiveness" mean in case of telehealth service?
   - Have you ever performed or used the results of any cost-effectiveness analyses before?
     If YES what for?
     If NO why not?

4. **Talking about telehealth service in general, could you explain;**
   - What savings would you expect to see in the organisation following the introduction of telehealth?
   - Are there any reasons you think might explain why telehealth might not generate savings?
   - What costs are important to consider in such service?

5. **Can we now focus on progression of the disease in case of COPD patients; can you tell me:**
   - What are the major "events" COPD patients' experiences in the course of the disease?
   - Which one of these would you consider especially important from the service perspective?

6. **I would now like to ask you some questions which focus on benefits of telehealth;**
   - What do you think the benefits of telehealth monitoring are for COPD patients?
   - What do you think the benefits of telehealth monitoring are for the service?
   - How else could the healthcare organisation benefit from introducing telehealth?
   - Which one of these would you consider important to be included in an analysis/evaluation?

7. **If we produce a tool for performing CE analyses:**
   - Who would use it in your organisation?
   - What aspects of the tool would be most useful?
   - How would you use the results of the analysis and what for?
   - Are there any changes to the services planned for the future which you might like to evaluate with the help of the tool? *Prompt* (cost depending on speed of deployment?)

8.   **Having a tool for performing cost-effectiveness analyses, what would be your requirements regarding:**

- Software platform?
- Operating the tool?
- User friendliness of the tool?
- Presentation of the results?

9.   **The last question I would like to ask is related to financial planning tools;**

- Firstly, can you tell me how do you understand what financial planning tools are?
- Have you used any financial planning tools in your telehealth service?
  If YES Where they useful? Which parts were good and which parts were bad?
  If NOT Why not?
- What calculations such tool should perform to be useful to you?
- Can you think of any 'what if' analyses you would want evaluating?

10.   **Thank you for giving up your time to participate in this interview, do you have anything further you would like to add?**

APPENDIX A.5 Thematic framework

1. Rationale for telehealth implementation

   1.1 Department of Health policy

   1.2 Local champions

   1.3 Evidence based practise

   1.4 Continue to use the technology left after trial was completed

   1.5 National incentives - telehealth part of QIPP

2. Indicators of telehealth effectiveness/ efficiency

   2.1 Reduced number of A&E visits

   2.2 Reduced number of hospital admissions

   2.3 Reduced Length of Stay

3. Reliability of available data

   3.1 Quality of data

   3.2 Robustness of evidence

   3.2 Relying on the NHS staff

   3.3 Critique of WSD

4. Identified reasons why the technology is not generating savings

   4.1 Lack of organisational changes in the NHS organisations

      4.1.1 Design of the care pathway

      4.1.2 Staff workload

4.1.3 Scale of deployment

4.1.4 Speed of deployment

4.3 Issue of double running costs

4.3 It takes time to generate savings

4.4 Savings not directly visible – flows in whole healthcare and social care systems

5. Categories of cost
   5.1 Implementation costs

           5.1.1 Capital cost of telehealth equipment

           5.1.2 Cost of installation

           5.1.3 Cost of training for staff and patients

           5.1.4 Cost of setting up the machines

           5.1.5 Cost of establishing new care pathways for use of telehealth – e.g. new data bases, methods of doing necessary activities

   5.2 Operational cost

           5.2.1 Cost of providing monitoring

           5.2.2 Cost of technical triage

           5.2.3 Cost of clinical triage

           5.2.4 Cost of service administration

           5.2.5 Cost of maintenance

           5.2.6 Ongoing requirement for training

           5.2.7 Ongoing costs of installation/ de installation/ recycling

6. <u>Methodological challenges in estimating cost</u>

      6.1 System wide costs

      6.2 Hidden cost

      6.3 Intangible cost

      6.4 Lack of data

      6.5 Getting a true picture

      6.6 Sunk costs

7<u>. Indicators of the severity of the COPD disease</u>

      7.1COPD exacerbations

      7.2 COPD related hospitalisations

      7.3 COPD related A&E visits

8. <u>Perceived benefits of using telehealth</u>

      8.1 Benefits in health service efficiency

            8.1.1 Reduced number of A&E visits

            8.1.2 Reduced number of GP visits

            8.1.3 Reduced number of nurse visits

            8.1.4 Reduced number of hospitalisations

            8.1.5 Reduced LoS

8.2 Benefits in quality of service

      8.2.1 Improved access to care

8.2.2 Faster time of response in case of exacerbations

8.2.3 Maintaining up-to-date patient record

8.2.4 Evidence in case of patient negligence

8.2.5 Safety umbrella for staff

8.3 Quality of life benefits for patients

8.3.1 Avoided hospitalisations

8.3.2 Improved medical compliance

8.3.3 Reduction of anxiety

8.3.4 Maintaining independence

8.4 Categories of intangible benefits

8.4.1 Saving cost and travel time for patients

8.4.2 Saving cost and travel time for carers

8.4.2 Saving costs and travel time for staff taken in face to face sessions

8.4.3 Moving care to community

9. Perceived application of the proposed model

9.1 Generating evidence that TH improves patient QoL

9.2 Generating evidence for TH cost-effectiveness

9.3 Strategic planning of allocation of resources

9.5 Evidence for savings in the long term

9.6 To inform business case

10. Requirements for the tool

    10.1 Flexible

    10.2 Transparent

    10.3 Excel based

11. What analyses stakeholders would like to perform?

    11.1 Investment per patient vs. predicted outcomes

    11.2 What is the optimal use of TH technology using existing resources?

    11.3 Duration of deployment

    11.2 Scale of deployment

    11.3 Target population

    11.4 Procurement scenarios

    11.5 Risk sharing and co-payment

12. Approach to financial planning tools

    12.1 Return on Investment

    12.2 Net Present Value

    12.3 Social Return on Investment

    12.4 Cash flows

APPENDIX A.6.1 Participant information sheet

PARTICIPANT INFORMATION SHEET

Version 3 08-07-15

1. **Research project title**

Mainstreaming Assisted Living Technologies (MALT) project - eliciting stakeholders' requirements for an economic model

I would like to invite you to take part in a research study. It is important that you understand why the research is being conducted and what your involvement would mean. Please take a time to read the following information. Ask questions if you find anything not clear or would like more information. Take time to decide whether or not to take part.

2. **What is the project purpose?**

This research forms a part of MALT study that aims to understand barriers and facilitators of telehealth adoption. We would like to explore what stakeholders expect from an economic model.

3. **Why have I been chosen?**

You have been asked to take part because you are involved in delivering or managing telehealth services.

4. **Do I have to take part?**

Taking part in this study is entirely voluntary. You are free to withdraw at any time and without giving a reason.

5. What will happen to me if I take part?

If you agree to take part, you will be asked to take part in an interview that will be arranged in a time and place suitable for you. We will ask your permission to audiotape the interview. The aim of the interview is to gather your thoughts on the features and components of the cost-effectiveness model for telehealth.

6. What do I have to do?
You will be asked to participate in an interview, which will last up to 60 minutes.

7. What are the possible disadvantages and risks of taking part?

There are no risks associated with taking part in this research.

8. What are the possible benefits of taking part?

The information you give us will help us to develop a cost-effectiveness model for telehealth. While there are no immediate benefits for participants; it is hoped that you will find this work interesting.

9. What if something goes wrong?

Should you wish to raise a complaint, please contact Prof Simon Dixon. Should you feel, your complaint has not been handled to your satisfaction you can contact University's Registrar and Secretary.

Office of the Registrar and Secretary

Firth Court

Western Bank

Sheffield S10 2TN

Telephone: 0114 222 1100

10. Will my participation in this project be kept confidential?

All the data collected from the interviews will be kept strictly confidential. You will be identified by a code rather than a name and the information will be stored in a password protected files. The

interviews will be audiotaped. The recordings from the interviews will be destroyed once they are transcribed verbatim. Transcription will take place within a month of the interview. Transcripts will be deleted one year after the study ended.

11. What will happen to the results of the research project?

We aim to publish the results of this study in health economics journals and present the findings at professional conferences. Participants will not be identified in any report or publication.

12. Who is organising and funding the research?

This research is funded by Technology Strategy Board as part of Mainstreaming Assisted Living Technologies Project (MALT).

13. Who has ethically reviewed the project?

This project has been ethically reviewed by NHS Research Ethics Committee.

14. Contact for further information:

If you have a concern about any aspect of this study, please contact Kinga Lowrie, research student for more information.

Kinga Lowrie

School of Health and Related Research

University of Sheffield

Regent Court

Sheffield S1 4DA

Email: k.lowrie@sheffield.ac.uk

The project is supervised by Prof Simon Dixon and Prof Gail Mountain

Thank you for taking part in this project!

APPENDIX A.6.2 Qualitative interview participant consent form

MAINSTREAMING ASSISTED LIVING TECHNOLOGIES

Staff Consent Sheet One

Please initial the box for each statement

1. I confirm that I have read and understand the information sheet (version 4; 05/09/12) for the above study. I have had the opportunity to consider information, ask questions and have had these answered satisfactorily.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time without my legal rights being affected.	
3. I understand that the information I provide will be securely stored and that access will be restricted to the researchers working on this project.	
4. I understand that, as part of the study, audio recordings of my speech will be made and data about my role in the delivery of telehealth services may be recorded.	
5. I understand that data collected during the study may be looked at by individuals from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to this data.	

THE FOLLOWING STATEMENTS ARE OPTIONAL. PLEASE CIRCLE THE RELEVANT BOX

6. I agree that my anonymised data can be used in future research studies.	YES	NO
7. I agree that any information that I have provided can be used if I wish to withdraw from the study.	YES	NO
8. I agree that elements of my work, such as interaction with patients, can be observed, if relevant.	YES	NO
9. I agree that I wish to be informed about the study results.	YES	NO
10. I agree that I can be informed about workshops related to this study.	YES	NO
11. I agree that I can be informed about other studies related to telehealth.	YES	NO

--------------------------------- ---------------- -------------------------------
Name of participant   Date   Signature

--------------------------------- ---------------- -------------------------------
Name of person taking consent   Date   Signature

--------------------------------- ---------------- -------------------------------
Witness   Date   Signature

One signed copy to be kept by the participant and one copy by the researcher.

Staff Consent Sheet One V3.0 (5 September 2012)

APPENDIX A.7 The type of benefits identified in the qualitative interviews

Table 36: The type of benefits identified in the qualitative interviews

Type of benefits	Telehealth Providers	NHS Providers	Clinical Commissioning Groups
Quality of Service	*"Another measure is data-the system is able to effectively pick up exacerbations as fast as you won't be able in ordinary services"*	*"[..] quality within the primary care – so you see them at the points when they need to see them rather than having to go out routinely and when you not necessary need to see them." (site C)*	
Efficiency of Service	*"the fact that your one clinician can have a bigger caseload, because she travels less and she can deal with more patients is the most important efficiency benefit in the service"*	*"You want to see the change on utilization – changes in the way people access care, so it's dealt in community rather than in emergency." (site B)*	*"If we can do things slightly differently, we are dealing with more patients but actually from the same financial envelope." (site A)*
QoL Benefits for Patients	*" (...) we generally look at reduction in unplanned admissions, you should see*	*"I'd think we'd want to look for an improvement in self-management,*	*"People becoming more confident in managing their own condition, (..)that would contribute to reducing the*

	big changes there, if telehealth is working effectively, what you should see , even if you have been admitted, you should see significantly less days in hospital. Other important is medication adherence."	*recognition of exacerbation and then reduction in emergency admissions and then reduced length of stay."* (site A)	*number of emergency admissions."* (site B)

APPENDIX A.8.1 Usability evaluation protocol

<div align="center">

Usability evaluation Protocol

Version 2 08-09-14

</div>

1. Aim of "think aloud" tests:

- To gain an understanding of users experience with the software and to identify areas of model improvement

2. Procedure:

- Introduce yourself and people present in the room (if any)

- Introduce the aim of the study
- Explain that the test will be conducted in order to understand flaws in the design of model interface.
- Reassure that responses will be kept strictly confidential and anonymous.

    - Confirm that the data from the interview will be only used by the researcher

    - No individual will be identified in the findings.

- Remind about length of interview – 45-60 minutes

- Introduce video recorder and explain that videos will be reviewed informally and voice data transcribed. Explain that time will be measured. Explain that data will be stored securely and destroyed after publication of findings

- Check to see if the participant is willing for the researcher to use a video recorder

    - If yes, proceed

    - If no, researcher to make audio recordings. Explain that this form of recording is essential for this type of research

- Check if participant has any questions at all at this stage

- Ask participant to sign consent form to provide written consent for participation

- Thank the person for agreeing to participate

- Perform "Quick Start" guide of the features of the software and explain how to use it. Present the manual and inform participants that they are free to use it during the task if they want to.

- Explain that participant can practise "thinking aloud" while performing an example task

  - If yes, start the mock up test
  - If not proceed to the main task

3. Task solving
- Explain task 1 to the participant.

Researcher remains silent, only intervenes if participant stop talking or asks questions
- Explain task 2 to the participant.

4. Thank the participant for their time and contribution and offer the voucher to student participants

APPENDIX A.8.2 Copy of an email used in the usability testing recruitment process

A copy of an email send to The University of Sheffield students

Dear all,

I am post-graduate student in ScHARR looking for volunteers to take part in a testing of an online economic model interface.

This study requires you to complete two tasks using the software and verbalising your thoughts - "talking through" as you are doing it. The test comprises of two tasks where you will be asked to imagine that you are working for a certain Primary Care Trust and need to use the tool to evaluate what is the best option for your service. You will be given data to input into an online spreadsheet. We will not be assessing how you analyse the results. Our main focus is to observe the problems which might arise when people are using the tool. Help will be available for the parts of the spreadsheet you might not understand. The tests will be video recorded to ease further analysis. If you do not wish to be recorded we will use audio recorder. One of these forms of recording is essential for this study.

This study is open to all of you who have worked with Excel spreadsheets, or have any other data inputting experience.

It will only take up to 45 minutes of your time to complete the test. All participants will be paid £10 vouchers for their time. The interview will be held in the University premises and at a time and date that is convenient for you. We plan to conduct these tests on the first week of September.

This research has been ethically reviewed and approved by the ScHARR ethics committee and is supervised by Prof Simon Dixon.

If you require further information regarding this study, you can contact me through email address: k.lowrie@sheffield.ac.uk or mobile phone number 0744 XXX XXXX.

Thank you for considering taking part in this study.

Kinga Lowrie

APPENDIX A.8.3 Usability testing participant task sheet

Usability test of the MALT model

Task 1

Access **https://malt.demo1.epigenesys.org.uk**

Please register and Log in.

Imagine you are a Commissioning Manager who is looking at this site for the first time after hearing about it. Take less than 5 minutes to explore the site.

<u>Scenario 1</u>

Imagine you are a Commissioning Manager working for Doncaster PCT. You have an ageing population under your care and there is growing demand for health and social services in your area. You decided to use telehealth for some of your COPD patients. You are in the process of going through tenders and two companies have been shortlisted: **Nova health and Kitco.**

You need to decide which company would provide value for money service for your population.

Both sets of equipment provide the same options (monitoring of blood oxygen saturation and blood pressure). It has been proven that they have the same clinical effectiveness. You plan to provide telehealth for 9 months. Your population size is 1200 in severity group A, 3000 in severity group B, 150 in severity C and 100 in severity D. Equipment can be used by the next patient after 3 months.

Task 2 Run the model for Nova health

Nova health is offering you a 5 years contract. You would have to buy the equipment for £45 a set per month, and you would be charged £20 per month for each connected patient. They would charge £35 per patient on monthly basis for the maintenance: machine maintenance, battery and consumable replacements: (replacements of broken devices, hubs) etc. Additionally you would have to pay £80 for each machine installation, machine removal. The company would do the initial review of patient peripheral readings to determine if they fall outside of the parameters that have been set for that patient by the clinical case manager (technical triage). That would cost £3000 per month for all your patients. If you decide to use Nova health there will be a one training sessions for your staff on how to use the equipment which comes with the additional charge of £3000.

Once the technical triage team has confirmed that the reading for a patient has fallen outside of the patient's usual parameters they will contact the patient's case manager who will deliver a clinical response (triage). This will aim to reduce the number of condition exacerbations and

associated emergency hospital admissions. As part of their jobs Matrons will contact the patient either via telephone or visit them at home. Each nurse home visit cost your Trust £150. The rest of the hospital costs remain the same as the national average. Doncaster PCT is a bearer of all your hospital and GP costs.

Task 3 Run the model for Kitco

Kitco is also offering the contract for 5 years. You have an option of leasing the equipment, which would be covered by a £65 per patient per month communication fee. This however, also includes equipment installation and removal in patients' homes. Additionally, you have to pay a total of £3000 for staff training. There is a one-off charge of £4500 for having the service reviewed.

APPENDIX A.8.4 Actions required to successfully complete task 1 and task 2

Number	Type of action	Steps	Possible errors
Task 1	Log in and Register	1. Select register category	Selecting sign in first
		2. Fill in the fields	Not assessed
		3. Go to personal email to confirm the account	Not assessed
		4. Go back to model to sign up	Not assessed
		5. Sign up using name and chosen password	Not assessed

Number	Type of action	Steps	Possible errors
Task 2	Scenario 1 Run the model for Nova health	1. Click run the new model	Problems with locating the button
	Navigation	2. Select regular model type from listed categories	Selecting Light or Advanced version
	Navigation/ Content	3. Select the correct PCT	Selecting other Trusts
	Navigation	4. Select disease type as COPD	Selecting CHF
	Click	5. Press 'Yes' for allocation to most ill (optional)	Pressing 'no' or leaving it blank

Number	Type of action	Steps	Possible errors
	Data input	6. Enter 60 as duration of model run	Any other value
		7. Enter 9 months as duration	Any other value
		8. Enter 3 as redeployment delay	Any other value
		9. Enter the size of population: 1200 in severe, 3000 in moderate and 150 in very severe state.	Any other value

Number	Type of action	Steps	Possible errors
		10. Enter the number of devices deployed each month or leave it as displayed (optional)	Any other value
		11. Type in Nova Health and specify it as supplier	Not specifying all suppliers
		12. Type in NHS and specify it as payer and corresponding NHS as a supplier*	This action was intuitive, not counted as error if omitted

Number	Type of action	Steps	Possible errors
		13. Select supplier and payer for all healthcare costs	Problems with locating appropriate buttons
		14. Enter 45 in device cost as 'variable cost'	Choosing different cost category
		15. Enter 20 in communication/ hosting as 'variable cost'	Choosing different cost category
		16. Enter 35 in maintenance as 'fixed cost'	Choosing different cost category

Number	Type of action	Steps	Possible errors
		17. Enter 80 in installation category as 'one-off cost'	Choosing different cost category
		18. Enter 80 in removal category as 'one – off cost'	Choosing different cost category
		19. Enter 3000 in technical triage category as 'one-off' cost	Choosing different cost category
		20. Enter 3000 in training and set up category as 'one-off cost'	Choosing different cost category

Number	Type of action	Steps	Possible errors
		21. Go up the screen, back to healthcare costs and enter 150 in nurse home visit category	Choosing different cost category
	Selecting from a drop - down menu	22. Choose supplier/payer	
	Click	23. Click 'run the model'	Problems with locating the button
	Data input	24. Fill in remaining fields with zeros if necessary (optional)	Not assessed

Number	Type of action	Steps	Possible errors
Task 3	Run the model for Kitco	1. Choose run the model from the same parameters if not follow steps from Task 2 (3-10)	Not assessed
	Click/ delete manually	2. Clear telemonitoring costs by pressing the button or manually	Not assessed
	Data input	3. Enter 65 in communication category as 'variable cost'	Choosing different cost category

Number	Type of action	Steps	Possible errors
		4. Enter 3000 in training category as 'one-off'	Choosing different cost category
		5. Enter 4500 in service review category as 'one-off'	Choosing different cost category
	Click	6. Click 'run the model'	Not assessed
	Navigation	7. Go to results panel	Not assessed

Number	Type of action	Steps	Possible errors
	Navigation/ Analysis	8. Find the table indicating total costs and assess which option (Nova health or Kitco) were cheaper	Not assessed

APPENDIX A.8.5 Usability study information sheet 1

STUDENT INFORMATION SHEET

Version 2 09-07-14

1. **Research project title**

Assessing the usability of the MALT (Mainstreaming Assisted Living Technologies project) economic model

I would like to invite you to take part in a research study. Please read the following information to decide if you would like to be involved. Please ask questions if anything is not clear or would like more information. Take time to decide whether or not to take part.

2. **What is the project purpose?**

This research forms a part of a PhD study which aims to understand how to enhance the usefulness of economic models to decision makers. The interviews are part of the qualitative workstream which seeks to explore usability issues in an online version of an economic model designed for health commissioners to use. The feedback from this study will be used by software developers to improve the model.

3. **Why have I been chosen?**

You have been asked to take part because you have experience working with Excel spreadsheets and data inputting. I aim to interview up to 5 students within the University of Sheffield and 5 professionals.

4. Do I have to take part?

Taking part in this study is entirely voluntary. You are free to withdraw at any time and without giving a reason. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. You will be rewarded £5high street voucher for your time for taking part in this study.

5. What will happen to me if I take part?

If you agree to take part, you will be approached by the researcher in order to arrange the interview in a time convenient for you on University premises. The interview will last no more than 60min. The test comprises of two tasks where you will be asked to imagine that you are working for a certain Primary Care Trust and need to use the tool to evaluate what is the best option for your service. You will be given data to input into an online spreadsheet. We will not be assessing how you analyse the results. Our main focus is to observe the problems which might arise when people are using the tool. Help will be available for the parts of the spreadsheet you might not understand. The interviews will be video recorded to ease further analysis. If you do not wish to be recorded, we will use audio recorder. One of these forms of recording is essential for this study.

6. What do I have to do?
You will be given a tutorial on how to use the software and asked to solve two tasks. You will be asked to verbalize your thoughts – "think aloud" as you solve these tasks while being video/audio recorded. There will be an opportunity to practise "thinking aloud" in a mock up test.

7. What are the possible disadvantages and risks of taking part?

There are no risks associated with taking part in this research.

8. What are the possible benefits of taking part?

Your involvement will help us to develop a model that is usable and therefore useful. To our knowledge this is the first of this type of test to be conducted in the field of modelling. Whilst there are no immediate benefits for participants; it is hoped that you will find this work interesting.

9. What if something goes wrong?

Should you wish to raise a complaint please contact Prof Simon Dixon. Should you feel your complaint has not been handled to your satisfaction you can contact University's Registrar and Secretary.

Office of the Registrar and Secretary

Firth Court

Western Bank

Sheffield S10 2TN

Telephone: 0114 222 1100

10. Will my taking part in this project be kept confidential?

All the data collected from the interviews will be kept strictly confidential. You will be identified by a code rather than a name and the information will be stored in a password protected files.

11. Will I be recorded and how will the recorded media be used?

The video recordings from the interviews will be reviewed by the research team and if necessary by software developers in order to study non-verbalised problems of usage, and the issues users struggled with. Voice data will be transcribed and analysed thematically. This is to develop recommendations for software developers so that the tool can be easily used. Both video and transcripts will be destroyed one year after the study ended.

12. What will happen to the results of the research project?

I aim to publish the results of this study in health economics journals and present the findings at professional conferences. Participants will not be identified in any report or publication.

13. Who is organising and funding the research?

This research is funded by Technology Strategy Board as part of Mainstreaming Assisted Living Technologies Project (MALT).

14. Who has ethically reviewed the project?

This project has been ethically reviewed by School of Health and Related Research Ethics Committee.

15. Contact for further information:

If you have a concern about any aspect of this study, please contact Kinga Lowrie, research student for more information.

Kinga Lowrie

School of Health and Related Research

University of Sheffield

Regent Court

Sheffield S1 4DA

Email: k.lowrie@sheffield.ac.uk

The project is supervised by Prof Simon Dixon and Prof Gail Mountain

Thank you for taking part in this project!

APPENDIX A.8.6 Usability study information sheet 2

<div align="center">

PROFESSIONAL INFORMATION SHEET

Version 2 09-07-14

</div>

1. **Research project title**

Assessing the usability of the MALT (Mainstreaming Assisted Living Technologies project) economic model

> *I would like to invite you to take part in a research study. Please read the following information that will help you to decide if you would like to be involved. Do not hesitate to ask questions if anything you read is not clear or would like more information. Take time to decide whether or not to take part.*

2. **What is the project purpose?**

This research forms a part of a PhD study which aims to understand how to enhance the usefulness of economic models to decision makers. The interviews are part of the qualitative workstream which seeks to explore usability issues in an online version of an economic model designed for health commissioners to use. The feedback from this study will be used by software developers to improve the model.

3. **Why have I been chosen?**

You have been asked to take part because you have experience working in telehealth industry or healthcare environment. I aim to interview up to 5 students within the University of Sheffield and 5 professionals.

4. Do I have to take part?

Taking part in this study is entirely voluntary. You are free to withdraw at any time and without giving a reason. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form.

5. What will happen to me if I take part?

If you agree to take part, you will be approached by the researcher in order to arrange the interview in a time and place convenient for you. The test will last no more than 60min. The test comprises of two tasks where you will be asked to imagine that you are working for a certain Primary Care Trust and need to use the tool to evaluate what is the best option for your service. You will be given data to input into an online spreadsheet. We will not be assessing how you analyse the results. Our main focus is to observe the problems which might arise when people are using the tool. Help will be available for the parts of the spreadsheet you might not understand. The tests will be video recorded to ease further analysis. If you do not wish to be recorded, we will use audio recorder. One of these forms of recording is essential for this study.

6. What do I have to do?

You will be given a tutorial on how to use the software and asked to solve two tasks. You will be asked to verbalize your thoughts – "think aloud" as you solve these tasks while being video/audio recorded. There will be an opportunity to practise "thinking aloud" in a mock up test.

7. What are the possible disadvantages and risks of taking part?

There are no risks associated with taking part in this research.

8. What are the possible benefits of taking part?

Your involvement will help us to develop a model that is usable and therefore useful. To our knowledge this is the first of this type of test to be conducted in the field of modelling. Whilst there are no immediate benefits for participants; it is hoped that you will find this work interesting.

9. What if something goes wrong?

Should you wish to raise a complaint please contact Prof Simon Dixon. Should you feel your complaint has not been handled to your satisfaction you can contact University's Registrar and Secretary.

Office of the Registrar and Secretary

Firth Court

Western Bank

Sheffield S10 2TN

Telephone: 0114 222 1100

10. Will my taking part in this project be kept confidential?

All the data collected from the interviews will be kept strictly confidential. You will be identified by a code rather than a name and the information will be stored in a password protected files.

11. Will I be recorded and how will the recorded media be used?

The video recordings from the interviews will be reviewed by the research team and if necessary by software developers in order to study non-verbalised problems of usage, and the issues users struggled with. Voice data will be transcribed and analysed thematically. This is to develop recommendations for software developers so that the tool can be easily used. Both video and transcripts will be destroyed one year after the study ended.

12. What will happen to the results of the research project?

I aim to publish the results of this study in health economics journals and present the findings at professional conferences. Participants will not be identified in any report or publication.

13. Who is organising and funding the research?

This research is funded by Technology Strategy Board as part of Mainstreaming Assisted Living Technologies Project (MALT).

14. Who has ethically reviewed the project?

This project has been ethically reviewed by School of Health and Related Research Ethics Committee.

15. Contact for further information:

If you have a concern about any aspect of this study, please contact Kinga Lowrie, research student for more information.

Kinga Lowrie

School of Health and Related Research

University of Sheffield

Regent Court

Sheffield S1 4DA

Email: k.lowrie@sheffield.ac.uk

The project is supervised by Prof Simon Dixon and Prof Gail Mountain

Thank you for taking part in this project!

APPENDIX A.8.7 Usability evaluation consent form

CONSENT FORM

Assessing the usability of the Mainstreaming Assistive Living Technologies project (MALT) economic model

Participant Identification Number for this project:

I confirm that I have read and understand the information sheet dated 09/07/2014 explaining the above research project and I have had the opportunity to ask questions about the project.

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline.

I understand that my responses will be kept strictly confidential. **I give permission to be video recorded** and for members of the research team to have access to my recordings and audio data. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.

I do not agree to be video recorded but I give permission for the researchers to record me "thinking aloud" while solving the tasks. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.

5. I agree to take part in the above research project.

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature

APPENDIX A.8.8 Ethical approval for usability evaluation

The
University
Of
Sheffield.

Kirsty Woodhead
Ethics Committee Administrator

Regent Court
30 Regent Street
Sheffield S1 4DA
Telephone: +44 (0) 114 2225453
Fax: +44 (0) 114 272 4095 (non confidential)
Email: k.woodhead@sheffield.ac.uk

Our ref: 0760/KW

22 July 2014

Kinga Lowrie
ScHARR

Dear Kinga

Assessing the usability of the MALT economic model

Thank you for submitting the above research project for approval by the ScHARR Research
Ethics Committee. On behalf of the University Chair of Ethics who reviewed your project, I am
pleased to inform you that on 22 July 2014 the project was approved on ethics grounds, on the
basis that you will adhere to the documents that you submitted for ethics review.

The research must be conducted within the requirements of the hosting/employing
organisation or the organisation where the research is being undertaken. You are also required
to ensure that you meet any research ethics and governance requirements in the country in
which you are researching. It is your responsibility to find out what these are.

If during the course of the project you need to deviate significantly from the documents you
submitted for review, please inform me since written approval will be required. Please also
inform me should you decide to terminate the project prematurely.

Yours sincerely

Kirsty Woodhead
Ethics Committee Administrator

APPENDIX A.9.1 Ethical approval for online questionnaire study

Kinga Sokolowska-Lowrie
Registration number: 100264106
School of Health and Related Research
Programme: PhD

Dear Kinga

PROJECT TITLE: An online survey to assess MALT model user satisfaction
APPLICATION: Reference Number 002277

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 16/12/2014 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 002277 (dated 15/12/2014).
- Participant information sheet 003620 version 1 (24/11/2014).
- Participant consent form 003621 version 1 (24/11/2014).

The following optional amendments were suggested:

Section B3: An URMS number is required as this is 'health services' research. Section D2: Please confirm that you are not asking for personal details to be forwarded to you. Section E1: It states on the information sheet that participants can withdraw at any point. Can they withdraw after the survey has been submitted? If not this needs to be made clear. External Documentation: Will participants know what the MALT study is? Does this need any further elaboration? (it's fine if they know about what you are referring to). Please would you update the complaints process: Simon in the first instance and Jon in the second (if the complaint isn't handled properly).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Jane Spooner
Ethics Administrator
School of Health and Related Research

APPENDIX A.9.2 EUCS responses

RESPONDENT 1: overall EUCS score 31

Does the model provide precise information you need?

Some of the time

Does the information content meet your needs?

Some of the time

Does the model provide sufficient information?

Some of the time

Is the model accurate?

Some of the time

Are you satisfied with the accuracy of the model?

Most of the time

Do you think the output is presented in a useful format?

About half of the time

Is the information clear?

Some of the time

Is the model user friendly?

Most of the time

Is the model easy to use?

Most of the time

Do you get the information you need in time?

About half of the time

Does the model provide up-to-date information?

About half of the time

RESPONDENT 2: overall EUCS score 44

Does the model provide precise information you need?

Most of the time

Does the information content meet your needs?

Most of the time

Does the model provide sufficient information?

Most of the time

Is the model accurate?

Most of the time

Are you satisfied with the accuracy of the model?

Most of the time

Do you think the output is presented in a useful format?

Most of the time

Is the information clear?

Most of the time

Is the model user friendly?

Most of the time

Is the model easy to use?

Most of the time

Do you get the information you need in time?

Most of the time

Does the model provide up-to-date information?

Most of the time

APPENDIX A.9.3 The MALT model feedback interview topic guide

TOPIC GUIDE
MALT model feedback
Version 1 10-06-15

1. Aim of interviews:

- To gather users' feedback on MALT model usefulness.

2. Introduction:

- Introduce yourself

- Introduce the aim of the study
- Explain that the interviews will be conducted with people who registered to use the model
- Reassure that data will be kept strictly confidential and anonymous.

    - Confirm that the data from the interview will be only used by the researcher

    - No individual will be identified in the findings.

- Remind about length of interview – max 30 minutes

- Introduce tape recorder and explain transcription, data storage and destruction (post publication of findings)

- Ask for the permission to use tape recorder

    - If yes, proceed

    - If no, researcher to record notes manually and explain that the written notes are to be used as an aid for the researcher

- Check if participant has any questions at all at this stage

- Ask participant to sign consent form to provide written consent for participation

- Thank the person for agreeing to participate

3. I would like to start by asking, how did you hear about MALT model?

4. Having heard of it, what motivated you to use it?

5. Can you tell me what your opinion about the models' features is? (e.g. options it provides)?

6. Concluding questions

- Were you satisfied with using the model?

- On the scale 1 to 4, were 4 means very satisfied, and 1 means dissatisfied, (very satisfied, satisfied, nor-satisfied-or dissatisfied, dissatisfied) how would you rate your overall satisfaction with the tool?

- What is your opinion about model's usefulness from your perspective?
- Do you have any other comments about the model?

Thank the participant for their time and contribution

APPENDIX A.9.4 Ethical approval for model feedback interviews

Ellen Nicolson

Ethics Committee Administrator

Regent Court

Telephone: +44 (0) 114 222 5446

Project title: Assessment of the MALT model user satisfaction

Reference Number: 0760

Dear Kinga,

Thank you for submitting the above amended research project for approval by the ScHARR Research Ethics Committee. On behalf of the University, I am pleased to inform you that the project with changes was approved. If during the course of the project you need to deviate significantly from the documents you submitted for review, please inform me since written approval will be required.

Yours sincerely,

Ellen Nicolson

On behalf of the ScHARR Research Ethics Committee

Literature

Abdinnour-Helm, S. F., Chaparro, B. S., & Farmer, S. M. (2005). Using the End-User Computing Satisfaction (EUCS) Instrument to Measure Satisfaction with a Web Site. *Decision Sciences, 36*(2), 341-364.

Adeogun, O., Tiwari, A., & Alcock, J. R. (2011a). Capabilities of Proprietary Intermediate Telehealth Devices *Telemedicine and e-Health, 17*(9), 712-721.

Adeogun, O., Tiwari, A., & Alcock, J. R. (2011b). Models of information exchange for UK telehealth systems. *Int J Med Inform., 80*(5), 359-370.

Aggelidis, V. P., & Chatzoglou, P. D. (2012). Hospital information systems: Measuring end user computing satisfaction (EUCS). *Journal of Biomedical Informatics, 45*(3), 566-579.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. NJ: Prentice-Hall.

Akpan, I. J., & Brooks, R. J. (2014). Experimental evaluation of user performance on two-dimensional and three-dimensional perspective displays in discrete-event simulation. *Decision Support Systems, 64*, 14-30.

Allen, M., Currie, L., M.,, Bakken, S., Patel, V. L., & Cimino, J. J. (2006). Heuristic evaluation of paper-based Web pages: a simplified inspection usability methodology. *Journal of biomedical informatics, 39*(4), 412-423.

Alter, S. (1977). A taxonomy of decision support systems. *Sloan Management Review, 19*(1), 39-56.

Andersson, F., Borg, S., Jansson, S. A., Jonsson, A. C., Ericsson, A., Prutz, C., . . . Lundback, B. (2002). The costs of exacerbations in chronic obstructive pulmonary disease (COPD). *Respiratory Medicine, 96*(9), 700-708.

Andree, R. V. (1988). Computer-Assisted Problem Solving *A Computer Science Reader* (pp. 317-337): Springer.

Anthonisen, N. R., Connett, J.E., Kiley, J.P. (1994). Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA, 272*, 1497-1505.

Audit, C. f. H., & Inspection. (2006). *Clearing the Air: A National Study of Chronic Obstructive Pulmonary Disease*: Commission for Healthcare Audit and Inspection.

Avouris, N., Tselios, N., Fidas, C., & Papachristos, E. (2003). Website evaluation: A usability-based perspective *Advances in Informatics* (pp. 217-231): Springer.

Azjen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl, Beckmann, J., (Ed.), *Action control: From cognition to behavior.* (pp. 11-39). Berlin, Heidelber,.

Azjen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes 50*(2), 179-211.

Bach, P. B., Brown, C., Gelfand, S. E., & McCrory, D. C. (2001). Management of acute exacerbations of chronic obstructive pulmonary disease: a summary and appraisal of published evidence. *Annals of Internal Medicine, 134*(7), 600-620.

Bagozzi, R. P. (2007). The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. *Journal of the Association for Information Systems, 8*(4), 3.

Bailey, J. E., & Pearson, S. W. (1983). Development of a Tool for Measuring and Analyzing Computer User Satisfaction. *Management Science, 29*(5), 530-545.

Bairamzadeh, S., & Bolhari, A. (2010). *Investigating factors affecting students' satisfaction of university websites.* Paper presented at the 3rd IEEE International Conference on Computer Science and Information Technology.

Balci, O. (1994). Principles of simulation model validation, verification, and testing.

Balci, O., Nance, R.E., Ormsby, W.F.,. (2002). *Improving the model development process: expanding our horizons in verification, validation, and accreditation research and practice.* Paper presented at the Proceedings of the 34th conference on Winter simulation: exploring new frontiers.

Barbour, R. S., & Schostak, J. (2005). Interviewing and focus groups. *Research methods in the social sciences, 1*, 41-48.

Barlow, J., Hendy, J., & Chrysanthaki, T. (2012). Scaling-up remote care in the United Kingdom. Lessons from a decade of policy intervention. In A. P. from Glascock, Kutzik, D.M., (Ed.), *Essential lessons for the success of telephone: why it's not plug and play* (Vol. 30): Ios Press.

Bartoli, L., Zanaboni, P., Masella, C., & Ursini, N. (2009). Systematic Review of Telemedicine Services for Patients Affected by Chronic Obstructive Pulmonary Disease (COPD). *Telemedicine Journal and E-health, 15*(9), 877-883.

Bastien, J. M. C. (2010). Usability testing: a review of some methodological and technical aspects of the method. *International journal of medical informatics, 79*(4), e18-e23.

Bastin, L., Cornford, D., Jones, R., Heuvelink, G. B. M., Pebesma, E., Stasch, C., . . . Williams, M. (2013). Managing uncertainty in integrated environmental modelling: The UncertWeb framework. *Environmental Modelling & Software, 39*, 116-134.

Battleson, B., Booth, A., & Weintrop, J. (2001). Usability testing of an academic library web site: a case study. *The Journal of Academic Librarianship, 27*(3), 188-198.

Bell, P. C., & O'Keefe, R. M. (1995). An experimental investigation into the efficacy of visual interactive simulation. *Management Science, 41*(6), 1018-1038.

Benbasat, I., Barki, H.,. (2007). Quo vadis TAM? *Journal of the association for information systems, 8*(4), 7.

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management, 39*(2), 151-163.

Bergmo, T. (2009). Can economic evaluation in telemedicine be trusted? A systematic review of the literature. *Cost Effectiveness and Resource Allocation, 7*(1), 18.

Bergmo, T., S.,. (2015). An Economic Model of Remote Specialist Consultations using Videoconferencing. *International Journal on Advances in Life Sciences Volume 7, Number 1 & 2, 2015*.

Bergmo, T. S. (2014). Using QALYs in telehealth evaluations: a systematic review of methodology and transparency. *BMC Health Services Research, 14*, 332.

Bevan, N. (1995). Usability is quality of use. *Advances in Human Factors Ergonomics, 20*, 349-349.

Blackmon, M. H., Kitajima, M., & Polson, P. G. (2003). *Repairing usability problems identified by the cognitive walkthrough for the web.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Booker, J., & Bent, A. (2004). Chronic obstructive pulmonary disease: national clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. *Thorax, 59*(1), 1-232.

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *Professional Communication, IEEE Transactions on, 43*(3), 261-278.

Borg, S., Ericsson, A., Wedzicha, J., Gulsvik, A., Lundbäck, B., Donaldson, G. C., & Sullivan, S. D. (2004). A Computer Simulation Model of the Natural History and Economic Impact of Chronic Obstructive Pulmonary Disease. *Value in health, 7*(2), 153-167.

Bowers, J., Ghattas, M., & Mould, G. (2012). Exploring alternative routes to realising the benefits of simulation in healthcare. *Journal of the Operational Research Society, 63*(10), 1457-1466.

Bradbury, H., & Reason, P. (2003). Action research: An opportunity for revitalizing research purpose and practices. *Qualitative social work, 2*(2), 155-175.

Brailsford, S., C.,. (2005). Overcoming the barriers to implementation of operations research simulation models in healthcare. *Clinical and Investigative Medicine, 28*(6), 312.

Brailsford, S., C.,. (2007). *Advances and challenges in healthcare simulation modeling: tutorial.* Paper presented at the Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come.

Brailsford, S. C., Bolt, T., Connell, C., Klein, J. H., & Patel, B. (2009). *Stakeholder engagement in health care simulation.* Paper presented at the winter simulation conference.

Brailsford, S. C., Bolt, T. B., Bucci, G., Chaussalet, T. M., Connell, N. A., Harper, P. R., . . . Taylor, M. (2011). Overcoming the barriers: a qualitative study of simulation adoption in the NHS. *J Oper Res Soc*.

Brailsford, S. C., Harper, P., Patel, B., & Pitt, M. (2016). An Analysis of the Academic Literature on Simulation and Modeling in Health Care *Operational Research for Emergency Planning in Healthcare: Volume 2* (pp. 231-251): Springer.

Brailsford, S. C., & Hilton, N. A. (2001). A comparison of discrete event simulation and system dynamics for modelling health care systems.

Brailsford, S. C., Lattimer, V. A., Tarnaras, P., & Turnbull, J. C. (2004). Emergency and on-demand health care: modelling a large complex system. *Journal of the Operational Research Society, 55*(1), 34-42.

Brennan, A., & Akehurst, R. (2000). Modelling in health economic evaluation. What is its place? What is its value? *Pharmacoeconomics, 17*(5), 445-459.

Brennan, A., Chick, S. E., & Davies, R. (2006). A taxonomy of the model structures for economic evaluations of health technologies. *Health Economics, 15*, 1295-1310.

Brettle, A. J., Brown, T.M.,Hardiker, N.R.,Radcliffe, J.N.,, & Smith, C. L. (2013). Telehealth: The effects on clinical outcomes, cost effectiveness and the patient experience: a systematic overview of the literature.

Brewster, L., Mountain, G., Wessels, B., Kelly, C., & Hawley, M. (2014). Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. *Journal of advanced nursing, 70*(1), 21-33.

Briggs, A., & Sculpher, M. (1998). An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics, 13*(4), 397-409.

Briggs, A. H., Claxton, K., & Sculpher, M. J. (2006). *Decision modelling for health economic evaluation*: Handbooks in Health Economic E.

Broens, T., Vollenbroek-Hutten, M., Hermens, H., van Halteren, A., & Nieuwenhuis, L. (2007). Determinants of successful telemedicine implementations: a literature study. *Journal of telemedicine and telecare, 13*(6), 303-309.

Bryan, S., Williams, I., & McIver, S. (2007). Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. *Health economics, 16*(2), 179-193.

Bryman, A. (2008). *Social reserach methods*: Oxford University Press.

Buist, A. S., Vollmer, W. M., Sullivan, S. D., Weiss, K. B., Lee, T. A., Menezes, A. M. B., . . . Burney, P. G. J. (2005). The Burden Of Obstructive Lund Disease Initiative (BOLD): Rationale and Design. *Journal of Chronic Obstructive Pulmonary Disease, 2*, 277-283.

Buxton, M., J.,. (2006). Economic evaluation and decision making in the UK. *Pharmacoeconomics, 24*(11), 1133-1142.

Buxton, M. J., Drummond, M. F., Van Hout, B. A., Prince, R. L., Sheldon, T. A., Szucs, T., & Vray, M. (1997). Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life. *Health Economics, 6*(3), 217-227.

Calvo, G. S., Gómez-Suárez, C., Soriano, J. B., Zamora, E., Gónzalez-Gamarra, A., González-Béjar, M., . . . Fernández, G. (2014). A home telehealth program for patients with severe COPD: the PROMETE study. *Respiratory medicine, 108*(3), 453-462.

Campbell, M., Egan, M.,Lorenc, T.,Bond, L.,Popham, F.,Fenton, C.,Benzeval, M.,. (2014). Considering methodological options for reviews of theory: illustrated by a review of theories linking income and health. *Systematic reviews, 3*(1), 1-11.

Car, J., Huckvale, K., & Hermens, H. (2012). Telehealth for long term conditions. *BMJ: British Medical Journal, 344*.

Caro, J. J., & Möller, J. (2016). Advantages and disadvantages of discrete-event simulation for health economic analyses: Taylor & Francis.

Caro, J. J., Möller, J., & Getsios, D. (2010). Discrete event simulation: the preferred technique for health economic evaluations? *Value in health, 13*(8), 1056-1060.

Caro, J. J., Möller, J., Karnon, J., Stahl, J., & Ishak, J. (2015). *Discrete event simulation for health technology assessment*: CRC Press.

Carroll, C., Marsden, P., Soden, P., Naylor, E., New, J., & Dornan, T. (2002). Involving users in the design and usability evaluation of a clinical decision support system. *Computer methods and programs in biomedicine, 69*(2), 123-135.

Castilla, D., Garcia-Palacios, A., Bretón-López, J., Miralles, I., Baños, R. M., Etchemendy, E., . . . Botella, C. (2013). Process of design and usability evaluation of a telepsychology web and virtual reality system for the elderly: Butler. *International Journal of Human-Computer Studies, 71*(3), 350-362.

Celli, B. R., Cote, C. G., Marin, J. M., Casanova, C., Montes de Oca, M., Mendez, R. A., . . . Cabral, H. J. (2004). The Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity Index in Chronic Obstructive Pulmonary Disease. *New England Journal of Medicine, 350*(10), 1005-1012. doi:doi:10.1056/NEJMoa021322

Celli, B. R., Thomas, N. E., Anderson, J. A., Ferguson, G. T., Jenkins, C. R., Jones, P. W., . . . Calverley, P. M. A. (2008). Effect of pharmacotherapy on rate of decline of lung function in chronic obstructive pulmonary disease: results from the TORCH study. *American journal of respiratory and critical care medicine, 178*(4), 332-338.

Chen, S. Y., & Macredie, R. D. (2005). The assessment of usability of electronic shopping: A heuristic evaluation. *International journal of information management, 25*(6), 516-532.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). *Development of an instrument measuring user satisfaction of the human-computer interface.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Chisman, J., Diller, K., & Walbridge, S. (1999). Usability testing: a case study. *College & research libraries, 60*(6), 552-569.

Christ-Neumann, M. L., Escrich, A., Anguita, A., Stenzhorn, H., Taylor, M., Ramay, H., . . . Graf, N. (2014). Usability on the p-medicine infrastructure: an extended usability concept. *ecancermedicalscience, 8*.

Chuck, A., Jacobs, P., Mayers, I., & Marciniuk, D. (2008). Cost-effectiveness of combination therapy for chronic obstructive pulmonary disease. *Canadian respiratory journal: journal of the Canadian Thoracic Society, 15*(8), 437.

Clark, M., & Goodwin, N. (2010). *Sustaining innovation in telehealth and telecare*. Retrieved from

Clarke, A., Taylor-Phillips, S., Swan, J., Gkeredakis, E., Mills, P., Powell, J., . . . Grove, A. (2013). Evidence-based commissioning in the English NHS: who uses which sources of evidence? A survey 2010/2011. *BMJ Open, 3*(5).

Claxton, K., Sculpher, M., & McCabe, C. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics, 14*, 339-347.

Coates, L., Nicholson, J., & Hawley, M. (2013). Overcoming the barriers to Mainstreaming Assisted Living Technologies (MALT): Methodological and practical issues in observing and effecting change in telehealth services. *International Journal of Integrated Care*(7).

Coghlan, D., & Casey, M. (2001). Action research from the inside: issues and challenges in doing action research in your own hospital. *Journal of advanced nursing, 35*(5), 674-682.

Connelly, L. G., & Bair, A. E. (2004). Discrete event simulation of emergency department activity: A platform for system-level operations research. *Academic Emergency Medicine, 11*(11), 1177-1185.

Constantine, L. L., & Lockwood, L. A. (1999). *Software for use: a practical guide to the models and methods of usage-centered design*: Pearson Education.

Cooke, L. (2010). Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication 53*(3), 202-215.

Coughlan, P., & Coghlan, D. (2002). Action research for operations management. *International journal of operations & production management, 22*(2), 220-240.

Courtney, J. F. (2001). Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS. *Decision Support Systems, 31*(1), 17-38.

Cruickshank, J. (2012). *What can the NHS learn from experience at the US Veterans Health Administration?* Retrieved from www.2020health.org/dms/2020health/downloads/reports/Telehealth-VA/ on the 01/07/2015

Cunliffe, D., Kritou, E., & Tudhope, D. (2001). Usability evaluation for museum web sites. *Museum Management and Curatorship, 19*(3), 229-252.

D'Souza, A. O., Smith,M.J.,Miller, L.A.,Kavookjian, J. (2006). An Appraisal of Pharmacoeconomic Evidence of Maintenance Therapy for COPD*. *Chest, 129*(6), 1693-1708. doi:10.1378/chest.129.6.1693

Dal Negro, R., Eandi, M., Prandelli, L., & Iannazzo, S. (2007). Cost-effectiveness and healthcare budget impact in Italy of inhaled corticosteroids and brochodilators for severe and very severe COPD patients. *International Journal of Chronic Obstructive Pulmonary Disease, 2*(2), 169-176.

Dansky, K. H., Vasey, J., Bowles, K.,. (2008). Impact of telehealth on clinical outcomes in patients with heart failure. *Clinical nursing research, 17*(3), 182-199.

Darkins, A., Kendall, S., Edmonson, E., Young, M., & Stressel, P. (2015). Reduced Cost and Mortality Using Home Telehealth to Promote Self-Management of Complex Chronic Conditions: A Retrospective Matched Cohort Study of 4,999 Veteran Patients. *Telemedicine and e-Health, 21*(1), 70-76. doi:10.1089/tmj.2014.0067

Darkins, A., Ryan, P., Kobb, R., Foster, L., Edmonson, E., Wakefield, B., & Lancaster, A. E. (2008). Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. *Telemedicine Journal & e-Health, 14*(10), 1118-1126.

Darzi, A. (2009). High quality care for all: NHS Next Stage Review final report, 2008. *Department of Health, London*.

Davalos, M. E., French, M. T., Burdick, A. E., & Simmons, S. C. (2009). Economic evaluation of telemedicine: review of the literature and research guidelines for benefit-cost analysis. *Telemedicine Journal & e-Health, 15*(10), 933-948.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3), 319-340.

Della Mea, V. (2001). What is e-health (2): the death of telemedicine? *Journal of Medical Internet Research, 3*(2).

DeLone, W., H.,, & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research 3*, 60-95.

DeLone, W., H.,, & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems, 19*(4), 9-30.

DeLone, W. H., & McLean, E. R. (2002). *Information systems success revisited.* Paper presented at the Proceedings of the 35th Annual Hawaii International Conference on Information Systems.

Denscombe, M. (2002). *Ground rules for good research*: Open University Press.

Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research*: Sage publications, inc.

Denzin, N. K., Lincoln, Y. S.,. (2003). *Strategies of Qualitative Inquiry.*: Sage.

Department of Health. (1997). The new NHS: modern, dependable: HMSO London.

Department of Health. (2002). Delivering 21st Century IT Support for the NHS. National Strategic Programme: Department of Health London, UK.

Department of Health. (2010). Building the National Care Service.

Department of Health. (2011). *Whole System Demonstrator Programme headline Findings*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215264/dh_131689.pdf:

Department of Health. (1998). *Information for Health: An Information Strategy for Modern NHS*. Retrieved from London:

Department of Health. (2011). NHS Trusts and PCTs combined reference cost schedules. Retrieved from

Dixon, P., Hollinghurst, S., Ara, R., Edwards, L., Foster, A., & Salisbury, C. (2016). Cost-effectiveness modelling of telehealth for patients with raised cardiovascular disease risk: evidence from a cohort simulation conducted alongside the Healthlines randomised controlled trial. *BMJ open, 6*(9), e012355.

Dodd, P. J., Thokala, P., Nicholson, J. T., Coates, E., Brennan, A., & Dixon, S. (2014). *MALT financial modelling of telemonitoring (internal report)*. Retrieved from

Doll, W. J., Deng, X., Raghaunathan, G. T., & Weidong, X. (2004). The meaning and measurement of user satisfaction: A multigroup invariance analysis of the end-user computing satisfaction instrument. *Journal of Management Information Systems, 21*(1), 227-262.

Doll, W. J., & Torkzadeh, G. (1988). The Measurement of End-User Computing Satisfaction. *MIS Quarterly, 12*(2), 259-274.

Drabble, L., Trocki, K. F., Salcedo, B., Walker, P. C., & Korcha, R. A. (2016). Conducting qualitative interviews by telephone: Lessons learned from a study of alcohol use among sexual minority and heterosexual women. *Qualitative social work, 15*(1), 118-133.

Drummond, M., F.,, Cooke, J., & Walley, T. (1997). Economic evaluation under managed competition: evidence from the UK. *Social science & medicine, 45*(4), 583-595.

Drummond, M. F., Schwartz, J., Jönsson, B., Luce, B., Neumann, P., Siebert, U., & Sullivan, S. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International journal of technology assessment in health care, 24*(3), 244-258.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*: Oxford university press.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B., & Stoddart, G. L. (2005). *Economic evaluation using decision analytic modelling.Methods for the economic evaluation of health care programmes.*

. New York: Oxford University Press.

Duguay, C., & Chetouane, F. (2007). Modeling and improving emergency department systems using discrete event simulation. *Simulation, 83*(4), 311-320.

Dumas, J. S., & Redish, J. C. (1999). *A Practical Guide to Usability Testing*: Intellect Books.

Dupuits, F. M., & Hasman, A. (1995). User satisfaction of general practitioners with HIOS+, a medical decision support system. *Computer methods and programs in biomedicine, 47*(2), 183-188.

Duranti, A. (2006). Transcripts, like shadows on a wall. *Mind, Culture, and Activity, 13*(4), 301-310.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology, 41*(3-4), 327-350.

Duthie, T., Trueman, P., Chancellor, J., & Diez, L. (1999). Research into the use of health economics in decision making in the United Kingdom - Phase II - Is health economics 'for good or evil?'. *Health policy, 46*(2), 143-157. doi:Doi 10.1016/S0168-8510(98)00057-8

Earnshaw, S. R., Wilson, M. R., Dalal, A. A., Chambers, M. G., Jhingran, P., Stanford, R., & Mapel, D. W. (2009). Cost-effectiveness of fluticasone propionate/salmeterol in the treatment of COPD. *Respiratory Medicine, 103*(1), 12-21.

Ebenezer, C. (2003). Usability evaluation of an NHS library website. *Health Information & Libraries Journal, 20*(3), 134-142.

Eddama, O., & Coast, J. (2008). A systematic review of the use of economic evaluation in local decision-making. *Health Policy, 86*(2), 129-141.

Eddama, O., & Coast, J. (2009). Use of economic evaluation in local health care decision-making in England: a qualitative investigation. *Health policy, 89*(3), 261-270.

Eddy, D. M. (2006). Accuracy versus transparency in pharmacoeconomic modelling. *Pharmacoeconomics, 24*(9), 837-844.

Eddy, D. M. (Ed.) (1985). *Assessing Medical Technologies, "Technology Assessment: The role of mathematical modelling."* Washington DC: National Academy Press.

Eddy, D. M., Hollingworth, W., Caro, J. J., Tsevat, J., McDonald, K. M., & Wong, J. B. (2012). Model transparency and validation a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. *Medical decision making, 32*(5), 733-743.

Edwards, P. J., Moloney, K. P., Jacko, J. A., & Sainfort, F. (2008). Evaluating usability of a commercial electronic health record: A case study. *International Journal of Human-Computer Studies, 66*(10), 718-728.

Ekeland, A. G., Bowes, A., & Flottorp, S. (2010). Effectiveness of telemedicine: a systematic review of reviews. *International Journal of Medical Informatics, 79*(11), 736-771.

Eldabi, T., Irani, Z., & Paul, R., J.,. (2002). A proposed approach for modelling health-care systems for understanding. *Journal of Management in Medicine, 16*(2/3), 170-187.

Eldabi, T., Paul, R., J.,, & Young, T. (2007). Simulation modelling in healthcare: reviewing legacies and investigating futures. *Journal of the Operational Research Society, 58*(2), 262-270.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis*: MIT-press.

Eysenbach, G. (2001). What is e-health? *Journal of medical Internet research, 3*(2).

Farmer, T., Robinson, K., Elliott, S. J., & Eyles, J. (2006). Developing and implementing a triangulation protocol for qualitative health research. *Qualitative Health Research, 16*(3), 377-394.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers, 35*(3), 379-383.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2): Stanford university press.

Finch, H., & Lewis, J. (2003). Focus groups. *Qualitative research practice: A guide for social science students and researchers*, 170-198.

Finkelstein, S. M., Speedie, S. M., Demiris, G., Veen, M., Lundgren, J. M., & Potthoff, S. (2004). Telehomecare: Quality, perception, satisfaction. *Telemedicine Journal and E-health, 10*(2), 122-128. doi:10.1089/1530562041641219

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fitzsimmons, D. A., Thompson, J., Hawley, M., & Mountain, G. A. (2011). Preventitive tele-health supported services for early stage chronic obstructive pulmonary disease: a protocol for a pragmatic randomized controlled trial pilot. *Trials, 12*(6).

Fleischmann, K. R., & Wallace, W. A. (2005). A covenant with transparency: Opening the black box of models. *Communications of the ACM, 48*(5), 93-97.

Folstad, A. (2007). Work-domain experts as evaluators: usability inspection of domain-specific work-support systems. *International Journal of Human-Computer Interaction, 22*(3), 217-245.

Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., . . . Palmer, S. (2003). Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health, 25*(4), 325-335.

Fook, J., & Gardner, F. (2007). *Practising critical reflection: A resource handbook: A handbook*: McGraw-Hill Education (UK).

Fossum, M., Ehnfors, M., Fruhling, A., & Ehrenberg, A. (2011). An evaluation of the usability of a computerized decision support system for nursing homes. *Applied clinical informatics, 2*(4), 420.

Franco, L., & Rouwette, E. (2011). Decision development in facilitated modelling workshops. *European Journal of Operational Research, 212*(1), 164-178.

Franco, L. A. (2013). Rethinking soft OR interventions: models as boundary objects. *European Journal of Operational Research, 231*(3), 720-733.

Franco, L. A., & Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research, 205*(3), 489-500. doi:http://dx.doi.org/10.1016/j.ejor.2009.09.030

Gable, G. G., Sedera, D.,Chan, T.,. (2008). Re-conceptualizing information system success: The IS-impact measurement model. *Journal of the association for information systems, 9*(7), 18.

Gani, R., Griffin, J., Kelly, S., & Rutten-van Mölken, M. (2010). Economic analyses comparing tiotropium with ipratropium

or salmeterol in UK patients with COPD. *Primary Care Respiratory Journal, 19*(1), 68-74.

Gellis, Z., D., Kenaley, B., McGinty, J., Bardelli, E., Davitt, J., Ten Have, T.,. (2012). Outcomes of a telehealth intervention for homebound older adults with heart or chronic respiratory failure: a randomized controlled trial. *The Gerontologist, 52*(4), 541-552.

Gillenson, M. L., & Sherrell, D. L. (2002). Enticing online consumers: an extended technology acceptance perspective. *Information & Management, 39*(8), 705-719.

Gillham, B. (2005). *Research Interviewing: The range of techniques: A practical guide*: McGraw-Hill Education (UK).

Giordano, A., Clark, M., & Goodwin, N. (2011). *Perspectives of telehealth and telecare, WSDAN briefing paper*. Retrieved from London:

Gold, M. R. (1996). *Cost-effectiveness in health and medicine*: Oxford university press.

Goodhue, D. L. (1998). Development and Measurement Validity of a Task-Technology Fit Instrument for User Evaluations of Information System. *Decision Sciences, 29*(1), 105-138.

Goodhue, D. L., & Thompson, R. L. (1995). Task Technology Fit and Individual Performance. *MIS Quarterly, 19*(2).

Goodwin, N. (2009). Will telehealth and telecare thrive or perish in a cold financial climate? : Taylor & Francis.

Goodwin, N. (2010). The state of telehealth and telecare in the UK: prospects for integrated care. *Journal of Integrated Care, 18*(6), 3-10.

Goodwin, N. (2011). Are the Whole System Demonstrator trial results a watershed moment for the rise of telehealth? http://www.kingsfund.org.uk/blog/2011/12/are-whole-system-demonstrator-trial-results-watershed-moment-rise-telehealth: The Kings Fund.

Gorman, G. E., Clayton, P.,. (2004). *Qualitative research for the information professional*: Facet publishing.

Gorst, S., Armitage, C., Brownsell, S., & Hawley, M. (2014). Home Telehealth Uptake and Continued Use Among Heart Failure and Chronic Obstructive Pulmonary Disease Patients: a Systematic Review. *Annals of Behavioral Medicine, 48*(3), 323-336. doi:10.1007/s12160-014-9607-x

Grady, H. M. (2000). *Web site design: a case study in usability testing using paper prototypes.* Paper presented at the Proceedings of the 18th annual ACM international conference on Computer documentation: technology & teamwork.

Granic, A., Mitrovic, I., & Marangunic, N. (2008). *Usability evaluation of web portals.* Paper presented at the Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on.

Grant, M., J.,, & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108.

Guba, E. G., Lincoln, Y.S., . (1994). Competing paradigms in qualitative reserach. . In N. K. Denzin, Lincoln, Y.S., (Ed.), *Handbook of qualitative research*: Thousands Oaks: Sage.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods, 18*(1), 59-82.

Haettenschwiler, P. (1999). *Neues anwenderfreundliches Konzept der Entscheidungsunterstützung. Gutes Entscheiden in Wirtschaft*.

Hämäläinen, R. P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research, 228*(3), 623-634.

Han, M. K., Agusti, A., Calverley, P.M.,Celli, B.R.,Criner, G.,Curtis, J.L.,Fabbri, L.M.,Goldin, J.G.,Jones, P.W.,MacNee, W.,. (2010). Chronic obstructive pulmonary disease phenotypes: the future of COPD. *American journal of respiratory and critical care medicine, 182*(5), 598-604.

Harper, B. D., Norman, K.L.,. (1993). *Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5.* Paper presented at the Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference.

Harper, P., R.,, & Pitt, M., A.,. (2004). On the challenges of healthcare modelling and a proposed project life cycle for successful implementation&star. *Journal of the Operational Research Society, 55*(6), 657-661.

Hart, F., & Bond, M. (1995). *Action research for health and social care: A guide to practice*: McGraw-Hill Education (UK).

Hasan, L., Morris, A.,Probets, S.,. (2012). A comparison of usability evaluation methods for evaluating e-commerce websites. *Behaviour & Information Technology, 31*(7), 707-737.

Haux, R. (2006). Health information systems–past, present, future. *International journal of medical informatics, 75*(3), 268-281.

Henderson, C., Knapp, M., Fernández, J. L., Beecham, J., & Hirani, S. P. (2013). Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. *BMJ: British Medical Journal, 346*.

Henderson, N. R. (1995). A practical approach to analyzing and reporting focus groups studies: lessons from qualitative market research. *Qualitative Health Research, 5*(4), 463-477.

Hendy, J., Chrysanthaki, T., Barlow, J., Knapp, M., Rogers, A., Sanders, C., . . . Bardsley, M. (2012). An organisational analysis of the implementation of telecare and telehealth: the whole systems demonstrator. *BMC health services research, 12*(1), 403.

Hertzum, M., & Clemmensen, T. (2012). How do usability professionals construe usability? *International Journal of Human-Computer Studies, 70*(1), 26-42.

HM Treasury. (Ed.) (2003). *The Green Book: Appraisal and Evaluation in Central Government, .* London HMSO.

Hoffmann, C., Stoykova, B. A., Nixon, J., Glanville, J. M., Misso, K., & Drummond, M. F. (2002). Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. *Value in health, 5*(2), 71-78.

Hoffmann, C., & von der Schulenburg, J. M. G. (2000). The influence of economic evaluation studies on decision making.: A European survey. *Health policy, 52*(3), 179-192.

Holden, R. J., & Karsh, B. T. (2010). The technology acceptance model: its past and its future in health care. *Journal of biomedical informatics, 43*(1), 159-172.

Holland, A. (2013). Telehealth reduces hospital admission rates in patients with COPD. *Journal of physiotherapy, 59*(2), 129.

Holliday, A. (2007). *Doing & writing qualitative research*: Sage.

Hoogendoorn, M., Rutten-van Molken, M., Hoogenveen, R. T., van Genugten, M. L. L., Buist, A. S., Wouters, E. F. M., & Feenstra, T. L. (2005). A dynamic population model of disease

progression in COPD. *Eur Respir J, 26*, 223–233.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies, 64*(2), 79-102.

Horton, K. (2008). The use of telecare for people with chronic obstructive pulmonary disease: implications for management. *J Nurs Manag, 16*(2), 173-180.

Howick, S., Eden, C., Ackermann, F., & Williams, T. (2008). Building confidence in models for multiple audiences: The modelling cascade. *European Journal of Operational Research, 186*, 1068-1083.

Hu, H. J., Yen, J., & Guan, S. S. (2008). *A study on the interface usability of B2C hypermarket E-commerce website.* Paper presented at the Proceedings of the Asia-Pacific Services Computing Conference, 2008. .

Hu, P. J. H., Ma, P. C., & Chau, P. Y. K. (1999). Evaluation of user interface designs for information retrieval systems: a computer-based experiment. *Decision Support Systems, 27*(1), 125-143.

Hubben, G. (2009). *The exploration of the combination of web technology and economic modelling to support decision making.* (PhD Thesis), University of Groningen.

Hubben, G. A. A., Bos, J. M., Glynn, D. M., Van Der Ende, A., Van Alphen, L., & Postma, M. J. (2007). Enhanced decision support for policy makers using a web interface to health-economic models—illustrated with a cost-effectiveness analysis of nation-wide infant vaccination with the 7-valent pneumococcal conjugate vaccine in the Netherlands. *Vaccine, 25*(18), 3669-3678.

Huberman, M., & Miles, B. (1994). Data management and analysis methods. Handbook of Qualitative Research. N. Denzin and Y. Lincoln London: Sage.

Hughes, J. A., & Sharrock, W. W. (1997). The philosophy of social research.

Hull and East Yorkshire Hospitals Nhs Trust. (2012). *Emergency Department Roconfiguration - Full Business Case*. Retrieved from http://www.hey.nhs.uk/content/files/corporate/trustBoard/120320/emergencyDeptFullBusCase.pdf

Hutton, J. L., & Ashcroft, R. (2000). What does systematic mean for reviews of methods. In N. Black (Ed.), *Health services research methods: a guide to best practice* BMJ Books.

Inglis, S. C., Clark, R.A., McAlister, F.A., Ball, J., Lewinter, C., Cullington, D., Stewart, S., Cleland, J.G.F. (2011). Structured telephone support or telemonitoring programmes for patients with chronic heart failure. *Cochrane Database of Systematic Reviews 2010*(6).

Ingwersen, P. E. (1992). *Information Retrieval Interaction*: Taylor Graham.

Irvine, A. (2011). Duration, dominance and depth in telephone and face-to-face interviews: A comparative exploration. *International Journal of Qualitative Methods, 10*(3), 202-220.

ISO. (1998). Ergonomic requirements for office work with visual display terminals - Part 11: guidance on usability (ISO 9241-11:1998).

Ives, B., Olson, M. H., & Baroudi, J. J. (1983). The measurement of user information satisfaction. *Commun. ACM, 26*(10), 785-793.

Jaana, M., Pare, G., & Sicotte, C. (2009). Home telemonitoring for respiratory conditions: a systematic review. *American Journal of Managed Care, 15*(5), 313-320.

Jacobsen, N. E. (1999). *Usability evaluation methods: the reliability and usage of cognitive walkthrough and usability test.* (PhD thesis), Department of Communication, Journalism and computer Science, Roskilde University.

Jahangirian, M., Taylor, S., J.,E.,, Eatock, J., Stergioulas, L., K.,, & Taylor, P., M.,. (2015). Causal study of low stakeholder engagement in healthcare simulation projects. *Journal of the Operational Research Society, 66*(3), 369-379.

Jankowski, P., Robischon, S.,Tuthill, D.,Nyerges, T.,Ramsey, K.,. (2006). Design considerations and evaluation of a collaborative, spatio-temporal decision support system. *Transactions in GIS, 10*(3), 335-354.

Jaspers, M., Steen, T., Bos, C., & Geenen, M. (2004). The think aloud method: a guide to user interface design. *International journal of medical informatics, 73*(11), 781-795.

Jaspers, M. W. M. (2009). A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *International journal of medical informatics, 78*(5), 340-353.

Jean, C., Jankovic, M., Stal-Le Cardinal, J., & Bocquet, J. (2015). Predictive modelling of telehealth system deployment. *Journal of Simulation, 9*(2), 182-194.

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. (1991). *User interface evaluation in the real world: a comparison of four techniques.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Jennett, P. A., Hall, L. A., Hailey, D., Ohinmaa, A., Anderson, C., Thomas, R., . . . Scott, R. E. (2003). The socio-economic impact of telehealth: a systematic review. *Journal of telemedicine and telecare, 9*(6), 311-320.

Jódar-Sánchez, F., Ortega, F.,Parra, C.,Gómez-Suárez, C.,Jordán, A.,Pérez, P.,Bonachela, P.,Leal, S.,Barrot, E.,. (2013). Implementation of a telehealth programme for patients with severe chronic obstructive pulmonary disease treated with long-term oxygen therapy. *Journal of Telemedicine and Telecare, 19*(1), 11-17.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction (TOCHI), 3*(4), 320-351.

Johnson, T., Hougland, J., & Clayton, R. (1989). Obtaining reports of sensitive behavior: A comparison of substance-use reports from telephone and face-to-face interviews. *Social Science Quarterly, 70*(1), 174.

Joseph, A. (2006). Care coordination and telehealth technology in promoting self-management among chronically ill patients. *Telemedicine Journal & e-Health, 12*(2), 156-159.

Joseph, V., West, R. M., Shickle, D., Keen, J., & Clamp, S. (2011). Key challenges in the development and implementation of telehealth projects. *Journal of Telemedicine & Telecare, 17*(2), 71-77.

Ju, B., & Gluck, M. (2005). User-process model approach to improve user interface usability. *Journal of the American Society for Information Science and Technology, 56*(10), 1098-1112.

Jubran, A., Gross, N., Ramsdell, J., Simonian, R., Schuttenhelm, K., Sax, M., . . . Sonnenberg, F. A. (1993). Comparative cost-effectiveness analysis of theophylline and ipratropium bromide in chronic obstructive pulmonary disease. A three-center study.

. *Chest, 103*, 678-684.

Jun, J., B.,, Jacobson, S., H.,, & Swisher, J., R.,. (1999). Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society, 50*(2), 109-123.

Kaltenthaler, E., Tappenden, P., Paisley, S., & Squires, H. (2011). Identyfying and reviewing evidence to inform conceptualisation and population of cost-effectiveness models. *NICE DSU Support Documents, 14*.

Kanungo, S., Sharma, S., & Jain, P. K. (2001). Evaluation of a decision support system for credit management decisions. *Decision Support Systems, 30*(4), 419-436.

Karat, C., M.,, Campbell, R., & Fiegel, T. (1992). *Comparison of empirical testing and walkthrough methods in user interface evaluation.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Karnon, J. (2003). Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health economics, 12*(10), 837-848.

Karnon, J., & Afzali, H., H.,A.,. (2014). When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics, 32*(6), 547-558.

Karnon, J., Stahl, J., Brennan, A., Caro, J. J., Mar, J., & Möller, J. (2012). Modeling using discrete event simulation a report of the ISPOR-SMDM modeling good research practices task force–4. *Medical decision making, 32*(5), 701-711.

Kidholm, K., Ekeland, A. G., Jensen, L. K., Rasmussen, J., Pedersen, C. D., Bowes, A., . . . Bech, M. (2012). A model for assessment of telemedicine applications: mast. *International journal of technology assessment in health care, 28*(01), 44-51.

Kilsdonk, E., Peute, L. W., Riezebos, R. J., Kremer, L. C., & Jaspers, M. (2013). From an expert-driven paper guideline to a user-centred decision support system: A usability comparison study. *Artificial intelligence in medicine, 59*(1), 5-13.

King, W., R., He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43*(6), 740-755.

Kirakowski, J., & Cierlik, B. (1998). *Measuring the usability of web sites.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Klecun-Dabrowska, E., & Cornford, T. (2000). Telehealth acquires meanings: information and communication technologies within health policy. *Information Systems Journal, 10*(1), 41-63.

Klopping, I. M., & McKinney, E. (2004). Extending the technology acceptance model and the task-technology fit model to consumer e-commerce. *Information Technology, Learning, and Performance Journal, 22*(1), 35.

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information systems research, 13*(2), 205-223.

Koutsabasis, P., Spyrou, T., & Darzentas, J. (2007). Evaluating usability evaluation methods: criteria, method and a case study *Human-Computer Interaction. Interaction Design and Usability* (pp. 569-578): Springer.

Kruse, C. S., Krowski, N., Rodriguez, B., Tran, L., Vela, J., & Brooks, M. (2017). Telehealth and patient satisfaction: a systematic review and narrative analysis. *BMJ Open, 7*(8), e016242.

Kushniruk, A. W., & Patel, V. L. (1998). Cognitive evaluation of decision making processes and assessment of information technology in medicine. *International journal of medical informatics, 51*(2), 83-90.

Kvale, S. (2008). *Doing interviews*: Sage.

Lacy, L. W., Randolph, W., Harris, B., Youngblood, S., Sheehan, J., Might, R., & Metz, M. (2001). *Developing a consensus perspective on conceptual models for simulation systems.* Paper presented at the Proceedings of the 2001 Spring Simulation Interoperability Workshop.

Larsen, T. J., Sorebo A., & Sorebo, O. (2009). The role of task-technology fit as users' motivation to continue information system use. *Computers in Human Behavior, 25*(3), 778-784.

Le Goff-Pronost, M., & Sicotte, C. (2010). The added value of thorough economic evaluation of telemedicine networks. *The European Journal of Health Economics, 11*(1), 45-55.

Lederer, A. L., Maupin, D.J.,Sena, M.P.,Zhuang, Y.,. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems, 29*(3), 269-282.

Lee, Y. J., & Lin, C. J. (2014). Evaluation and Satisfaction Survey on the Interface Usability of Online Publishing Software. *Mathematical Problems in Engineering 2014*.

Legard, R., Keegan, J., & Ward, K. (2003). In-depth interviews. *Qualitative research practice: A guide for social science students and researchers*, 138-169.

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management, 40*(3), 191-204.

Leslie, S. J., Hartswood, M., Meurig, C., McKee, S., P., , Slack, R., Procter, R., & Denvir, M. A. (2006). Clinical decision support software for management of chronic heart failure: Development and evaluation. *Computers in biology and medicine, 36*(5), 495-506.

Lewis, C. (1982). *Using the "Think Aloud" method in cognitive interface design*. Retrieved from

Lewis, C., Polson, P.G., Wharton, C., . (1990). *Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Lewis, J. (2009). Redefining qualitative methods: Believability in the fifth moment. *International Journal of Qualitative Methods, 8*(2), 1-14.

Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 36*(2), 368-378.

Li, A. C., Kannry, J. L., Kushniruk, A., Chrimes, D., McGinn, T. G., Edonyabo, D., & Mann, D. M. (2012). Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. *International journal of medical informatics, 81*(11), 761-772.

Lin, C. A., Neafsey, P. J., & Strickler, Z. (2009). Usability testing by older adults of a computer-mediated health communication program. *Journal of health communication, 14*(2), 102-118.

Lin, Y. L., & Lin, D. (2009). Usability Evaluation of Dynamic RSVP Interface on Web Page *Human Interface and the Management of Information. Information and Interaction* (pp. 55-64): Springer.

Lowry, P. B., Spaulding, T., Wells, T., Moody, G., Moffit, K., & Madariaga, S. (2006). *A theoretical model and empirical results linking website interactivity and usability satisfaction.* Paper presented at the System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on.

Luoma, J. (2016). Model-based organizational decision making: A behavioral lens. *European Journal of Operational Research, 249*(3), 816-826.

Ma, Q., & Liu, L. (2004). The technology acceptance model: A meta-analysis of empirical findings. *Journal of Organizational and End User Computing (JOEUC), 16*(1), 59-72.

Mahmood, M. A., Burn, J. M., Gemoets, L. A., & Jacquez, C. (2000). Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature. *International Journal of Human-Computer Studies, 52*(4), 751-771.

Maniadakis, N., Tzanakis, N., Fragoulakis, V., Hatzikou, M., & Siafakas, N. (2006). Economic evaluation of tiotropium and salmeterol in the treatment of chronic obstructive pulmonary disease (COPD) in Greece. *Current Medical Research and Opinion, 22*(8), 1599-1607.

Mannino, D. M., & Buist, A. S. (2007). Global burden of COPD: risk factors, prevalence, and future trends. *The Lancet, 370*(9589), 765-773.

March, J. G., & Simon, H. A. (1958). Organizations.

Margolis, E., & Pauwels, L. (2011). *The Sage handbook of visual research methods*: Sage.

Maxwell, J. A. (1996). *Qualitative research design: An interactive approach: An interactive approach*: Sage.

May, C., Harrison, R., MacFarlane, A., Williams, T., Mair, F., & Wallace, P. (2003). Why do telemedicine systems fail to normalize as stable models of service delivery? *Journal of Telemedicine and Telecare, 9*(suppl 1), 25-26.

May, T. (2001). *Social research*: McGraw-Hill Education (UK).

McCabe, C., & Dixon, S. (2000). Testing the validity of cost-effectiveness models. *Pharmacoeconomics, 17*(5), 501-513.

McHaney, R., & Cronan, T. P. (1998). Computer Simulation Success: On the Use of the End-User Computing Satisfaction Instrument: A Comment. *Decision Sciences, 29*(2), 525-535.

McHaney, R., Cronan, T. P., Hightower, R., Pearson, J., & White, D. (1999). EUCS test–retest reliability in representational model decision support systems. *Information & Management, 36*(2), 109-119.

McIntosh, E., & Cairns, J. (1997). A framework for the economic evaluation of telemedicine. *Journal of Telemedicine & Telecare, 3*(3), 132-139.

McKinstry, B., Pinnock, H.,Sheikh, A.,. (2009). Telemedicine for management of patients with COPD? *The Lancet, 374*(9691), 672-673.

McLean, S., Nurmatov, U., LyLiu, J., Pagliari, C., Car, J., Sheikh, A., . . . Aziz, S. (2011). Telehealthcare for long term conditions. *Bmj, 342*, d120.

McLean, S., Nurmatov, U.,Liu, J., Pagliari, C.,Car, J.,Sheikh, A.,. (2012). Telehealthcare for chronic obstructive pulmonary disease: Cochrane Review and meta-analysis. *British Journal of General Practice, 62*(604), e739-e749.

Medical Research Council. (2008). *Developing and evaluating complex interventions: new guidance*. Retrieved from www.mrc.ac.uk/complexinterventionsguidance:

Mersey Care NHS Trust. (2012). Guidance on completing a business case.   Retrieved from http://www.merseycare.nhs.uk/Library/What_we_do/Corporate_Services/Service_Development_Delivery/Document_Library/Business%20Case%20Guidance.pdf

Mertens, D. M. (2005). Quality criteria in Qualitative research. *Pattons version*.

Mielczarek, B., & Uziałko-Mydlikowska, J. (2010). Application of computer simulation modeling in the health care sector: a survey. *Simulation*, 0037549710387802.

Mistry, H. (2012). Systematic review of studies of the cost-effectiveness of telemedicine and telecare. Changes in the economic evidence over twenty years. *Journal of Telemedicine & Telecare, 18*(1), 1-6.

Mistry, H., Garnvwa, H., & Oppong, R. (2014). Critical appraisal of published systematic reviews assessing the cost-effectiveness of telemedicine studies. *Telemedicine and e-Health, 20*(7), 609-618.

Molich, R., & Nielsen, J. (1990). Improving a human- computer dialogue. *Communications of the ACM, 33*(3), 338-348.

Molla, A., & Licker, P. S. (2001). E-Commerce Systems Success: An Attempt to Extend and Respecify the Delone and MaClean Model of IS Success. *J. Electron. Commerce Res., 2*(4), 131-141.

Monks, T., Robinson, S., & Kotiadis, K. (2014). Can involving clients in simulation studies help them solve their future problems? A transfer of learning experiment. *European Journal of Operational Research*.

Morris, A. (2015). *A practical introduction to in-depth interviewing*: Sage.

Morris, D. E., Oakley, J. E., & Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software, 52*, 1-4.

Mun, Y. Y., Hwang, Y.,. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies, 59*(4), 431-449.

Murphy, S. P., & Perera, T. (2002). Successes and failures in UK/US development of simulation. *Simulation Practice and Theory, 9*(6), 333-348.

Naik, S., Kamal. K., Keys, P., & Mattei, T. (2010). Evaluating the cost-effectiveness of tiotropium versus salmeterol in the treatment of chronic obstructive pulmonary disease. *ClinicoEconomics and Outcome Research, 2*, 25-36.

National Institute for Health and Clinical Excellence. (2010). *Chronic heart failure: costing template*. Retrieved from http://guidance.nice.org.uk/CG108/CostingTemplate/xls/English

National Institute for Health and Clinical Excellence. (2011). *Chronic obstructive pulmonary disease costing template* Retrieved from http://guidance.nice.org.uk/CG101/CostingTemplate/xls/English

NHS. (1998). *Information for health: an information strategy for the modern NHS 1998-2005: a national strategy for local implementation*: NHS Executive.

NHS. (2001). Building the information core: implementing the NHS Plan. *NHS Executive*.

NHS England. (2014). Commissioning Simulation Model.

NHS Lothian. (2010). *Royal Victoria Building - Full Business Case*. Retrieved from http://www.nhslothian.scot.nhs.uk/GoingToHospital/Locations/RVB/NewBuilding/Documents/RVH-FullBusinessCase.pdf

NICE. (2014). *Interim methods guide for developing service guidance* Retrieved from retrieved from http://publications.nice.org.uk/interim-methods-guide-for-developing-service-guidance-2014-pmg8 29.07.2015:

NICE. (2015).

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175.

Nielsen, J. (1994). *Usability Engineering*. San Franscisco: Morgan Kaufman.

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Commun. ACM, 37*(4), 66-75.

Nielsen, J., & Mack, R. L. (1994). *Usability Inspection Methods*. New York: Willey.

Nielsen, R., Johannessen, A., Benediktsdottir, B., Gislason, T., Buist, A. S., Gulsvik, A., . . . Lee, T. A. (2009). Present and future cost of COPD in Iceland and Norway: results from the BOLD study. *Eur Respir J, 34*, 850-857.

Nishimura, K., & Tsukino, M. (2000). Clinical course and prognosis of patients with chronic obstructive pulmonary disease. *Current Opinion in Pulmonary Medicine, 6*(2), 127-132.

Norgaard, M., & Hornbæk, K. (2006). *What do usability evaluators do in practice?: an explorative study of think-aloud testing.* Paper presented at the Proceedings of the 6th conference on Designing Interactive systems.

Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them*. Retrieved from http://eprint.ihmc.us/5/

Oba, Y. (2009). Cost-effectiveness of salmeterol, fluticasone, and combination therapy for COPD. *The American journal of managed care, 15*(4), 226.

Office Cabinet. (1999). Modernising government. *London: The Stationery Office*.

Ohinmaa, A., Hailey, D., & Roine, R. (2001). Elements for assessment of telemedicine applications. *International journal of technology assessment in health care, 17*(2), 190-202.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, 460-469.

Oostenbrink, J. B., Rutten-van Mölken, M., Monz, B. U., & FitzGerald, J. M. (2005). Probabilistic Markov Model to Assess the Cost-Effectiveness of Bronchodilator Therapy in COPD Patients in Different Countries. *Value in health, 8*(1), 32-46.

Ormerod, R. J. (2013). The mangle of OR practice: towards more informative case studies of/technical'projects. *Journal of the Operational Research Society, 65*(8), 1245-1260.

Ossebaard, H. C., Seydel, E. R., & van Gemert-Pijnen, L. (2012). Online usability and patients with long-term conditions: a mixed-methods approach. *International journal of medical informatics, 81*(6), 374-387.

Oxford University Hospital. (2012). *Expansion of Newborn Facilities - Full Business Case*. Retrieved from http://www.ouh.nhs.uk/about/trust-board/2012/may/documents/TB201235a_ExpansionofNeonatalFacilities2.pdf

Pagani, M. (2006). Determinants of adoption of high speed data services in the business market: evidence for a combined technology acceptance model with task technology fit model. *Information & Management, 43*(7), 847-860.

Pandor, A., Thokala, P., Gomersall, T., Baalbaki, H., Stevens, J. W., Wang, J., . . . Fitzgerald, F. (2012). Home telemonitoring or structured telephone support programmes after recent discharge in

patients with heart failure: Systematic review and economic evaluation. *Health Technol Assess.*

Paradowski, M., & Fletcher, A. (2004). Using task analysis to improve usability of fatigue modelling software. *International Journal of Human-Computer Studies, 60*(1), 101-115.

Patton, M. Q. (2002). Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative social work, 1*(3), 261-283.

Pauwels, R. A., Buist, A. S., Calverley, P.M., Jenkins, C.R.,Hurd, S.S.,. (2014). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine, 163*(5).

Petter, S., DeLone, W.,McLean, E.,. (2008). Measuring information systems success: models, dimensions, measures, and interrelationships. *European journal of information systems, 17*(3), 236-263.

Peute, L., W., P., , & Jaspers, M. (2005). Usability evaluation of a laboratory order entry system: cognitive walkthrough and think aloud combined. *Studies in health technology and informatics, 116*, 599.

Philips, Z., Ginnelly, L., Sculpher, M., Claxton, K., Golder, S., Riemsma, R., . . . Glanville, J. (2004). Review of guidelines for good practice in decision-analytic modelling in health technology assessment.

Pike, R. H., Tayles, M. E., & Mansor, N. N. A. (2011). Activity-based costing user satisfaction and type of system: A research note. *The British Accounting Review, 43*(1), 65-72.

Polinski, J. M., Barker, T., Gagliano, N., Sussman, A., Brennan, T. A., & Shrank, W. H. (2016). Patients' satisfaction with and preference for telehealth visits. *Journal of general internal medicine, 31*(3), 269-275.

Polisena, J., Coyle, D., Coyle, K., & McGill, S. (2009). Home telehealth for chronic disease management: A systematic review and an analysis of economic evaluations. *International journal of technology assessment in health care, 25*(3), 339-349.

Polisena J., Tran K., Cimon K., Hutton B., McGill S., & Palmer K. (2009). Home telehealth for diabetes management: a systematic review and meta-analysis. *Diabetes Obes. Metab., 10*, 913-930.

Polson, P., & Lewis, C. (1990). Theory- based design for easily learned interfaces. *Human Computer Interaction, 5*(2&3), 191-220.

Power, D. J. (2002). *Decision support systems: concepts and resources for managers.*: Quorum Books.

Proudlove, N., C.,, Black, S., & Fletcher, A. (2007). OR and the challenge to improve the NHS: modelling for insight and improvement in in-patient flows. *Journal of the Operational Research Society, 58*(2), 145-158.

QRS International. (2012). Nvivo qualitative data analysis Software.

Rabe, K. F., Hurd, S., Anzueto, A., Barnes, P. J., Buist, S. A., Calverley, P., . . . Zielinski, J. (2007). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine, 176*(6), 532-555.

Reardon, T. (2005). Research Findings and Strategies for Assessing Telemedicine Cost. *Telemedicine and e-health, 11*(3), 348-369.

Reason, P., & Bradbury, H. (2001). *Handbook of action research: Participative inquiry and practice*: Sage.

Rhenius, D., & Deffner, G. (1990). *Evaluation of concurrent thinking aloud using eye-tracking data.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Ritchie, J., & Lewis, J. (Eds.). (1994). *Qualitative Research Practice: A Guide for Social Science Students and Researchers* SAGE.

Ritchie, J., & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & B. Burgess (Eds.), *Analyzing Qualitative Data* (pp. 173-194). London and New York: Routledge.

Roberts, L., V., , & Fels, D. I. (2006). Methods for inclusion: Employing think aloud protocols in software usability studies with individuals who are deaf. *International Journal of Human-Computer Studies, 64*(6), 489-501.

Roberts, M., Russell, L. B., Paltiel, A. D., Chambers, M., McEwan, P., & Krahn, M. (2012). Conceptualizing a Model A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. *Medical Decision Making, 32*(5), 678-689.

Robinson, S. (2008). Conceptual modelling for simulation Part II: a framework for conceptual modelling. *Journal of the Operational Research Society, 59*(3), 291-304.

Robinson, S. (Ed.) (2011). *Conceptual modeling for simulation*. New York: Wiley.

Robinson, S., Nance, R. E., Paul, R. J., Pidd, M., & Taylor, S. J. E. (2004). Simulation model reuse: definitions, benefits and obstacles. *Simulation modelling practice and theory, 12*(7), 479-494.

Robinson, S., & Pidd, M. (1998). Provider and customer expectations of successful simulation projects. *Journal of the Operational Research Society*, 200-209.

Rodriguez-Roisin, R. (2000). Toward a Consensus Definition for COPD Exacerbations. *Chest, 117*(5, Supplement 2), 398S-401S. doi:http://doi.org/10.1378/chest.117.5_suppl_2.398S

Rodriguez-Roisin, R. (2006). COPD exacerbations· 5: Management. *Thorax, 61*(6), 535-544.

Rodriguez-Ulloa, R., & Paucar-Caceres, A. (2005). Soft system dynamics methodology (SSDM): combining soft systems methodology (SSM) and system dynamics (SD). *Systemic Practice and Action Research, 18*(3), 303-334.

Rogers, E. M. (2003). *Diffusion of Innovations.* . New York: Free Press.

Rouwette, E., Korzilius, H., Vennix, J., & Jacobs, E. (2011). Modeling as persuasion: the impact of group model building on attitudes and behavior. *System Dynamics Review, 27*(1), 1-21.

Rubin, H., & Rubin, J. (2005). Qualitative interviewing: The art of hearing data. *Thousand Oaks, CA: Sage*.

Rubin, J. (1994). Handbook of Usability Testing, 1994. New York: John Wiley and Sons, .

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: howto plan, design, and conduct effective tests*: John Wiley & Sons.

Ruland, C. M. (2004). A survey about the usefulness of computerized systems to support illness management in clinical practice. *International Journal of Medical Informatics, 73*(11), 797-805.

Rutten-van Molken, M., & Lee, T. A. (2006). Economic Modeling in Chronic Obstructive Pulmonary Disease. *Proc Am Thorac Soc, 3*(7), 630-634.

Rutten-van Mölken, M., Oostenbrink, J., Miravitlles, M., & Monz, B. (2007). Modelling the 5-year cost effectiveness of tiotropium, salmeterol and ipratropium for the treatment of chronic obstructive pulmonary disease in Spain. *The European Journal of Health Economics, 8*(2), 123-135.

Rycroft-Malone, J., Harvey, G., Seers, K., Kitson, A., McCormack, B., & Titchen, A. (2004). An exploration of the factors that influence the implementation of evidence into practice. *Journal of clinical nursing, 13*(8), 913-924.

Saarinen, T. (1996). An expanded instrument for evaluating information system success. *Information & Management, 31*(2), 103-118.

Sabatier, P. A., & Weible, C. M. (2007). *Theories of the policy process*: Westview Press.

Sanders, C., Rogers, A., Bowen, R., Bower, P., Hirani, S., Cartwright, M., . . . Hendy, J. (2012). Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. *BMC health services research, 12*(1), 220.

Sanders, G. L. (1984). MIS/DSS success measure. *Systems, Objectives, Solutions, 4*(1), 29-34.

Sayer, A. (2000). *Realism and social science*: Sage.

Scanlon, P. D., Connett, J.E., Waller, L.A. (2000). Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung

Health Study. *Am J Respir Crit Care Med, 161*, 381-390.

Seddon, P., Kiew, M.Y.,. (1996). A partial test and development of DeLone and McLean's model of IS success. *Australasian Journal of Information Systems, 4*(1).

Seddon, P. B. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information systems research, 8*(3), 240-253.

Seffah, A., Donyaee, M., Kline, R. B., & Padda, H. K. (2006). Usability measurement and metrics: A consolidated model. *Software Quality Journal, 14*(2), 159-178.

Serry, T., & Liamputtong, P. (2013). The in-depth interviewing method in health. *Research methods in health: Foundations for evidence-based practice*, 39-53.

Shackel, B., Richardson, S. J., . (1991). *Human factors for informatics usability*: Cambridge university press.

Shaikh, A. A., Karjaluoto, H.,. (2015). Making the most of information technology & systems usage: A literature review, framework and future research agenda. *Computers in Human Behavior, 49*(0), 541-566. doi:http://dx.doi.org/10.1016/j.chb.2015.03.059

Silverman, D. (Ed.) (2000). *Doing Qualitative Research: A Practical Handbook* Sage Publications.

Simon, H. A. (1977). Scientific discovery and the psychology of problem solving *Models of discovery* (pp. 286-303): Springer.

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations research, 6*(1), 1-10.

Simpson, C. R., Hippisley-Cox, J.,Sheikh, A.,. (2010). Trends in the epidemiology of chronic obstructive pulmonary disease in England: a national study of 51 804 patients. *British Journal of General Practice, 60*(576), e277-e284.

Sin, D., Golmohammadi, K., Jacobs, P. (2004). Cost-effectiveness of inhaled corticosteroids for chronic obstructive pulmonary disease according to disease severity. *The American Journal of Medicine, 116*(5), 325-331. doi:10.1016/j.amjmed.2003.09.027

Sisk, J. E., Sanders, J.H.,. (1998). A proposed framework for economic evaluation of telemedicine. *Telemedicine Journal, 4*(1), 31-37.

Snoswell, C., Smith, A., C.,, Scuffham, P., A.,, & Whitty, J., A.,. (2016). Economic evaluation strategies in telehealth: obtaining a more holistic valuation of telehealth interventions. *Journal of Telemedicine and Telecare*.

Spencer, M., Briggs, A. H., Grossman, R. F., & Rance, L. (2005). Development of an Economic Model to Assess the Cost Effectiveness of Treatment Interventions for Chronic Obstructive Pulmonary Disease. *Pharmacoeconomics, 23*(6), 619-637.

Spool, J., & Schroeder, W. (2001). *Testing web sites: Five users is nowhere near enough.* Paper presented at the CHI'01 extended abstracts on Human factors in computing systems.

Squires, H. (2014). *A methodological framework for developing the structure of Public Health economic models.* (PhD thesis).

Srivastava, A., & Thomson, S., B.,. (2009). Framework analysis: a qualitative methodology for applied policy research.

Standfield, L., Comans, T., & Scuffham, P. (2014). Markov modeling and discrete event simulation in health care: a systematic comparison. *International journal of technology assessment in health care, 30*(02), 165-172.

Starkie, H. J., Briggs, A. H., & Chambers, M. G. (2008). Pharmacoeconomics in COPD: lessons for the future. *International Journal of Chronic Obstructive Pulmonary Disease, 3*(1), 71.

Stephens, N. (2007). Collecting data from elites and ultra elites: telephone and face-to-face interviews with macroeconomists. *Qualitative Research, 7*(2), 203-216.

Sterman, J. D. (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review, 18*(4), 501-531.

Steventon, A., Bardsley, M., Billings, J., Dixon, J., Doll, H., Hirani, S., . . . Henderson, C. (2012). Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *Bmj, 344*.

Stewart, D. W., & Shamdasani, P. N. (2014). *Focus groups: Theory and practice* (Vol. 20): Sage publications.

Stinson, J., McGrath, P., Hodnett, E., Feldman, B., Duffy, C., Huber, A., Tucker, L., Hetherington, R., Tse, S., Spiegel, L.,. (2010). Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. *Journal of medical Internet research, 12*(3).

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*: Sage Publications, Inc.

Swaak, M., de Jong, M., & de Vries, P. (2009). *Effects of information usefulness, visual attractiveness, and usability on web visitors' trust and behavioral intentions.* Paper presented at the Professional Communication Conference, 2009. IPCC 2009. IEEE International.

Tako, A. A., & Kotiadis, K. (2015). PartiSim: A multi-methodology framework to support facilitated simulation modelling in healthcare. *European Journal of Operational Research, 244*(2), 555-564.

Tako, A. A., & Robinson, S. (2010). Model development in discrete-event simulation and system dynamics: An empirical study of expert modellers. *European Journal of Operational Research, 207*(2), 784-794.

Tanriverdi, H., & Iacono, C. S. (1999). Diffusion of telemedicine: a knowledge barrier perspective. *Telemedicine Journal, 5*(3), 223-244.

Tappenden, P. (2012). Conceptual modelling for health economic model development. *HEDS Discussion Paper 12/05*.

Taylor, J., Coates, E., Brewster, L., Mountain, G., Wessels, B., & Hawley, M. S. (2015). Examining the use of telehealth in community nursing: identifying the factors affecting frontline staff acceptance and telehealth adoption. *Journal of advanced nursing, 71*(2), 326-337.

Taylor, J., Oguntuase, O., Gorst, S., Coates, E., & Armitage, C. J. (2014). Does telehealth promote self-care? A qualitative study examining the experience of older people using telehealth for long-term condition management. *International Journal of Integrated Care, 14*(8).

Taylor, R. S., & Iglesias, C. P. (2009). Assessing the Clinical and Cost-Effectiveness of Medical Devices and Drugs: Are They That Different? *Value in health, 12*(4), 404-406.

Taylor, S., Brailsford, S. C., Chick, S., L'Ecuyer, P., Macal, C., & Nelson, B. (2013a). Modelling and simulation grand challenges: an OR/MS perspective. *Proceedings of the 2013 Winter Simulation Conference*.

Taylor, S., J.,E.,, Brailsford, S., Chick, S., E.,, L'Ecuyer, P., Macal, C., M.,, & Nelson, B., L.,. (2013b). *Modeling and simulation grand challenges: an OR/MS perspective.* Paper presented at the Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World.

Taylor, S. J. E., Chick, S. E., Macal, C. M., Brailsford, S. C., L'Ecuyer, P., & Nelson, B. L. (2013c). *Modeling and simulation grand challenges: An OR/MS perspective.* Paper presented at the Simulation Conference (WSC), 2013 Winter.

Technology Strategy Board. (2011). Mainstreaming Assistive Technology for Users, Families and Carers. ALIP Project Guide.

Tessier, S. (2012). From field notes, to transcripts, to tape recordings: Evolution or combination? *International Journal of Qualitative Methods, 11*(4), 446-460.

The Ipswich Hospital. (2012a). *Compliant Sterilisation Services Department - Full Business Case*. Retrieved from http://www.ipswichhospital.nhs.uk/aboutourhospital/Documents/Publication%20Scheme/IHT%20Compliant-SSD%20FBC-FINAL-v1.1-PUBLIC.pdf

The Ipswich Hospital. (2012b). *Invasive Cardiology Service Developement - Full Business Case,* . Retrieved from

The King's Fund. (2006). *Combined Predictive Model. Final Report*. Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_document/PARR-combined-predictive-model-final-report-dec06.pdf:

The MALT Study Consortium. (2014). *MALT- Overcoming the Barriers to Mainstreaming Assisted Living Technologies. Summary research report*. Retrieved from http://malt.group.shef.ac.uk/summary/home:

The Map of Medicine. (2014).

The Royal Wolverhampton Hospitals NHS Trusts. (2011). *Integrated Pathology Full Business Case*. Retrieved from

The Shrewsbury and Telford Hospital NHS Trust. (2012). The Future Configuration of Hospital Services. Securing High Quality, Safe and Sustainable Hospital Services in Shrewsbury and Telford. Full Business Case Summary

Thokala, P., Dodd, P., Brennan, A., Dixon, S., Lowrie, K., . (2018). Developing Markov models based on administrative hospitalisation data - a case study of chronic heart failure. *in print*.

Thokala, P., Hassan, B., Abdullah, P., Stevens, J., & Alan, B. (2013). Cost-effectiveness modelling of telemonitoring after discharge from hospital with heart failure. *International Journal of Integrated Care, 13*(7).

Totten, A. M., Womack, D. M., Eden, K. B., McDonagh, M. S., Griffin, J. C., Grusing, S., & Hersh, W. R. (2016). Telehealth: mapping the evidence for patient outcomes from systematic reviews.

Trafton, J., Martins, S., Michel, M., Lewis, E., Wang, D., Combs, A., Scates, N., Tu, S., Goldstein, M. K.,. (2010). Evaluation of the acceptability and usability of a decision support system to encourage safe and effective use of opioid therapy for chronic, noncancer pain by primary care providers. *Pain Medicine, 11*(4), 575-585.

Trappenburg, J., Niesink, A.,de Weert-van Oene, G.H.,van der Zeijden, H.,van Snippenburg, R.,Peters, A.,Lammers, J.W., Schrijvers, A.,. (2008). Effects of telemonitoring in patients with chronic obstructive pulmonary disease. *Telemedicine and e-Health, 14*(2), 138-146.

Udsen, F. W., Hejlesen, O., & Ehlers, L. H. (2014). A systematic review of the cost and cost-effectiveness of telehealth for patients suffering from chronic obstructive pulmonary disease. *Journal of telemedicine and telecare, 1*(9).

Ulgen, O. (1991). Proper management techniques are keys to a successful simulation project. *Industrial Engineering, 23*(8), 37-41.

van Gool, K., Gallego, G., Haas, M.,Viney, R.,Hall, J.,Ward, R.,. (2007). Economic Evidence at the Local Level. *Pharmacoeconomics, 25*(12), 1055-1062.

Van Hulst, M., Hubben, G. A. A., Sagoe, K. W. C., Promwong, C., Permpikul, P., Fongsatitkul, L., . . . Postma, M. J. (2009). Blood donors and blood collection: Web interface–supported transmission risk assessment and cost-effectiveness analysis of postdonation screening: a global model applied to Ghana, Thailand, and the Netherlands. *Transfusion, 49*(12), 2729-2742.

van Velden, M. E., Severens, J. L., & Novak, A. (2005). Economic evaluations of healthcare programmes and decision making. *Pharmacoeconomics, 23*(11), 1075-1082.

Venkatesh, V., Davis,F.D.,. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science, 46*(2), 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Venkatesh, V., Morris, M., Davis, . (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly, 27*(3), 425-478.

Vontetsianos, T., Giovas, P., Katsaras, T., Rigopoulou, A., Mpirmpa, G., Giaboudakis, P., . . . Tsoulkas, B. (2005). Telemedicine-assisted home support for patients with advanced chronic obstructive pulmonary disease: preliminary results after nine-month follow-up. *Journal of telemedicine and telecare, 11*(suppl 1), 86-88.

Vredenburg, K., Mao, J.Y., Smith, P.W., Carey, T., . (2002). *A survey of user-centered design practice.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Wakefield, B. J., Holman, J.E., Ray, A., Scherubel, M., Burns, T. L., Kienzle, M.G., Rosenthal, G.E.,. (2009). Outcomes of a home telehealth intervention for patients with heart failure. *Journal of Telemedicine and Telecare, 15*(1), 46-50.

Wang, S. K., Yang, C.,. (2005). The interface design and the usability testing of a fossilization web-based learning environment. *Journal of Science Education and Technology, 14*(3), 305-313.

Ward, D. J., Furber, C., Tierney, S., & Swallow, V. (2013). Using framework analysis in nursing research: a worked example. *Journal of advanced nursing, 69*(11), 2423-2431.

Ward, J. L., Hiller, S., . (2005). Usability Testing, Interface Design, and Portals. *Journal of Library Administration, 43:1-2*, 155-171.

Weber, M. (1964). *The theory of social and economic organization* New York: Free Press.

Weinstein, M. C. (1990). Principles of cost-effective resource allocation in health care organizations. *International journal of technology assessment in health care, 6*(1), 93-103.

Weiss, C. H. (1979). The many meanings of research utilization. *Public administration review, 39*(5), 426-431.

Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British medical bulletin, 96*(1), 5-21.

WHO. (1998). *A Health Telematics Policy in Support of WHO'S Health-For-All Strategy for Global Development: Report of the WHO Group Consultation on Health Telematics* Geneva: World Health Organization.

WHO. (2015). International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10).

Willemain, T., Wallace, W., Fleischmann, K., Waisel, L., & Ganaway, S. (2003). Bad numbers: coping with flawed decision support. *Journal of the Operational Research Society, 54*(9), 949-957.

Willemain, T. R. (1989). Graphical adjustment of statistical forecasts. *International Journal of Forecasting, 5*(2), 179-185.

Willemain, T. R. (1995). Model formulation: What experts think about and when. *Operations research, 43*(6), 916-932.

Williams, I., & Bryan, S. (2007a). Understanding the limited impact of economic evaluation in health care resource allocation: a conceptual framework. *Health policy, 80*(1), 135-143.

Williams, I., McIver, S., Moore, D., & Bryan, S. (2008). *The use of economic evaluations in NHS decision-making: a review and empirical investigation*: Gray Pub.

Williams, I., P.,, & Bryan, S. (2007b). Cost-effectiveness analysis and formulary decision making in England: findings from research. *Social science & medicine, 65*(10), 2116-2129.

Williams, T., May, C., Mair, F., Mort, M., & Gask, L. (2003). Normative models of health technology assessment and the social production of evidence about telehealth care. *Health Policy, 64*(1), 39-54.

Woolrych, A., Cockton, G., . (2001). *Why and when five test users aren't enough.* Paper presented at the Proceedings of IHM-HCI 2001 conference.

Worsley, G. (2013). Assisted Living Innovation Platform and the Dallas programme.

Wrubel, L., Schmidt, K.,. (2007). Usability testing of a metasearch interface: A case study. *College & research libraries, 68*(4), 292-311.

Wu, Y. L., Tao, Y. H. , Yang, P. C., . (2009). *The discussion on influence of website usability towards user acceptability.* Paper presented at the Management and Service Science, 2009. MASS'09. International Conference on.

Yeh, Y. T., Wu, S. J., Liu, C.T., Lee, T.I., . (2008). *Assessment of user satisfaction with an Internet-based integrated patient education system for diabetes management.* Paper presented at the 10th International Conference on e-health Networking, Applications and Services, 2008.

Yen, P. Y., & Bakken, S. (2009). *A comparison of usability evaluation methods: heuristic evaluation versus end-user think-aloud protocol–an example from a web-based communication tool for nurse scheduling.* Paper presented at the AMIA Annual Symposium Proceedings.

Yi, M., Kim, J., Noh, D. Y., Lee, J. L., Yoo, K. Y., Hwang, K. T., Chung, H., D., . (2008). Evaluation of the satisfaction and usefulness of a web-based educational program for breast cancer patients. *The open medical informatics journal, 2*, 129.

Zaniolo, O., Iannazzo, S.,Pradelli, L.,Miravitlles, M. (2010). Pharmacoeconomic evaluation of tiotropium bromide in the long-term treatment of chronic obstructive pulmonary disease (COPD) in Italy. *The European Journal of Health Economics*, 1-10.

Zohar, E., Mandelbaum, A.,Shimkin, N.,. (2002). Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science, 48*(4), 566-583.

Zwart-van Rijkom, J. E. F., Leufkens, H. G. M., Busschbach, J. J. V., Broekmans, A. W., & Rutten, F. F. H. (2000). Differences in attitudes, knowledge and use of economic evaluations in decision-making in The Netherlands. *Pharmacoeconomics, 18*(2), 149-160.