Challenges encountered in the economic evaluation of medical devices

by

Ruth Pulikottil Jacob

Submitted for consideration for the degree of

Doctor of Philosophy by published work – Health Sciences

Warwick Medical School

University of Warwick

November 2016

Table of Contents

List of illustrations and tables

Tables

Figures

Acknowledgements

I would like to thank all the co-authors and my former colleagues from Warwick Evidence who have supported me during my publications- in particular Dr. Kandala, Ms. Freeman, Ms. Grove and Ms. Court. I would also like to thank Dr. Sutcliffe who was instrumental in igniting the idea of PhD by published work in me and putting me in touch with Dr. Madan, who has been a constant support for last 4 years of my PhD studies. I also thank Dr. Mistry whose contribution to my research as a line manager, co-author and supervisor was significant.

This thesis could not have been possible without the generous support and encouragement from Prof Clarke and Dr. Connock who have been a great source of inspiration.

Finally, I thank my family for their support throughout this venture.

This thesis is dedicated to my son John as a benchmark.

Submission declaration

I declare that the submitted material as a whole is not substantially the same as published or unpublished material that I have previously submitted, or am currently submitting, for a degree, diploma, or similar qualification at any university or similar institution. No parts of the works have been submitted previously for any aforementioned qualification.

List of publications

A list of the 4 papers submitted are shown below and the author's (Ruth Pulikottil Jacob) contribution is shown in Appendix A.

Title of Paper	**First Author and Co-authors**	**Publication and Date**
Has metal-on-metal resurfacing been a cost-effective intervention for health care providers? - A registry based study	First Author: **Ruth Pulikottil Jacob** Co-authors: Connock M, Kandala N-B, Mistry H, Grove A, Costa M, Sutcliffe P, Clarke A.	PLOS one November 2016
Cost-effectiveness of total hip replacements for women and men with hip osteoarthritis: comparison of devices with differing bearing surfaces and modes of fixation	First Author: **Ruth Pulikottil Jacob** Co-authors: Connock M, Kandala N-B, Mistry H, Grove A, Costa M, Sutcliffe P, Clarke A.	The Bone & Joint Journal April 2015
Comparative cost-effectiveness of HeartWare versus HeartMate II left ventricular assist devices used in the United Kingdom National Health Service bridge-to-transplant program for patients with heart failure	First Author: **Ruth Pulikottil Jacob** Co-authors: Suri G, Connock M, Kandala NB, Sutcliffe P, Maheswaran H, Banner NR, Clarke A.	The Journal of Heart and Lung Transplantation April 2014
Cost-effectiveness of left ventricular assist devices (LVADs) for patients with advanced heart failure: Analysis of the British NHS bridge to transplant (BTT) program	First Author: Aileen Clarke Co-authors: **Ruth Pulikottil Jacob**, Connock M, Suri G, Kandala NB, Maheswaran H, Banner NR, Sutcliffe P	International Journal of Cardiology February 2014

Word count

Summary: 740

Background: 2044

Rationale for this study: 698

Summary of the published work: 2096

Discussion: 4321

Total: 9,899

List of abbreviations

AIC	Akaike information criterion
ASA	American Society of Anaesthetologists
BTT	Bridge to Transplant
BTDB	British Blood and Transplant Database
CCA	Cost-consequence analysis
CE	Conformité Européenne
CeCoP	Cemented Ceramic on Polythene bearing surface
CeLCoC	Cementless Ceramic on Ceramic bearing surface
CeCoP	Cemented Ceramic on Polythene bearing surface
CeMoP	Cemented Metal on Polythene bearing surface
CEAC	Cost-effectiveness acceptability curves
EMBED	Economic model and price benchmarking
EQ-5D	EuroQol-5D
EU	European Union
HMII	HeartMate II
HRQoL	Health related quality of life
HT	Heart transplant
HTA	Health technology assessment
HW	HeartWare
HyPoM	Hybrid Metal Head on Cementless Hydroxyapetite Coated Metal cup
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient data
KM	Kaplan-Meier
LVAD	Left ventricular assist device
LYG	Life-years gained
MAGLEV	Magnetic levitation
MM	Medical Management
MHRA	Medicines & Healthcare Regulatory Agency
MTEP	Medical Technologies Evaluation Programme
NICE	National Institute of Health and Care Excellences
NHS	National Health Service
NJR	National Joint Registry
NYHA	New York Heart Association
PAS	Patient Access Scheme
PSS	Personal social service
PROMS	Patient reported outcome measures
QALY	Quality-adjusted life years
QoL	Quality of life
RCT	Randomised controlled trial
REMATCH	Randomised Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure
RS	Resurfacing

SHFM	Seattle Heart Failure Model
TA	Technology appraisal
THR	Total hip replacement
VAD	Ventricular assist device
WTP	Willingness to pay

1. Summary

This thesis concerns the linking together of the challenges encountered in the economic evaluation of medical devices and credible ways of performing economic evaluation in such a scenario. Although the standard methods of estimating cost-effectiveness have gained widespread acceptance, there are concerns around the methods for conducting economic evaluation in the health technology assessment of devices. Currently, the lack of appropriate comparators and evidence generation (i.e. quantity and quality of the clinical and economic evidence) have been identified as the main challenges.

The four publications associated with this thesis explore the validity and applicability of methods which have the potential to overcome these challenges. These methods include: undertaking economic evaluation by classifying devices based on their granularity and clinical relevance; construction of artificial comparator; time-to-event analysis; propensity score matching; covariate adjustments; and sensitivity analysis for device cost and health related quality of life (HRQoL) estimates. The first and second publications (studies I & II) draw upon the British Blood and Transplant Database (BTDB), which holds individual patient level data collected from the six designated British centres responsible for left ventricular assist device (LVAD) implantation and heart transplant (HT). The database holds the medical histories of patients with advanced heart failure on the waiting list for HT, and recipients of LVAD implants and post-HT between May 2002 and December 2011. The economic evaluation for LVADs used a Markov multi-state model to assess the cost-effectiveness of the devices. For this model, transition probabilities between health states were modelled using Kaplan–Meier (KM) time-to-event analyses with extrapolation beyond the observed data. Health outcomes were measured in quality-adjusted life years (QALYs) and NHS and personal social service (PSS) resource use and costs were included. Transition probabilities for the control arm were also estimated by constructing an artificial “control arm”. This artificial control arm was used to attain the statistical impact of a two-armed randomised trial. The two studies revealed that LVADs considered as a Bridge to Transplant (BTT) yields incremental cost-effectiveness ratio (ICERs) of £55,173 per QALY, when compared with Medical Management (MM). Although LVADs were not

considered cost-effective at the National Institute of Health and Care Excellence's (NICE) standard willingness to pay (WTP) thresholds, they clearly demonstrated an improvement in HRQoL and functional status for patients who survive implantation of a LVAD as a BTT. Our study estimated the cost-effectiveness of HW versus HMII for patients with advanced heart failure and showed that HW was more effective and cost-effective than HMII. The analyses demonstrated the credibility of economic evaluation of LVAD devices and show an acceptable degree of robustness in estimating the cost-effectiveness of second and third generation LVADs.

The third and fourth publications (studies III & IV) used individual patient data from the National Joint Registry (NJR) for patients with osteoarthritis undergoing total hip replacement (THR) and resurfacing (RS) between April 2003 and December 2012. The database contained several different combinations of prosthesis components and these components were stratified by frequency of use and on expert clinical opinion. This process identified five categories of THR prosthesis, which were compared against each other; and compared against RS. The economic evaluation of hip prosthesis also used a Markov model to assess the cost-effectiveness of the above mentioned comparisons. For this model, transition probabilities between health states were modelled using KM time-to-event analyses with extrapolation beyond the observed data. Health outcomes were measured in QALYs and NHS and PSS resource use and costs were used. To reduce selection bias due to the use of non-randomised data, we propensity (age-sex) matched patients from the NJR database. We identified that at a WTP £20,000 per QALY a cemented prosthesis with metal-on-polyethylene or ceramic-on-polyethylene bearings had the greatest probability of being cost-effective for all groups of age and sex over a lifetime. Furthermore, our study found that at a WTP of £20,000/QALY, only the Birmingham Hip (RS) device had a reasonable probability of being cost-effective, and even then only for the youngest group of men, but alternative THRs with ceramic bearing surface devices appeared about equally cost-effective. Our study also found that RS is not cost-effective over a lifetime for healthy patients and THR is shown to be the most cost-effective option for all women and men aged over 50 years old.

The approaches described above improve the credibility of the cost-effectiveness estimate of devices; thereby assist decision makers to make informed decisions as to adopt or reject the technology for use in the National Health Service (NHS).

2. Background

2.1. Medical device industry and role of regulatory bodies

The UK medical device industry is an innovative industry with a long history of developing safe products to improve people's health and benefit the health care sector. This industry was valued at $17 billion USD in 2013 and is characterised by approximately 3,300 medical manufacturers (ABPI, 2013), which represent some key manufacturers developing implantable medical devices, like Smith & Nephew, Johnson and Johnson, Medtronic and Thoratec Corporation to name a few (EMERGO, 2006). Medical Device Directive classify devices into four categories (class I, IIa, IIb, and III). Class III devices are mainly implantable devices and considered high risk and subject to additional conformity procedures (French-Mowat E, 2012). An active implantable medical device is defined by the European Union (EU) directive as "*any active medical device which is intended to be totally or partially introduced, surgically or medically, into the human body or by medical intervention into a natural orifice, and which is intended to remain after the procedure*" (EEC Council directive, 1990). The devices are regulated under the EU law, and in the UK the Medicines & Healthcare Regulatory Agency (MHRA) is legally delegated by the EU to regulate medical devices. Therefore, MHRA is responsible to undertake post-market surveillance, whereby monitor any reports of adverse incidents or associated problems with use identified in the post-marketing phase (MHRA, 2008). However, the pre-market assessment of medical devices are not conducted by MHRA, but through organisations called notified bodies, which are accredited by member states to assess medical devices to grant conformity approval i.e the Conformité Européenne (CE) mark. Currently, there are 76 notified bodies across different EU member states to assess conformity of medical devices, which then can be marketed across all EU member states.

2.2. Common differences between devices and pharmaceuticals

Though pharmaceuticals and devices share same common objectives "to improve people's health and benefit the health care sector", devices differ from drugs in their mechanism of action, as drugs directly interact with patients by generating biochemical reactions; but devices rely on their physicality for treatment. A new device would undergo continuous

alteration throughout the lifecycle of the product, which considerably reduces the average product lifetime of any device. For example, it has been estimated that there are 500,000 medical devices belonging to 10,000 generic groups with an average product lifecycle of 18 months (Parvizi N, 2014). On the contrary, the pharmaceutical industry requires more intense research to develop innovative chemical entities with great research and development spend providing longer average product life. However, the device manufacturer is only required to demonstrate the safety and performance of the new device, as part of conformity assessment to CE mark, and is not required to conduct RCT studies to demonstrate the clinical efficacy of the new device. This is radically different to how drugs are regulated, which requires a stricter, transparent trial design to demonstrate the safety and efficacy of the new product before being approved by the drug regulators (Govin, 2006). Moreover, the EU directives on medical devices regulate devices, in contrast to the case with pharmaceuticals, which are regulated through a centralised regulatory body like the European Medicines Agency, rather than through different notified bodies across different EU member states (Clinical trials regualtions, 2004).These differences hold the basis for the dissimilarity in the process, industry composition and technology assessment between drugs and devices (Chapman AM, 2014).

2.3. Reimbursement of health technologies in the UK

In the UK, decisions on reimbursement of technologies are undertaken through health technology assessments (HTA). HTA bodies like NICE, produce guidance to publically funded health care systems like the NHS in England; and are also endorsed as applicable by Wales and Northern Ireland on the effectiveness and cost-effectiveness of new and existing technologies. NICE produces guidance on all technologies (i.e. pharmaceutical, medical devices, diagnostics and public health) through four separate appraisal programmes: technology, medical technologies, diagnostics and public health (NICE, 2014).

Appraisal of medical technologies through the Medical Technologies Evaluation Programme (MTEP) was launched in 2009/2010. This is the newest NICE committee formed to develop medical technology guidance for suitable technologies. The technology is considered to be eligible to be appraised by the MTEP, if the technology has a CE mark or expected to receive regulatory approval and to be used in the NHS. Technologies appraised through the

MTEP committee use cost-consequence analysis (CCA) (NICE, 2011). CCA is a form of economic analysis which measures cost and benefits in a single unit and lists the incremental costs and benefits, without combining these results. At times, devices could be considered to be appropriate to be reviewed by other guidance producing programmes such as a multiple technology appraisal (MTA) or non-NICE programme (HTA programme with no funding mandate). Technologies appraised through the other evaluation programmes are appraised using cost-utility analysis, where the primary outcome is cost per QALY. QALYs provide a common currency to assess the survival and quality of life benefits gained from an intervention (Weinstein MC, 2009). The cost-effectiveness of the intervention is then expressed using an ICER, where the increased/decreased marginal costs are divided by the increased/decreased marginal gains in QALYs.

$$ICER = \frac{Cost_B - Cost_A}{QALY_B - QALY_A}$$

The NICE reference case also recommends decision-analytical models as a standard framework for cost-effectiveness analysis.

In summary, HTA bodies like NICE have adopted different approaches in evaluating medical devices due to acknowledgment of the fact that evaluations of devices are challenging in comparison to pharmaceuticals (Chapman AM, 2014).

2.4. Challenges encountered in the assessment of medical devices

As discussed earlier, medical devices are less regulated than pharmaceuticals, hence, the European Union launched the Methods for Health Technology Assessment of Medical Devices: A European perspective (MedtechHTA) project (Drummond M, 2016). The three main project recommendations were to align regulatory and HTA processes for evidence generation, by conducting joint scientific advice for regulatory and reimbursement bodies to optimise the study design; and conduct studies that enable data collection that jointly satisfy the requirements of regulatory and reimbursement bodies. Secondly, to improve existing methods for collecting and incorporating clinical and economic data to address the complexities associated with medical devices. Thirdly, identify the factors driving the diffusion of medical devices (Tarricone R, 2014). Thus, the recommendations from Medtech

HTA project could address the challenges encountered in the HTA of medical devices, however, there is still a strong need to understand and address the challenges encountered in the economic evaluation of medical devices.

2.5. Challenges encountered in the economic evaluation of devices

2.5.1. Feasibility of conducting clinical trials

Device industries are highly innovative and continuously changing, with mostly small and medium sized manufacturers, and a few large companies. The implications of having to perform a clinical trial program could have a substantial financial impact on small and medium sized companies. Moreover, data from observational studies are commonly used for regulatory approval in the EU, as randomised controlled trials (RCTs) are not mandatory for EU approval. Therefore, the minimal evidence required for a regulatory approval makes it less encouraging for manufacturers to run a clinical trial programme. Unlike the pharmaceutical industry, lack of patent protection and the short lifecycle of medical devices discourages manufacturers from conducting additional research for market access and are only incentivised to conduct adequate research required for pre-marketing and post-marketing surveillance.

2.5.2. Lack of comparator evidence

Randomising patients to intervention or control groups permits balancing for unknown or unmeasured covariates i.e. demographic variables, severity of disease condition and professional use/preference bias. However, randomising patients to a comparative clinical trial might not be always feasible for the following reasons: i) ethical issues with recommending subjects to sham procedures; ii) denying patient access to the most appropriate treatment; iii) the design of the device could impose difficulty to randomly assign subjects to intervention/control group; iv) repeating clinical trials for every design modification might be challenging; v) and identifying the right comparator and recruiting the required sample size could be problematic with clinical practice varying from centre to centre. Therefore, medical devices commonly use observational study design for the comparison of safety and effectiveness of devices.

2.5.3. Issues around generating good quality and quantity evidence

HTA of medical devices is hugely dependent on good quality evidence on clinical and economic effectiveness, and safety of the concerned device. The drug approval process requires appropriate, adequate and reliable evidence for marketing authorisation. However, such stringent pre-requisite of appropriate evidences are not mandatory for devices.

- *Confounding and blinding*

In the case of implantable medical devices, where the device is compared to medical therapy, it could be difficult to blind the subject and the investigator to the assigned intervention. In these situations, there could be no means of controlling for known and unknown variations across the groups, leading to inaccurate conclusions on device safety and effectiveness. The value of medical device is determined by the patient outcome and this outcome is directly related to the device performance and clinician contribution in implanting this device. Therefore, a superior outcome of a device is a culmination of device performance, surgical technique and operational efficiency of the device. Hence, it is important to adjust for any bias in user characteristics i.e. surgical technique, clinician skills and any learning curve effects on the study outcomes.

The treatment effect estimated using observational studies are subject to treatment – selection bias in which the treatment group could differ substantially from the control group. Several traditional adjustment methods (i.e. matching, stratification and covariate adjustment) has several limitations and propensity score matching is considered to be the most widely used method for bias reduction in comparison of a treatment to a non – randomised control group. Propensity score is defined as the probability of treatment assignment conditional on measured baseline covariates; and the treatment status is independent of measured baseline covariates (Rosenbaum PR, 1983). The three most common techniques of propensity score are matching, sub-classification and regression adjustment (D'Agostino, 1998).

- *Limitations in device class and its effect*

A drug class is defined as "*the class of compounds which share a similar structure and mechanism of action. As a result of their similar mechanisms of action, drugs of a particular class produce similar pharmacologic effects and clinical outcomes ("class effects")*" (Soares I, 2002). This definition could also apply for devices in which evidence from studies using one or more devices within a class is assumed to be the same as other devices of the same class. For example, THR can be classified by their fixation method, implant components and femoral head sizes (Clarke A P.-J. R., 2015). The DePuy Articular Surface Replacement (ASR), a Metal–on–Metal Hip Implant consists of ASR Acetabular Cup System (manufactured by DePuy Orthopaedics); Metal Transcend Articulation System (manufactured by Wright Medical Technology); and Ultima Unipolar Head and Adapter Sleeves (manufactured by Johnson & Johnson Professionals) (Ardaugh M B, 2013). Out of three components, two are manufactured by different companies, showing that a hip prosthesis is a unique combination of different components manufactured by different companies, and the device performance depends on one or more components across manufacturers. Therefore, to accept a priori that clinical outcomes noticed within a class could be extrapolated to other THR implants would not be a reasonable assumption to make. This could also infer to other devices, therefore, it is important to compare the clinical and cost-effectiveness of devices within and between device classes.

- *Less mature HRQoL data*

HRQoL is a standardised patient reported outcome measure used to assess the impact of the treatment on patient's quality of life (Weinstein MC, 2009). EQ-5D is the most preferred measure to calculate QALYs and is the outcome measure recommended by NICE for economic evaluation. However, EQ-5D may not always be reported and in such a situation mapping can be used to predict utilities for each health states using data from other health measures i.e. mapping of NYHA to predict EQ-5D utility values (Longworth L, 2011). Nevertheless, HRQoL measurements are not stipulated by regulators and reimbursement bodies, leading to its limited diffusion in research and decision making of

medical devices. Technologies appraised through the MTEP committee use a CCA and do not require QALY as a measure to assess device outcomes (NICE, 2011). Lack of a standardised approach to compare treatment benefits across different health-care programs prevents to establish whether the device represents an efficient allocation of resources within the NHS setting, thereby a major limitation.

- *Inconsistencies in pricing*

Medical devices are procured through different methods and they include: procurement by trusts independently; centralised procurement processes i.e. the NHS supply chain; and NHS collaborative procurement hubs established in some NHS regions. Each NHS trust has also formulated its own procedure to procure devices in accordance with their respective trust contract regulations (NHS Foundation Trust, 2014). Likewise, devices could also be procured and delivered centrally by the NHS Supply chain, which is owned and managed by the Department of Health, and offers effective procurement practices. To summarise, the model and approach to procure devices varies by trust and region, and there is no consistency in device price across the NHS (Department of Health, 2011).

- *Absence of mature clinical data on treatment benefit*

Time to event analysis are commonly used in economic evaluation to model the impact of treatment benefit and to define the treatment related health state. Survival modelling approaches most widely used in economic evaluations are restricted means analysis, proportional hazard modelling and parametric modelling. The restricted means analysis and proportional hazard model is only used when there is relatively little censoring in the survival data and when the extrapolation is only based upon the observed data. Censored data are one form of missing data, where patients are either lost to follow-up or might not experience the event of interest before the end of the trial. Therefore, the censored data prevents a simple estimate of mean survival from the trial outcome. However, clinical trial data are often censored and the final outcome of interest is not observed for the randomised population. Hence, in the presence of censoring it is unlikely to estimate mean survival based on trial data alone, therefore, parametric survival analysis is greatly favoured in economic evaluation because the model estimates the baseline hazard function, thus

drawing information from the whole data and dealing with issues of data censoring. Different parametric models can be used to extrapolate survival/time-to-event data to estimate transition probabilities for economic models. The most commonly used parametric models in economic evaluation are exponential, Weibull and Gompertz models.

Parametric model

Parametric models are models were a specific probability distribution is assumed for the survival function using the K-M estimate. For the economic analysis, several parametric models are fitted to the observed data and the model extrapolated based on the different modelled hazard function. For example, the observed data is shown in figure 1 and exponential fit to observed data is shown in figure 2.

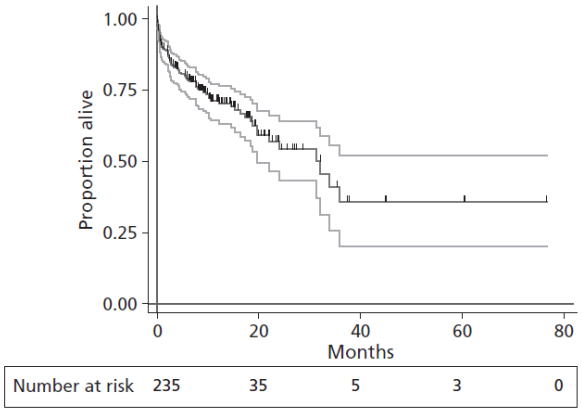

Figure 1 Illustrates observed survival and 95% CI while supported on a VAD

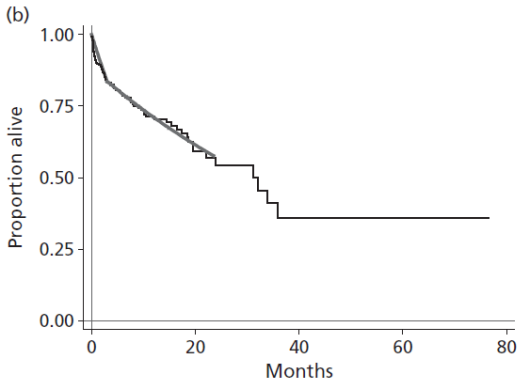

Figure 2 Exponential fits to observed data month 3 and from month 3 to 23 while supported by a VAD

An example of the different modelled hazard are described in table 1 and shown in figure 3.

Table 1 The parametric models

Modelled hazard	Description
Exponential distribution	The exponential distribution is the simplest parametric model for survival data because it assumes the hazard function is constant over time.
Weibull distribution	The Weibull distribution is more flexible than an exponential distribution, as the hazard function can either increase or decrease over time.
Gompertz	The Gompertz proportional hazard model is the most widely used model for human mortality.
Accelerated failure time (AFT) model	AFT model can be used when the hazard function changes direction, and allows us to measure the direct effect of an explanatory variable on the survival time.
Log normal distribution	The log normal distribution is an accelerated failure time model and has a hazard function which can be non-monotonic with time.
Bath-tub model	The failure rate function in a Bath tub model is characterised by a U- shaped curve comprised of three stages; an initial stage with a decreasing hazard function, a middle stage with a constant hazard function and a final stage with an increasing hazard function.
Flexible parametric model	A flexible parametric model using restricted cubic splines offers greater flexibility to model the baseline hazard compared to standard parametric models.

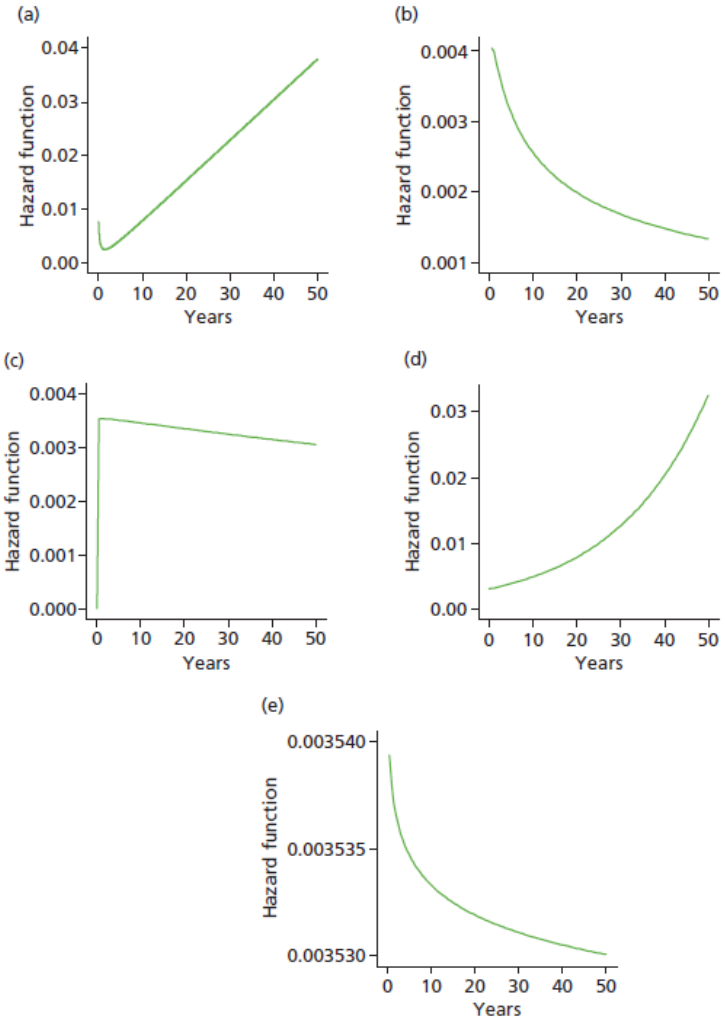

FIGURE 26 Differing modelled hazard on extrapolation beyond observation for THR HyMoP (category D) female patients aged < 65 years. (a) Bathtub model; (b) log-normal model; (c) log-logistic model; (d) Gompertz model; and (e) Weibull model.

Figure 3 a) Bathtub model b) log-normal model c) log-logistic model d) Gompertz model e) Weibull model from Clarke A et al 2015

The NICE technical support document on survival analysis recommends using parametric models to extrapolate data in health technology assessment (Latimer, N, 2011). In keeping with the NICE methods guide, it is important to assess the suitability of the survival model, the goodness of fit to the observed data, and the clinical and biological plausibility of the extrapolated data. The NICE Decision support unit (DSU) provides guidance on the model selection process, and have formulated a Survival Model Selection for Economic Evaluation Process chart to demonstrate the logical process needed to select a preferred model. However, in some instances the NICE reference case/DSU document would not seem applicable to all health technology assessment, where high-quality evidence is a challenge

to health economists, as seen with medical devices. Nevertheless, it is important to demonstrate that a logical and clear process was followed to select the preferred survival model, and the impact of model choice on cost-effectiveness results. The choice of a survival model could considerably vary the ICER, and therefore the decision to whether accept or reject the technology by the decision maker.

In summary, the challenges encountered in economic evaluation of medical devices are:

- Lack of appropriate comparators for the intended subject population
- Issues around generating good quality and quantity evidence due to
    - Limitations in device class and its effects
    - Confounding and blinding
    - Less mature HRQoL data
    - Inconsistencies in pricing
    - Absence of mature clinical data on treatment benefit

3. Rationale for this study

This thesis explores the challenges described in the previous section using two specific devices – hip replacement prostheses and left ventricular assist device.

3.1. Hip replacement prostheses

People with hip problems due to osteoarthritis are surgically treated with a THR or RS. Both these procedures involve replacing the hip joint with a synthetic body part called a hip prosthesis, which consist of a femoral stem, head and acetabular cup made from different materials (i.e. metal, ceramic or polyethylene); and these implant components are fixed permanently to the pelvis and femur using different fixation methods (i.e. cemented, cementless, hybrid or reverse hybrid prostheses) (Clarke A, 2015). Therefore, based on different fixation methods for the femoral stem and acetabular cup in THR, they are further classified into cemented, cementless, hybrid and reverse hybrid THR (See figure 1).

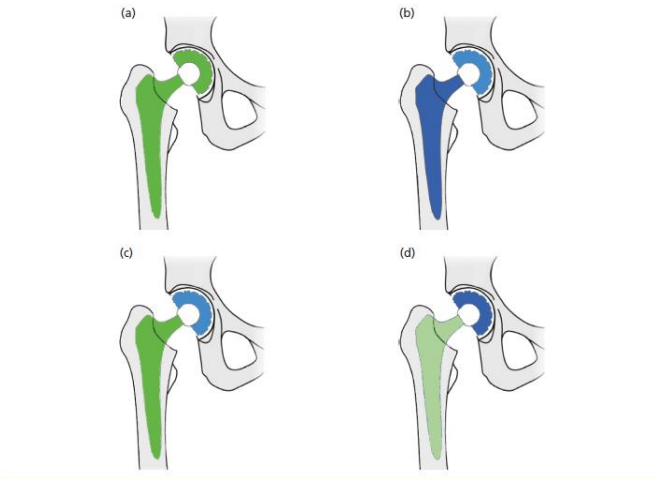

Figure 4 Different THR fixation methods (a) cemented THR (b) cementless THR (c) reverse hybrid THR (d) hybrid THR from Clarke A et al 2015

Based on the frequency of use, combination of bearing and articulation surface; and method of fixation, the THR can be also characterised into five frequently used categories (see table 1) (Clarke A, 2015).

Table 2 Summary of THR prosthesis by five frequently used categories

Device type	Components
A (CeMoP)	**M**etal head (cemented stem) **o**n **Ce**mented **p**olyethylene cup
B (CeLMoP)	**M**etal head (cementless stem) **o**n **Ce**ment**l**ess hydroxyapetite coated metal cup (**p**olyethylene liner)
C (CeLCoC)	**C**eramic head (cementless stem) **o**n **Ce**ment**l**ess hydroxyapetite coated metal cup (**c**eramic liner)
D (HyMoP)	**Hy**brid **M**etal head (cemented stem) **o**n cementless hydroxyapetite coated metal cup (**p**olyethylene liner)
E (CeCoP)	**C**eramic head (cemented stem) **o**n **Ce**mented **p**olyethylene cup

3.1.1. Hip resurfacing (RS)

Hip resurfacing is an alternative surgical technique to THR, mostly advocated for younger, active male patients with osteoarthritis (see figure 2) (NICE, 2000).

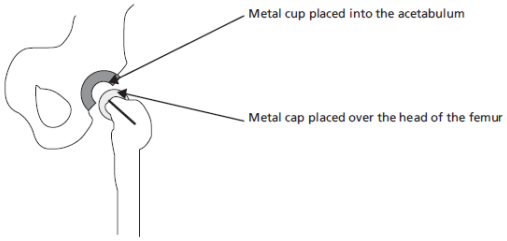

Figure 5 Hip resurfacing from Clarke A et al 2015

3.2. Left ventricular assist device (LVAD)

LVAD is a mechanical circulatory device used to pump blood from the left ventricle to the rest of the body in patients with heart failure waiting for HT. This device is further classified into first, second and third generation LVADs (see figure 3); first generation LVADs consist of a pulsatile volume displacement device to pump blood; second and third generation device use magnetic continuous flow rotary pump. (Sutcliffe P, 2013) For the purpose of this thesis, the focus would be on CE approved second and third generation LVADs.

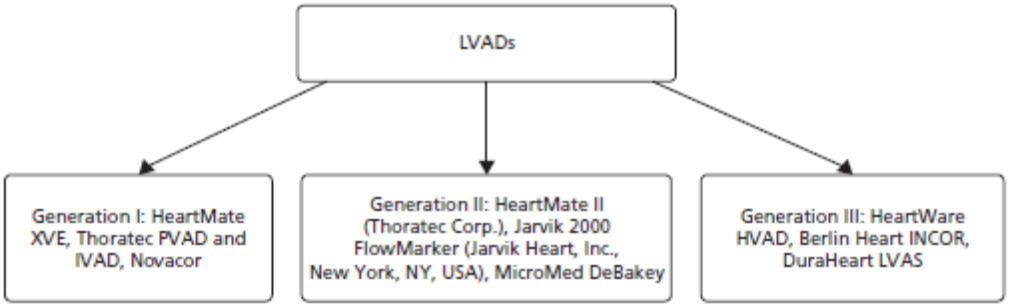

Figure 6 Types of LVADs from Sutcliffe P et al 2013

Against this background, we describe the key challenges faced in the economic evaluation of devices and present the causes of these challenges in Table 2.

Table 3 Challenges encountered in the economic evaluation of devices: Application to hip replacement prosthesis and left ventricular assist device

Challenges	Description	Causes
Lack of comparator evidence		

Feasibility of conducting clinical trial	No appropriate control for the intended subject population
	
	Standard of care for :
	• Patients with advanced heart failure - MM with inotrope
	• Patients with osteoarthritis - medical management, weight control and exercise.
		• Control group did not represent the subject population.
		• Historical control did not reflect current standard of care.
		• It was considered unethical to randomise subjects to intervention (VAD/Hip) because-
		  • It would be challenging to recruit subjects to receive medical management, when VAD/Hip device group were expected to receive a benefit.
		  • Extent of subject knowledge on VAD/Hip prosthesis for subjects awaiting surgery could be perceived to be denying treatment to control group subjects.

	Randomised clinical trial design is not mandatory to obtain market approval.	• Manufacturers of these devices have less incentive to conduct a reasonably powered RCTs, as it is not a mandatory requirement for pre or post–market assessment.
Issues around generating good quality and quantity evidence		
Limitations in device class and its effect	Recommendation of devices based on class effect could be flawed.	• Technology advancements • Improvements in clinical management of patients over time • Second and third generation VADs showed important differences in the mechanism of action, though approved for the same clinical indication. Similar difference was also noticed in hip prosthesis.
Confounding and blinding	Difficulty to blind devices due to difference in design and procedure	• Both VADs and Hip prosthesis were compared to medical management; therefore, it was difficult to blind the subject and the investigator.

		• Surgical technique, clinician skills and any learning curve effects could act as a confounding variable and pose challenges. • Person's behaviour on hip prosthesis will affect the device performance i.e. patients physical activity could lead to friction and wear out the prosthesis.
Less mature HRQoL data	Immature HRQoL data to support device outcome.	• HRQoL data not mandatory to support reimbursement • Lack of QoL from RCTs
Inconsistencies in pricing	There is no centralised procurement process for devices and price paid for the same device could differ across different NHS trusts.	• VADS and Hip prostheses are not procured centrally; therefore, devices are purchased through different processes in different NHS hospitals. Thus, the cost of the same device differs across centres, where special rebates and confidential volume based discounts are negotiated and offered, thereby reducing the generalisability of device cost.
Absence of mature clinical data on treatment benefit	Lack of long term data on median survival after VAD, MM with inotrope and HT.	• The BTDB had missing values for some of the patients. • The database had no information about patients prior to VAD implantation. • There were no report of adverse events based on the type of VAD device.

3.3. Aim

This thesis aims to identify and assess the most credible way of conducting economic evaluation when faced with the challenges mentioned earlier, using hip replacement prosthesis and left ventricular assist devices as examples.

3.4. Objectives

- To determine and describe the challenges encountered in the economic evaluation of medical devices.
- To describe methods for robust economic evaluation of devices.

3.5. Methods

3.5.1. Data sources

This research used the BTDB, which holds IPD from the six designated British centres responsible for VAD implantation and heart transplant. The database, overseen by the British Cardiothoracic Transplant Audit Group, holds the medical histories of patients with advanced heart failure on the waiting list for HT; and recipients of VAD implants and HT between May 2002 and December 2011.

Likewise, this research also used IPD from the NJR for England and Wales for patients with osteoarthritis undergoing THR and resurfacing between April 2003 and December 2012.

3.5.2. Analysis

The economic evaluation for each type of device (LVADs and THR/RS) used a Markov multi-state model to assess the cost-effectiveness of the device. For these models, three sets of model inputs were used: transition probabilities, costs and quality of life (utilities) associated with each health state. Transition probabilities between health states were modelled using KM time-to-event analyses with extrapolation beyond the observed data. Health outcomes were measured in QALYs and NHS and PSS resource use and costs were used.

3.5.3. Main outcomes

- *Observed and modelled survival*

BTDB patient level data was used to model time to death on VAD support and MM, time to heart transplant and cumulative incidences of HT.

NJR data was used to model lifetime revision for men and women with THR and RS.

- *Other main outcomes*

Lifetime costs, lifetime QALYS and the probability of a device being cost-effective at the NICE WTP threshold.

4. Summary of the published work

The overall aim of this thesis is addressed using four publications, and the body of research are linked to each other as shown in figure 4. First, we published a cost-effectiveness analysis of LVADs for patients with advanced heart failure, based on a Markov model. A second publication compared the cost-effectiveness of third generation VAD (i.e. HW) versus second generation VAD (i.e. HMII). The above two publications were used to address the challenges encountered in economic evaluation of VAD devices. In the next two publications, we used a Markov model to explore the cost-effectiveness of five frequently used categories of THR; and cost-effectiveness of metal-on-metal resurfacing versus THR in patients with osteoarthritis.

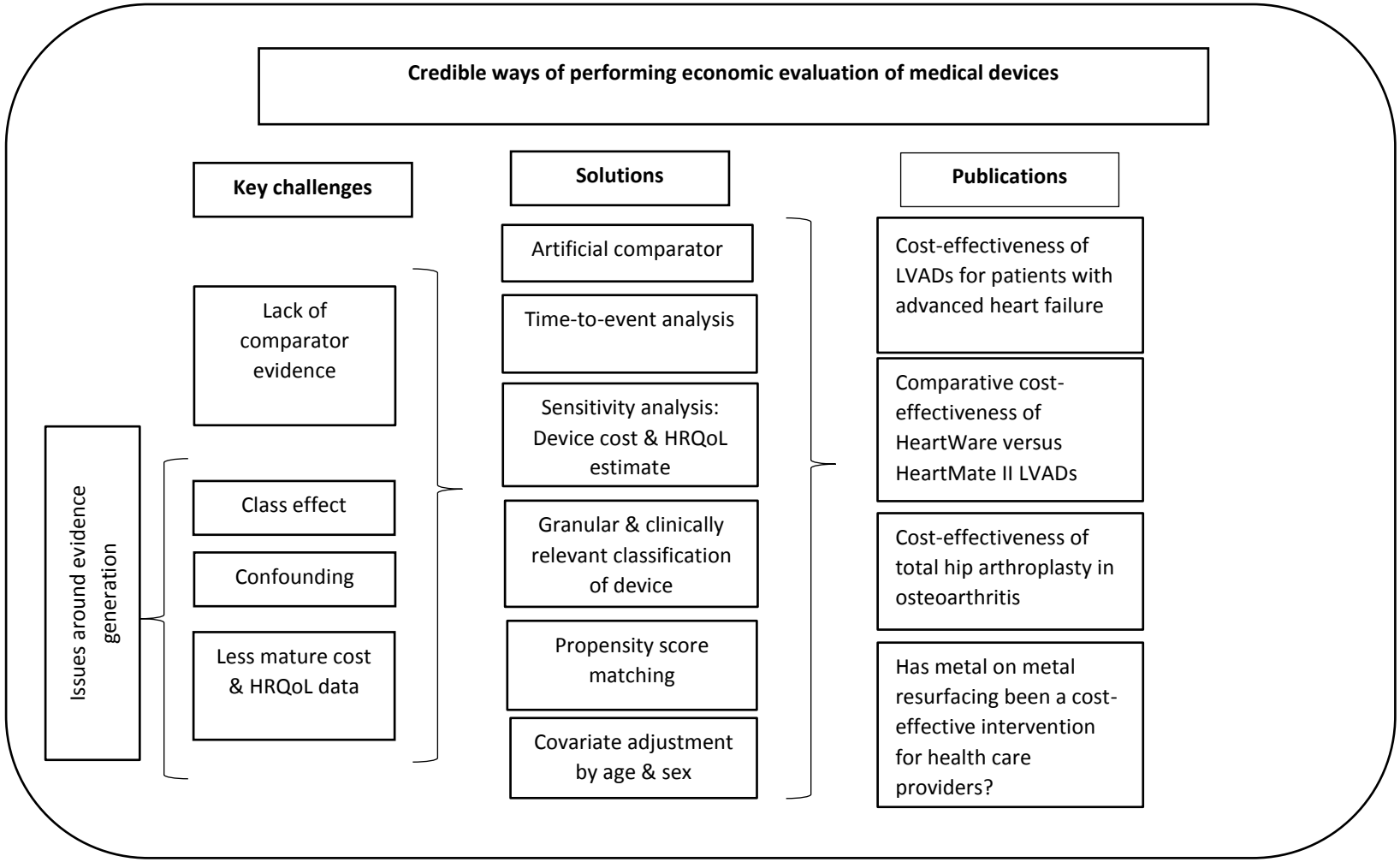

Figure 7 Logical representation of the publications used to illustrate credible ways of performing economic evaluation of medical devices

4.1. Study 1- Cost-effectiveness of left ventricular assist devices (LVADs) for patients with advanced heart failure: Analysis of the NHS bridge to transplant (BTT) program (Clarke, 2014).

Introduction

LVADs are an implantable mechanical device used as a BTT for patients with advanced heart failure. Based on the specific characteristics of the device, LVADs are further classified into first, second and third generation devices. In 2002, the NHS BTT program was formed to support VAD therapy in the UK. The clinical and cost-effectiveness analysis of first generation LVADs showed them not to be cost-effective at NICE WTP threshold.

Objective

To investigate the cost-effectiveness of second and third generation LVADs as a BTT, compared to MM with inotrope support in the NHS BTT program.

Methods

Clinical outcomes

The BTDB contains individual patient medical history for all advanced heart failure patients on the waiting list for HT; LVAD patients; and HT patients. All patients who received a second or third generation LVAD between May 2002 and December 2011 as a BTT, bridge to decision for HT, or bridge to myocardial recovery were included in this analysis. Parametric models were used to extrapolate beyond the observed data.

Economic model

A semi-Markov multistate structure was built in which each patient exists in one of three mutually exclusive health states: alive on LVAD or MM support; alive after HT; dead - the absorbing health state. Individual patient data (IPD) was used for KM time-to-event analyses to estimate the transition probabilities for recipients of LVAD implants as BTT, recipients of heart transplants and all patients on MM for advanced heart failure. The model was run for a lifetime horizon (50 years), and shorter time horizons of 3 and 10 years were explored in sensitivity analyses. Probabilistic sensitivity analysis was undertaken to

account for uncertainty in individual patient outcomes and parameter uncertainty. The costs and benefits were evaluated from the UK NHS perspective. An annual discount rate of 3.5% was applied to costs and benefits.

Health outcomes

We used the New York Heart Association (NYHA) patient level data from BTDB to obtain EQ-5D utility scores.

Cost estimates

The device cost for second and third generation LVADs were obtained from six NHS hospitals responsible for implementation of BTT program. LVAD implant, heart transplant procedure and post-operative support costs were based on a previous analysis and inflated to current prices by applying the projected Health Service Cost Index (PSSRU, 2010).

Results

Clinical outcome

235 patients were registered on the database as having had a second- or third- generation device. Amongst these, 125 received HW, 82 HMII, 23 Jarvik 2000 FlowMaker, and 5 Micromed Heart Assist devices, respectively. At 2 months, survival was apparently similar for patients receiving an LVAD compared to those receiving MM with inotropes (LVAD 89% vs. MM 83%). Modelled survival and time to transplant varied considerably for patients on MM with inotropes relative to patients supported with an LVAD. Survival with a transplant was good, with over 75% survival at 24 months.

Economic analysis

The ICERs for the 3 year, 10 year and lifetime time horizons were £120,510/QALY (95% CI: £79,560 to £251,285), £67,119/QALY (95% CI: £38,756 to £116,681) and £53,527/QALY (95% CI: £31,802 to £94,853) respectively. According to the current UK threshold of £20,000 to £30,000/QALY recommended by NICE, LVADs cannot be considered to be cost-effective. At a higher WTP threshold of £50,000 /QALY (end of life criteria), the probability that LVADs are cost-effective is 0%, 13.3% and 40.7%, over the 3 year, 10 year and lifetime horizons respectively.

4.2. Study II- Comparative cost-effectiveness of HeartWare versus HeartMate II left ventricular assist devices used in the United Kingdom National Health Service bridge to transplant program for patients with heart failure (Pulikottil-Jacob, 2014).

Introduction

HMII (HMII, Thoratec, Pleasanton, CA) and HW (HW, HeartWare International, Framingham, MA) are the two most frequently used LVADs in the UK. HMII is a second generation magnetic continuous flow rotary pump implanted below the diaphragm in the abdomen. HW is a third generation LVAD with a small centrifugal pump implanted in the pericardial space. The clinical benefits and risks of LVAD implantation has been extensively studied, however, there have been no comparative studies between the two devices. Therefore, it is important to evaluate the clinical and cost-effectiveness of these devices to support the right choice for the indicated population.

Objective

To investigate the clinical outcomes for individuals implanted with HMII and HW devices in the UK; and to estimate the cost-effectiveness of the HW compared to the HMII for patients treated in the UK NHS BTT program.

Methods

Clinical outcomes

All patients who received an HW or HMII as a BTT, bridge to decision for HT, or bridge to myocardial recovery between May 2002 and December 2011 were included in this analysis. Principal outcomes recorded in the BTDB after implantation of an LVAD were HT, explantation of the device, remaining alive with the originally implanted device, and death while supported with the originally implanted device. These outcomes were investigated with KM time-to-event analyses and with cumulative incidence methodology for competing outcomes.

Economic model

A semi-Markov multistate structure was built using Excel. Each patient in the model exists in 1 of 3 mutually exclusive health states: alive on LVAD; alive after HT; dead. IPD was used for KM time-to-event analyses to estimate the transition probabilities for death amongst those

who received a LVAD or an HT, and time to HT. Parametric models were used to extrapolate beyond the observed data.

Health outcomes

NYHA classification of patients in BTDB was used to determine EQ-5D utility scores.

Cost estimates

The mean cost for the HW was £80,076 and ranged from £76,774 to £98,160; the mean cost for HMII was £89,830 and ranged from £78,877 to £126,702.

Results

Clinical outcomes

Observed survival was superior for the HW recipients (63%) compared with HMII (40%) at 750 days. According to KM analysis at 6 months, there was little difference in survival between groups. By 12 months, an estimated 24% of HW patients and 34% of HMII recipients had died.

Economic analysis

The economic analysis found individuals who received HW incurred more costs but accrued more life-years and QALYs than did HMII recipients. The base-case probabilistic analysis indicates that implanting HW compared with HMII would cost the NHS on average an additional £24,379 over the lifetime of an individual (95% CI, dominates to £108,940), with an ICER of £20,799 per QALY gained (95% CI, dominant to £79,837).

4.3. Study III- Cost-effectiveness of total hip arthroplasty in osteoarthritis: a comparison of devices with differing bearing surfaces and modes of fixation (Pulikottil-Jacob R, 2015)

Introduction

THR is a commonly undertaken procedure, and with increasing demand for THR more attention is being paid to its cost-effectiveness and the choice between designs and modes of fixation. For osteoarthritis, more than 80,000 surgeries are performed each year in the UK, and over 100 different brands of prosthesis are available, manufactured by at least 20 companies with varying performance and cost. It is worth noting that each brand has individual purchasing arrangements with orthopaedic centres, rather than a NHS wide procurement. Hence, identification of the most cost-effective device would allow significant cost savings for the NHS.

Objective

To examine and compare the cost-effectiveness of frequently used combinations of components in THR, including the type of fixation and bearing surfaces.

Methods

Clinical outcomes

Patients undergoing THR for osteoarthritis between April 2003 and December 2012 were selected from the NJR for England and Wales. The registry provided IPD of hip prosthesis and combinations of components by usage. This study focussed on the most commonly used components by stratifying the various combinations of components using an iterative cross-tabulation procedure. This procedure selected four combinations based on the highest recorded usage, and one additional combination was added based on expert opinion. KM analysis was used to estimate time to revision for the above five prosthesis combinations. NJR data was stratified by sex and controlled for age. Parametric distributions were fitted to KM data and models extrapolated to a lifetime horizon.

Economic model

A multi-state Markov model was built to model the cost-effectiveness of each THR combination. In this model, patients exist in one of four mutually exclusive states: successful primary THR; revision surgery (patients can move into this state more than once but stay in

this state for one annual cycle only); successful revision surgery; death (patients may enter this state both due to operative mortality or due to death from other causes). Model inputs were yearly transition probabilities of converting from successful THR to revision surgery or death; costs associated with different prostheses, hospital costs and follow–up costs.

Health outcomes

Quality of life data were obtained from Patient reported outcome measures (PROMs) database for patients who had a THR from January 2009 to December 2012. EQ-5D data were available for patients post successful THR and revision THR by age and sex.

Cost estimates

Costs of prostheses were obtained from NHS supply chain and mean cost of each prosthesis was estimated using list price from the most commonly used suppliers.

Results

The NJR included records of 386,556 patients with osteoarthritis, and 62% of osteoarthritis patients accounted for five frequently-used combinations of prosthesis (239,089 records). The overall rates of revision were higher in men than women. The distribution of age varied between categories, but was similar for men and women within each category. In all categories, apart from category CeCoP, the difference in rates by sex was statistically significant ($p<0.05$) both with and without stratification for age. Revision rates for the lifetime model were lowest for category CeCoP and highest for category CeLCoC, irrespective of sex and age.

Economic analysis

The base-case results for men and women aged 60, 70 and 80 years showed marginal differences in costs and QALYs between the five different types of prosthesis. Although, we used alternative cost and QoL inputs in a sensitivity analysis, the difference in cost-effectiveness between different types of prosthesis were minimal.

4.4. Study IV- Has metal on metal resurfacing been a cost-effective intervention for health care providers? - A registry based study (Pulikottil-Jacob R, 2016)

Introduction

RS accounted for about 7.5% of UK implants in 2007, but due to poor revision performance and concerns about metal debris, the use of RS had declined in 2012 to only about a 1% share of UK hip procedures. Currently, some still advocate RS for young and active men with large femoral head size. The NICE guidance for hip replacement recommends a 10 year revision rate benchmark of <5%. Like THR devices, RS devices produced by different manufacturers differ with respect to revision rate, and currently RS devices from sixteen different manufacturers have been used in the UK. Therefore, it is worth exploring whether implantation of any RS devices or the best RS devices in highly selected patients such as active young men with large femoral head size could be considered cost-effective.

Objective

To assess the cost-effectiveness of RS versus THR in device-patient combinations likely to satisfy the NICE benchmark of less than 5% revisions at 10 years.

Methods

Clinical outcomes

The NJR for England and Wales supplied IPD for RS and THR surgery for osteoarthritis recorded from April 2003 to December 2012. A flexible parametric modelling method was used to identify combinations of patients and devices that had satisfied the NICE benchmark. Data were stratified according to sex and pooled by device manufacturer and patient subgroups. All THR devices have shown revision rates within the NICE benchmark, hence, the five most frequently used THR devices were selected for comparison with RS on the basis of frequency of use of components.

Economic model

A multi-state Markov model was built to model the cost-effectiveness of the most used THR for comparison with RS. Model inputs were yearly transition probabilities for time to revision and to death, costs, and utilities of health states. The age-related utilities were assumed to be the same for the comparison of RS with THR.

Results

Clinical outcomes

KM analyses indicated that for all six manufacturers, men experienced lower revision rates than women. Flexible parametric models predicted a low probability of satisfying the 10 year NICE benchmark. The only device delivering an overall revision rate within or near the 10 year benchmark was the Birmingham Hip. For women, no manufacturer's device delivered a revision rate within the benchmark. All exceeded 5% revision within only 5 years, and by 10 years, predicted rates in women were all greater than 10%. For men the revision rate was poor for most manufacturers and flexible parametric models predicted a low probability of satisfying the 10 year NICE benchmark.

Economic analysis

For the base-case economic analysis we compared the most used RS (Birmingham hip, within NICE benchmark) with the most used THR device (CeMoP) in male grade ASA 1 plus ASA grade 2 patients. This analysis indicated that Birmingham Hip accumulated fewer lifetime QALYs for 40, 50 and 60 year old men, lifetime costs were greater for RS by at least £2,900 and RS was dominated by THR for 60 year old men, and the probability that RS was cost-effective remained low (<4%) compared to THR at a WTP of £20,000/QALY. The probability that RS was cost-effective was < 6% at a WTP of £20,000/QALY for the Biomet RS and Finsbury devices when compared with the best performing THR device (CeCoP) in male grade ASA 1 plus ASA grade 2 patients.

5. Discussion

This thesis tries to address the several challenges encountered in the economic evaluation of medical devices and ways to deal with these challenges. This enables estimates of cost-effectiveness to be derived in situations where sufficient comparative data, utility and cost data are lacking to do standard economic evaluation. However, the challenges discussed here are not exclusive to devices and could be applicable to both drugs and devices, due to certain overlaps noticed and accepted across both the groups.

5.1. Lack of comparator evidence

Lack of comparator evidence is a major challenge in the assessment of clinical and cost-effectiveness of devices. We propose constructing an artificial "control arm" to overcome this challenge. The survival estimates for the control arm were explored using the application of the Seattle Heart Failure Model (SHFM) to the baseline characteristics of the intervention group (constructing an artificial "control arm") (Levy WC, 2006). This artificial control arm could be used to attain the statistical impact of a two-armed randomised trial. Therefore, in our analysis, we constructed an artificial MM using patient characteristics of VAD patients, so VAD patients could act as their own control. We acknowledge that this approach has its own limitations, but in absence of an appropriate comparator this forms the best approach we could use to test the robustness of patient selection in the economic analysis. However, similar method have been used in the past to draw causal inferences from the study of a treatment or intervention in an event of a lack of an experimental design (Schaffer et al, 2009). This study concluded that SHFM scores most accurately predicted survival on MM and could be used to identify patients who might benefit from LVAD implant (Schaffer JM, 2009). Likewise, Aaronson et al developed a predictive model (i.e. the heart failure survival score) to accurately identify individuals with advanced heart failure and most likely to require urgent transplantation (Aaronson KD, 1997). However, comparison with artificial controls can yield biased results due to sampling bias. Nonetheless, we undertook extensive sensitivity analysis to test the uncertainty associated with our comparator group.

Firstly, we sort expert opinion, and clinical advisors asserted that the median survival observed in real life was close to survival observed for MM patients in the REMATCH trial. The REMATCH trial (Rose EA, 2001) was the only multicentre RCT study for patients with end-stage heart failure who received VAD as long-term circulatory support. Therefore, in view of a possible overestimate of survival for the BTDB inotrope control group, the MM group from the REMATCH trial (Stevenson LW, 2004) was explored as a control group in a sensitivity analysis. 75% of patients in the REMATCH trial (Stevenson LW, 2004) were treated with inotropes and was considered to be too unwell for HT with median survival of 4.9 versus 9.1 months for the BTDB inotrope control group. Hence, the MM group from the REMATCH trial (Stevenson LW, 2004) was considered to be the optimum MM group due to older age and higher disease severity. This analysis showed that the ICER barely changed from the base-case results, therefore, revealing that the base-case ICER was reasonably robust. The reason for this is that even though the poorer survival of the MM arm results in an increase in the difference in QALYs between BTT and MM, this poorer survival also results in lower costs in the MM group. Thus, an increase in the difference in costs between BTT and MM tend to cancel out the difference in QALYs when calculating the ICER.

Secondly, we identified appropriate comparators based on their clinical relevance and granularity of the prosthesis. This iterative process enabled us to categorise the five most frequently used THR prostheses for patients with osteoarthritis. However, publication by Pennington et al (Pennington M, 2013) compared three types of prosthesis i.e. cemented, cementless and hybrid prostheses for THR and reported that hybrid prostheses was the most cost-effective device. Our results varied from the published evidence and showed that the lifetime rates of revision were lowest for CeCoP and highest for category CeLCoC. This difference in result could be due to the difference in methodology used to examine the types of prosthesis, as Pennington et al (Pennington M, 2013) did not examine the differences in the bearing surface.

Finally, we used propensity score matching technique and covariate adjustment to address biases encountered in the economic evaluation of medical devices for the following reason- RS implants are widely perceived to be suitable for younger active patients (Pollard TC, 2006). There is no long-term RCT data to validate this benefit; however, the age distribution of patients implanted with RS supported this perception i.e. NJR database showed that most RS patients were male (70%) with mean age of 55 years compared to

30% of THR male recipients with mean age of 61 years. We noticed observed difference in sex mix of THR and RS population; twice the proportion of women received THR than RS. The literature also indicated that the revision rate for RS was higher in women than in men. Hence, the difference in age and sex between RS and THR patients was a major limitation of our analysis, and the only viable option to adjust for this difference was to model THR revision rates outside of the observed age distribution i.e. controlled for age as a covariate. Therefore, we used propensity matching to stratify data according to age and sex; and compared devices by their manufacturer, femoral head size, patient ASA grade and age at the time of primary intervention. In this comparison our results indicate that RS is unlikely to be cost-effective compared to THR at a WTP of £20,000/QALY. Recent trends in clinical practice suggest that RS is becoming less popular among patients with severe arthritis of the hip.

To summarise, in circumstances where head-to–head randomised comparison of devices are absent; artificial comparator, granular and clinical relevant classification of prosthesis, and sensitivity/scenario analysis assists to improve the credibility of the cost-effectiveness estimate.

5.2. Limitations in device class and its effect

Devices undergo constant modification from their initial form to overcome initial shortcomings. Therefore, there is a need to clearly distinguish between devices by class/generation. It could be clearly seen from our analysis that second and third generation LVADs have improved survival prospects in comparison to first generation devices. The difference between the second and third generation devices is that third generation device uses non-contact bearing also known as magnetic levitation (MAGLEV) which allows rotation without fiction or wear, whereas second generation devices use contact bearings. Therefore, it is broadly believed that third generation devices could reduce thrombotic events and improve durability (Rodriguez L, 2013). Although, the new technology has been widely used for the management of patients with end stage heart failure, but there have been no good quality randomised/non-randomised trials evaluating the benefits of second versus third generation LVADs (Girling AJ, 2007). The technology has not been fully evaluated using prospective studies, therefore, doubts do remain in the

technological advancement of third generation pumps (i.e. MAGLEV), if they could generate significant benefit.

Our study findings support clinicians and decision makers to make more rationalised decision on which of the different LVADs to consider within the NHS-supported BTT program. Had we pooled the LVADs together with the aim of estimating cost-effectiveness relative to MM might have seriously mislead clinicians/decision makers to reject LVADs based on poor performance.

The NJR dataset contained 387,667 records of THR and 31,222 patients' records of RS. Comparing all THR and RS categories with each other would have been an impractical task to consider, therefore, we compared between five identified THR categories; and the best RS devices that satisfy the NICE benchmark (i.e. 5% revision rate at 10 years) versus five frequently used THR categories. Our analysis clearly showed that the five THR categories differed in the observed revision rates. Ceramic head (cemented stem) on cemented polyethylene cup (CeCoP) had the lowest observed revision rate, however, the difference in cost and QALYs were minimal between the two categories.

NICE technology appraisal (TA2) on the selection of prostheses for primary THR did not differentiate THRs based on their bearing surface and fixation, and recommended cemented prostheses for primary THR (NICE, 2000). Conversely, TA44 recommended metal–on–metal hip RS in June 2002 (NICE, 2002), however, in June 2012 MHRA issued medical alert for all metal-on-metal hip RS implants following concerns on the revision rate (MHRA, 2012). Identification and distinguishing devices across different RS prostheses by manufacturer and fixation method could have prevented recommending such devices for continued use.

Therefore, in our study, we classify devices based on their granularity and clinical relevance and provide a clear indication of important differences across different categories of hip prostheses.

5.3. Absence of mature HRQoL and cost data

The lack of individual patient level EQ-5D data to estimate QALYs could impede robust economic evaluation of LVADs. The BTDB did not provide EQ-5D data, therefore, we used HRQoL data from the literature. Two sources of EQ-5D utility scores were available from published literature (Sharples LD, 2006) (Gohler A, 2009), and the evidence from these literatures highlight a strong relationship between NYHA class and utility (i.e. NYHA class deterioration was associated with utility loss). Therefore, we used NYHA class recorded for patients at registration; 3 months post LVAD and heart transplant to estimate EQ-5D utility scores. Nonetheless, we also undertook univariate sensitivity analysis using utility scores from Sharples et al (Sharples LD, 2006); however, the ICER barely differed from the base-case.

Outcomes for the primary THR, RS and revision THR were modelled in QALYs. Utility data were obtained from the PROMS database, however, we were unable to link NJR and PROMS dataset by patient identifiers. In light of this limitation, we adjusted EQ-5D scores by age and sex to overcome population differences. Based on this approach, our analysis demonstrated slightly better utility scores for men than women. The base-case results for ten year and lifetime horizons for men and women aged 60, 70 and 80 years showed very small differences in accumulated QALYs between prosthesis categories. These small differences were driven by the disutility associated with revision. Although, we undertook extensive sensitivity analysis to test the robustness of the utility data, we acknowledge lack of mature HRQoL to be a major limitation in our study.

The main source of cost inputs in our economic evaluation of BTT with LVADs compared to MM was from Sharples et al (Sharples LD, 2006). It is worth mentioning that the cost data reflected in the Sharples et al (Sharples LD, 2006) publication were based on a mixture of first and second generation LVAD devices. Although, our economic analysis were based on second and third generation devices; we still used resource use data from Sharples et al (Sharples LD, 2006), as it was the only cost data available from the NHS perspective. We obtained the cost of LVADs from five centres across the UK operating the BTT program. The centres with low number of LVAD usage showed greater purchase price in comparison to centres associated with a greater number of LVAD implants. Moreover, the cost towards device maintenance (i.e. replacing batteries, cables or any hardware associated with the

device) was largely unclear, as some centres specified that the maintenance cost would be included in the purchase agreement; some centres reported the cost to be trivial; and two centres reported £4,000 per year towards cost of maintenance. Another limitation was lack of short and long-term cost data on adverse events; and evidence on requirement of LVAD replacement. All these model assumptions were explored in the sensitivity analysis. Our results indicate that the cost of the device was a key driver of cost-effectiveness, as a 15% reduction of device cost reduced the ICER threshold to £50,000 per QALY; and a substantial reduction in device price (i.e. 76%) was required to reduce the ICER to £30,000 per QALY. Our study findings support clinicians and decision makers to identify the maximum price they may be willing to accept for LVADs within the NHS-supported BTT program.

Cost estimates of THR and RS prosthesis by type were obtained from two sources the NHS Supply Chain (June 2013) and from the respective manufacturers. These sources differed considerably, making it difficult to ascertain what would be the approximate price paid by NHS trust in reality. A price of prosthesis can vary from £800-£2150 per trust and there is a clear lack of transparency and consistency over pricing of prosthesis, as price paid by trusts vary considerably due to wide variation in discounts based on individual purchase agreements (National Audit Office, 2000). Report by the National Audit Office state that three quarters of trusts have negotiated a price discount for hip prosthesis based on the volume of purchase (National Audit Office, 2000). In 2014, NJR introduced the Economic–model-and–price-benchmarking (EMBED) service to support NHS commissioners to reduce their orthopaedic implants spend without compromising on patient outcomes. EMBED is expected to provide NHS trusts with an overview of their usage and total spend on hip prosthesis in comparison with other NHS trusts, as a means of cost-saving (National Joint Registry, 2014). However, even now (November 2016?) no such scheme is in place to support procurement of VAD implants across different NHS trusts.

In 2016, NHS England launched the centralised supply chain for the procurement of high cost tariff excluded devices, and NHS Supply Chain was appointed to operate this newly centralised supply chain on behalf of NHS England (NHS England, 2016). Currently, this has been only rolled out to high cost tariff excluded devices, and one would anticipate similar services would be beneficial to address the inconsistency noticed over pricing of other devices i.e. devices included in the National Tariff Payment System.

5.4. Absence of mature clinical data on treatment benefit

We applied time-to-event analysis to improve the robustness of the economic evaluation. Our base case analysis revealed that the ICER was critically dependent on the probability of receiving a heart transplant. Time to HT was very different between BTDB MM and BTT patients, as the median time to transplant in MM patients were 3.25 months compared to 44.7 months for LVAD patients. Moreover, our clinical expert confirmed that MM patients are prioritised over LVAD for transplant when donor heart is available. Hence, time-to-event analysis was used to model a high probability of receiving a HT for the MM group, than patients supported on LVAD. For the MM patients, the log-normal distribution provided the lowest AIC and the best fitting model, however, for the BTT patients, exponential distribution was the best fitting model. Therefore, the survival benefits were estimated using different possible model fits (i.e. log-normal fit to time to HT for MM patients; and exponential fit to time to HT for BTT patients) and extrapolated to predict long term survival. This analysis showed that inclusion of different model fits to the observed data changed the resulting cost-effectiveness estimate i.e. lifetime ICER for LVAD was considerably larger, and MM was more effective and less costly than BTT. Although, the DSU guidance for modelling time to event recommends to adopt the same parametric form for the intervention being compared, however, our results demonstrate that for observational data, parametric fits for different interventions groups may not be well described by a single parametric form (Latimer, N, 2011). Thus, further research is needed to support the relaxation of this assumption.

Likewise, study III highlights the application and effect of different parametric models on cost-effectiveness estimates. In the base case analysis, bath tub models were used to predict life time revision rates for 60, 70 and 80 year old patients. Bathtub models predict an initially decreasing hazard for revision (i.e. after primary surgery) followed by an increasing hazard as the prosthesis ages, and the quality of the patient's bone deteriorates. We used a direct approach in which a single bath tub model was employed, whereas Pennington et al. 2013 combined two Weibull models (one for early and one for late THR failures) in order to achieve the bath tub hazard profile; this may have led to difficulties in extrapolation so that "capping" of revision rates was sometimes necessary. Therefore, the direct bathtub modelling avoids the problems seen in analyses of combining two separate

Weibull distributions. Lognormal models adjusted for patient's age were used in sensitivity analyses to predict the revision rates to 100 years of age. With lognormal models the predicted revision rates were reduced, as were the magnitude of the differences between prostheses. However, the relative performance of prostheses was similar to that with the bath tub models. Therefore, we tested the extrapolation generated from well-fitting models predicting a gradual increase in the rate of revision with time (i.e. bathtub model) and models predicting a gradual decrease in rate of revision with time (i.e. log-normal model). The trajectory of revision rates using our predicted model was compared with the Swedish registry data (reported follow-up data up to 20 years) supported an increasing hazard, therefore, demonstrates the validity of our chosen model.

To summarise, across four publications there were moderate differences in ICERs for the base case and the sensitivity analysis, however, all the analyses led to the same decision regarding the cost-effectiveness of these devices.

5.5. Value of information analysis

The uncertainty surrounding the cost-effectiveness of a LVAD was represented as a cost-effectiveness acceptability curve (CEAC). The CEAC illustrated that the LVAD had a very low probability of being cost-effective i.e. less than 15% at NICE WTP threshold of £30K/QALY; and 25% at £50K/QALY (using NICE end of life criteria). Therefore, the decision to reject or adopt the device based on the existing information could be uncertain; and the cost of uncertainty was explored using value of information analysis (VOI). The UK population expected value of perfect information (EVPI) was £670, 000 at £30K/QALY; and £13 million at £50K/QALY. However, the expected cost of conducting future research is more likely to exceed £700,000, due to the inherent cost of the required type of research; and the potential cost associated with the research (i.e. device cost, operative cost and treatment and follow-up visit). However, LVADs do tend to satisfy the NICE end of life criteria, where NICE tends to apply a threshold of £50K/QALY. At this threshold, the expected upper limit on the returns to further research is approximately £13 million (10 year life time). Therefore, for deciding to conduct more research, the expected value of sample information analysis would be worthwhile to fully inform the research design and the optimal sample size required to conduct further research. Likewise, EVPI analyses based on the evidence available from the REMATCH trial was explored by Girling et al (Girling et al,

2007), and the study concluded that a future trial was not expected to be cost-effective at the existing device cost (cost of the device was estimated to be £60,000). Indeed, the actual price of a device is a grey area and would continue to be so due to the competition within the market. Moreover, it would not be possible to examine the effectiveness of VAD alongside a clinical trial for ethical reasons, as randomising patients to a trial offering equal probability of HT for each group would not be workable.

The CEAC for the comparison between different types of THR; and THR vs RS at NICE WTP threshold of £20K/QALY was greater than 95%. As THR has shown to be a cost-effective technology, it is unlikely to benefit from additional research.

5.6. Study strengths

The approaches described in this thesis are comprehensive, and it provides robust information that payers need to make decisions on the cost-effectiveness of these devices. Furthermore, the studies project the long-term clinical outcomes in a cohort of patients with implanted hip prostheses and LVAD devices, as well as a detailed cost-effectiveness estimates.

A key strength of these studies is the use of real world evidence of hip prostheses and LVADs from the NJR and BTDB databases. BTDB provided key clinical characteristics for greater than 1000 individual patients receiving a relevant LVAD and MM over 9 year period. This study also used a virtual control to assess the choice of the comparator population. This approach was critical is testing the robustness of the cost-effectiveness estimate in absence of direct comparative evidence.

Similarly, the NJR database provided good quality data on time to revision after primary THR for 386,556 patients with osteoarthritis spanning 10 years. Using the NJR dataset, we further identified five frequently used types of prosthesis, classified according to both bearing surface and mode of fixation. This approach increased the granularity and clinical relevance of this study. This is the first lifetime cost-effectiveness analysis comparing five frequently used types of THR prosthesis, and comparing THRs with RS in young active patients. The main strength of this study lies in the flexible parametric and direct bathtub

modelling used to predict revision rate. The bath tub model predicts an initial decreasing hazard for revision after primary surgery followed by an increasing hazard as the prosthesis ages and the quality of the patient's bone deteriorates. This modelling approach avoided the problems seen in other publication where two separate Weibull distributions were combined to model revision.

Another major strength of this work is the application of propensity score matching to reduce confounding, therefore, assisting in obtaining the most credible cost-effectiveness estimates for devices.

5.7. Study limitations

There are limitations in the economic evaluation approaches, for example all approaches listed in this thesis require collection of observational data to allow estimates of cost-effectiveness. However, short and long-term clinical outcomes data are not routinely captured within an electronic health record data for patients with implanted devices. This may mean that the approaches illustrated in this thesis could only be used to assess the cost-effectiveness of medical devices, where there are existing clinical databases for post-marketing devices.

A major limitation of study I and II are, like all other studies on VAD devices, is around the nonrandomised comparison across devices. Therefore, the survival difference seen between the two devices could be due to external factors and not due to the difference in performance of the actual device. The length of follow-up of patients supported by either MM or a LVAD was short and no randomised or controlled evidence were available to inform the choice of an appropriate MM population to act as a comparator to BTT with a VAD. In addition, extrapolation beyond the observed data was required to model survival, and this unavoidably leads to bias regarding the estimation of transition probabilities in the longer term. The lack of short-term and long-term EQ-5D utility scores, and absence of reliable long-term EQ-5D data from published literature may be a possible limitation.

The BTDB contained limited clinical variables for the patients in this dataset. Although, the NJR contains good individual patient data for time to revision after primary THR, some key data were still missing (i.e. patient level activity data, body mass index; and adverse

events). We acknowledge that the lack of randomisation may lead to unexplored confounding and that observational studies could be susceptible to bias due to the limited ability to adjust for hidden covariates (i.e. level of physical fitness and athleticism may dictate patient selection for RS versus THR). For the cost-effectiveness analysis, we sourced model inputs from several different sources and were unable to link the three data sources i.e. PROMS, hospital episode statistics and NJR using patient records. This approach could have provided a better cost and QoL estimate by prosthesis type.

5.8. Study implications and recommendations

The findings from our studies revealed that LVADs considered as a BTT yield ICERs of £55,173, when compared with MM. Although LVADs were not considered cost-effective at NICE's standard WTP thresholds, they clearly demonstrated an improvement in QoL and functional status for patients who survive implantation of a LVAD as BTT. Our findings showed that patients on the HW device had superior survival than those on the HMII. However, it is clear that third generation LVADs are evolving and show a greater advantage over first and second generation LVADs. Therefore, LVADs clearly support survival benefits for BTT patients and a reduction in device costs would bring the ICER to the NICE WTP threshold. The cost of VAD would need to be reduced by 15% to bring the ICER to £50,000 per QALY and by 76% to bring the ICER to £30,000 per QALY. Pharmaceutical companies in agreement with Department of Health and NICE operate patient access schemes (PAS) to improve the cost-effectiveness of drugs. A PAS is proposed (i.e. a confidential discount rate) by the company with the aim of linking the price of the drug to decision makers WTP. No such transparent schemes are available to improve the cost-effectiveness of devices. Clearly, operating a flexible pricing scheme in any form will be a useful approach in considering LVAD as a BTT on the grounds of cost-effectiveness.

Our study was part of the supporting evidence for NICE hip replacement and resurfacing multiple technology assessment. As a result of this work, NICE revised its current benchmark set at a 10% revision rate at 10 years to 5%. Our study found that the likely factors to influence revision are gender, age, head size, and device manufacturer. Our study also found that RS is not cost-effective over a lifetime for healthy patients and THR is

undoubtedly the most cost-effective option for all women and for men aged over 50 years old. The implication of our findings was that the use of suboptimal RS devices were removed or slowly phased out from routine NHS use.

5.9. Future studies

Some of the challenges faced in the economic evaluation of medical devices are rooted in the diversity of devices, as they encompass a wide range of products across a variety of health care settings. This diversity with rapid technological advancement pose challenges to centralise regulatory assessment of medical devices, and has allowed a large number of devices to slip into routine use without rigorous monitoring. Hence, existing regulatory framework should be centralised and improved for better patient safety.

There are a number of technical challenges which would benefit from further work. Firstly, the accuracy of the long-term extrapolation of time-to-event data is constrained by the quality of the IPD used as the basis of extrapolation. In this thesis, we concentrated on commonly used parametric survival models, such as the exponential, lognormal and Weibull models, which make restrictive assumptions of the baseline hazard function, such as monotonicity. Although this thesis used Bathtub and flexible parametric models, work using other modelling approaches i.e. piece wise exponential framework using either a Bayesian or classical approach could offer greater flexibility to capture the observed data.

Secondly, new designs of THRs and RSs have been introduced in the market without any comparative evidence demonstrating their superiority to existing devices. Hence, there is a strong need for randomised studies demonstrating the superiority of the different types of THR and RS prosthesis. When undertaking these randomised studies, emphasis should be on collecting better quantity and quality data for different type of prosthesis i.e. rate of revision; QoL and cost data; and patient and surgeon's preference.

National joint registries have played a significant part in monitoring and improving the outcomes for hip replacement and is one of the largest databases for hip replacement in

the world (National Joint Registry, 2016). Likewise, Australian Orthopaedic Association National Joint Replacement Registry and Swedish Hip Arthroplasty Register are two other established large national joint registries. However, there has been no noticeable joint working programme (i.e. joint publications; exchange programmes to learn on the data quality strategies; and learn from experience to reduce overall level of missing data) across these established national registries. Such future joint working programmes could be beneficial if considered.

A major limitation of this study is that no RCT has yet been conducted comparing the long-term survival of BTT with MM for patients eligible for HT. Therefore, we strongly recommend undertaking a prospective randomised trial comparing BTT with MM for patients eligible for HT. However, an RCT offering equal probability of HT for each group would not be feasible for ethical reasons. Hence, in the absence of such data, it is crucial to have a strong commitment to collect high quality real world evidence, including data on long-term patient survival, QoL, functional ability, adverse events and costs associated with second and third generation LVADs. It is necessary to give due consideration to ensure that the collection of real world evidence is fully updated with low overall level of missing data.

5.10. Conclusions

Our studies report the challenges faced in the economic evaluation of medical devices, and test the credibility of the cost-effectiveness estimates in a HTA framework. We have demonstrated methods to address these challenges: applications to the cost-effectiveness of BTT with LVAD compared to MM; and cost-effectiveness of five frequently used categories of THR; and cost-effectiveness of RS versus THR. However, practical implementation of these approaches to other devices could be limited due to the lack of good quality and quantity data on comparative treatment effectiveness, making economic evaluations challenging. The VAD and hip devices had long-term observational data for economic evaluation; whether or not establishing similar observational database becomes a regular component of device evaluation process depends on the affordability and practicality of establishing and maintaining such a database. Recently, the Department of Health has established “The Breast and Cosmetic Implant Registry following the Poly Implant Prosthese breast implant scandal in 2010 (Register, 2016). The DePuy RS device

did face similar challenges, and was recently withdrawn after unacceptably high revision rates. However, such an incident was well monitored and managed, as the registry provided an opportunity for the hospital to notify patients on the performance of their implant and to create awareness through patient organisations (Smith AJ, 2012). Whether or not mandatory database for all surgical implants will improve decision making and ultimately improve patient outcomes is not a question that could be answered until more such economic evaluations have been undertaken.

6. References

Aaronson KD, Slaughter MS, Miller LW et al (2012). Use of an intra-pericardial, continuous flow, centrifugal pump in patients awaiting heart transplantation. *Circulation*, 191-200.

Aaronson KD, Schwartz S, Chen T, Wong K et al (1997). Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. Circulation, 95: 2660-2667.

Association of British Healthcare Industries (2013). Retrieved September 2016, from http://www.abhi.org.uk/key-issues/uk-market/valueofmedtech.aspx

Annemans L, Aristides M, Kubin M. (2007). *Real-Life Data: A Growing Need*. Retrieved from ISPOR: www.ispor.org/news/articles/Oct07/RLD.asp

Ardaugh MB, Graves ES, Redberg R. (2013). The 510(k) Ancestry of a Metal-on-Metal Hip Implant. *N Engl J Med* , 97-100.

Chapman AM, Taylor CA, Girling AJ. (2014). Are the UK systems of innovation and evaluation of medical devices compatible? The role of NICE's Medical Technologies Evaluation Programme (MTEP). *Appl Health Econ Health Policy, 12*(4), 347-57.

Clarke A, Pulikottil-Jacob, Freeman K, Mistry H et al. (2015). Total hip replacement and surface replacement for the treatment of pain and disability resulting from end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation. *Health Technol Assess, 19*(10), 1-668.

Clarke A, Pulikottil-Jacob, Connock M, Suri G et al. (2014). Cost-effectiveness of left ventricular assist devices (LVADs) for patients with advanced heart failure: Analysis of the British NHS bridge to transplant (BTT) program. *International Journal of Cardiology*, 338-345.

The Medicines for Human Use (Clincial Trials) Regulations (2004). legislation.gov.uk.

http://www.legislation.gov.uk/uksi/2004/1031/contents/made

Costa ML, Achten J, Parsons N, Edlin R et al. (2012). Total hip arthroplasty versus resurfacing arthroplasty in the treatment of patients with arthritis of the hip joint: single centre, parallel group, assessor blinded, randomised controlled trial. *BMJ*.

Cowger J, Sundareswaran K, Rogers JG, Park J S et al. (2013). Predicting survival in patients receiving continuous flow left ventricular assist device: The HeartMate II risk score. Journal of the American College of Cardiology, 16(3) 313-21.

D'Agostino R. (1998). Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomised control group. Statistics in medicine. 17, 2265-2281.

Department of Health. (2011). *Procurement matters: A best practice procurement diagnostic* http://www.ukvisionstrategy.org.uk/sites/default/files/dh_125076%20(1).pdf.

Department of Health. (2013–2014). *NHS Reference Costs 2013–2014 London: Department of Health, 2010 https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014. (date last accessed 16 August 2015).* Retrieved from www.gov.uk: https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014

Drummond M, Tarricone R, Torbica A. (2016). Challenges in the assessment of medical devices: Recommendations from the Medtechta project. Value in Health, 19 (7), A860.

European Economic Community Council directive. (1990). *Emergogroup*. Retrieved September 2016, from http://www.emergogroup.com/sites/default/files/file/europe-consolidated-90-385-eec.pdf

The Economist Intelligence Unit. (2014). *A white paper from The Economist Intelligence Unit Healthcare. Health outcomes and cost: A 166-country comparison*. Retrieved from http://docplayer.net/1573183-A-white-paper-from-the-economist-intelligence-unit-healthcare-health-outcomes-and-cost-a-166-country-comparison-www-eiu.html

EMERGO. (2006). *UNITED KINGDOM - UK Medical Device Industry and Demographic Data*. Retrieved September 2016, from http://www.emergogroup.com/resources/market-united-kingdom

Food and Drug Administration. (2013). *Design Considerations for Pivotal Clinical Investigations for Medical Devices*. Retrieved from http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm373750.htm

French-Mowat E, Burnett J. (2012). How are medical devices regulated in the European Union? *J R Soc Med*, S22-S28.

Girling AJ, Freeman G, Gordon JP et al. (2007). Modeling payback from research into the efficacy of left-ventricular assist devices as destination therapy. *Int J Technol Assess Health Care*, 269-77.

Gohler A, Geisler BP, Manne JM et al. (2009). Utility estimates for decision-analytic modeling in chronic heart failure—health states based on New York Heart Association classes and number of rehospitalizations. *Value in Health, 12*(1), 185-7.

Altenstetter C, Permanand G. (2006). EU Regulation of Medical Devices and Pharmaceuticals in Comparative Perspective. *Review of Policy Research*, 24(5), 385-405.

Heintzbergen S, Kulin NA, Ljzerman MJ et al. (2013). Cost-utility of metal-on-metal hip resurfacing compared to conventional total hip replacement in young active patients with osteoarthritis. *Value Health*, 942-52.

Higgins JPT, Green S. (2011). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 . *The Cochrane Collaboration*.

John R, Naka Y, Smedira NG. (2011). Continuous flow left ventricular assist device outcomes in commercial use compared with the prior clinical trial. *Ann Thorac Surg*, 1406-13.

Kandala NB, Connock M, Pulikottil-Jacob R et al. (2014). Setting benchmark revision rates for total hip replacement: analysis of registry evidence. *BMJ*.

Kastrati A, Mehilli J, Pache J et al. (2007). Analysis of 14 trials comparing sirolimus -eluting stents with bare-metal stents. *N Engl J Med*, 1030-9.

Lambert PC, Royston P. (2009). Further development of flexible parametric models for survival analysis. *The Stata journal* , 265-290.

Lao, C. S. (1995). Statistical considerations for survival analysis from medical device clinical studies. *Journal of Biopharmaceutical Statistics, 5*(2), 159-170.

Latimer, N. (2011). NICE DSU Technical support document 14: Survival analysis for

economic evaluations alongside clinical trials - extrapolation with patient-level data.

Levy WC, Mozaffarian D, Linker DT et al. (2006). The Seattle Heart FailureModel: prediction of survival in heart failure. *Circulation* , 1424-43.

Longworth L, Rowen D. NICE DSU Technical Support Document 10: the use of mapping methods to estimate health state utility values 2011.

MHRA. (2008). www. *gov.uk*. Retrieved September 2016, from Medicines and Healthcare Products Regulatory Agency: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/403818/Directives_Bulletin_no3.pdf

MHRA. (2012). *Medical Device Alert: All Metal-on-Metal(MoM) Hip Replacements (MDA/2012/036).* MHRA. Retrieved from Medicines and Healthcare Products Regulatory Agency.

MHRA. (2015). *www.gov.uk*. Retrieved September 2016, from Medicines and Healthcare products Regulatory Agency: https://www.gov.uk/topic/medicines-medical-devices-blood/medical-devices-regulation-safety

Moreno SG, Novielli N, Cooper NJ. (2012). Cost-effectiveness of the implantable HeartMate II left ventricular assist device for patients awaiting heart transplantation. *J Heart Lung Transplant*, 450-8.

National Audit Office, NHS Executive. (2000). *Hip replacements: Getting it right first time.* London: The Stationery Office.

National Joint Registry. (2014). *National Joint Registry*. Retrieved from WWW. njrcentre.org.uk: http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Pricebenchmarking/NJR%20EMBED%20Information%20leaflet%20(enhanced%20price-benchmarking%20service).pdf

National Joint Registry. (2014). *National Joint Registry for England and Wales: 11th Annual Report.*

National Joint Registry. (2016). *National Joint Registry*. Retrieved from www.njrcentre.org.uk: http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Reports/13th%20Annual%20Report/07950%20NJR%20Annual%20Report%202016%20ONLINE%20REPORT.pdf

NHS England. (2016). *NHS England*. Retrieved from WWW. england.nhs.uk: https://www.england.nhs.uk/2016/04/nhs-devices-deal/

The Newcastle Upon Tyne Hospitals NHS Foundation Trust. (2014). *www. newcastle-hospitals.org.uk*. Retrieved from http://www.newcastle-hospitals.org.uk/downloads/policies/Operational/MedicalDevicesProcurementPolicy201602.pdf

National Institute for Health and Care Excellence. (2000). NICE: *Guidance on the Selection of Prostheses for Primary Total Hip Replacement*. Retrieved from National institute for health and care excellence.

National Institute for Health and Care Excellence. (2000). NICE: Guidance on the Selection of Prostheses for Primary Total Hip Replacement. *Technology appraisal guidance 2*.

National Institute for Health and Care Excellence. (2002). NICE: *Guidance on the Use of Metal on Metal Hip Resurfacing Arthroplasty*. Retrieved from NICE: National Institute for Health and Care Excellence: http://guidance.nice.org.uk/TA44

National Institute for Health and Care Excellence. (2011). *NICE: Medical technologies evaluation programme—methods guide*. Retrieved from www.NICE.org.uk: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-medical-technologies/Medical-technologies-evaluation-programme-methods-guide.pdf

National Institute for Health and Care Excellence. (2014). NICE: *Developing NICE guidelines: the manual*.

Pagani FD, Miller LW, Russell SD, et al. (2009). Extended Mechanical Circulatory Support With a Continuous-Flow Rotary Left Ventricular Assist Device. *J AM Coll Cardiol*, 312-21.

Parvizi N, Woods K. (2014). Regulations of medicines and medical devices: contrasts and similarities. *Clinical medicine*, 6-12.

Pennington M, Grieve R, Sekhon JS, et al. (2013). Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis. *BMJ*, 1-14.

Pollard TC, Baker RP, Eastaugh-Waring SJ. Bannister GC. (2006). Treatment of the young active patient with osteoarthritis of the hip. A five- to seven-year comparison of hybrid total hip arthroplasty and metal-on-metal resurfacing. *J Bone Joint Surg Br*, 592-600.

Personal Social Service Research Unit. (2010). *Unit costs of health & Social care*. Retrieved from PSSRU: http://www.pssru.ac.uk/project-pages/unit-costs/

Pulikottil-Jacob R, Connock M, Kandala N-B, Mistry H, et al. (2015). Cost effectiveness of total hip arthroplasty in osteoarthritis. *Boint Joint J*, 449-57.

Pulikottil-Jacob R, Connock M, Kandala N-B, Mistry H, et al. (2016). Has Metal-On-Metal Resurfacing Been a Cost-Effective Intervention for Health Care Providers?ÐA Registry Based Study. *PLoS ONE*, 1-13.

Pulikottil-Jacob, Suri G, Connock M, Kandala N-B, et al. (2014). Comparative cost-effectiveness of the HeartWare versus HeartMateII left ventricular assist devices used in the United Kingdom National Health Service bridge-to-transplant program for patients with heart failure. *The Journal of Heart and Lung Transplantation*, 350-357.

The Register. (2016, Octoebr). *www.theregister.co.uk*. Retrieved from The Register: http://www.theregister.co.uk/2016/10/17/national_silicon_boob_database/

Rodriguez L, Suarez EE, Loebe M, Bruckner BA. (2013). Ventricular Assist Devices (VAD) Therapy: New Technology, New Hope? *Methodist Debakey Cardiovasc J*, 32-37.

Rose EA, Gelijins AC, Moskowitz AJ, et al. (2001). Long-Term Use of a Left Ventricular Assist Device for End-Stage Heart Failure. *The New England Journal of Medicine*, 1435-1443.

Rosenbaum PR, Rubin DB. (1983). The central role of propensity score in observational studies for causal effects. Biometrika, 70:41-55.

Schaffer JM, Allen JG, Weiss ES, et al. (2009). Evaluation of Risk Indices in Continuous-Flow Left Ventricular Assist Device Patients. *Ann Thorac Surg* , 1889-96.

Sedrakyan A, Normand S-L, Dabic S et al. (2011). Comparative assessment of implantable hip devices with different bearing surfaces: Systematic appraisal of evidence. *BMJ*, 1-12.

Shah KK, Tsuchiya A, Wailoo AJ. (2014). Valuing health at the end of life: an empirical study of public preferences. *The European Journal of Health Economics*, 389-399.

Sharples LD, Dyer M, Cafferty F, et al. (2006). Cost-effectiveness of ventricular assist device use in the United Kingdom: results from the evaluation of ventricular assist device programme in the UK (EVAD UK). *J Heart Lung Transplant*, 1336-43.

Slaughter MS, Rogers J. (2012). Editorial commentary: determining the cost-effectiveness of mechanical circulatory support. *J Heart Lung Transplant*, 448-9.

Smith AJ, Dieppe P, Vernon K, Porter M et al. (2012). On behalf of the National Joint Registry for England and Wales. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet*.

Soares I, Carneiro AV. (2002). Drug Class Effects: Definitions and Practical Applications. *Rev Port Cardiol* , 1031-1042.

Stevenson LW, Miller LW, Desvigne-Nickens P, et al. (2004). Left ventricular assist device as destination for patients undergoing intravenous inotropic therapy: a subset analysis from REMATCH (Randomized Evaluation of Mechanical Assistance in Treatment of Chronic Heart Failure). *Circulation* , 975-81.

Strueber M, O'Driscoll G, JansZ P, Khaghani A, et al. (2011). Multicenter Evaluation of an Intrapericardial Left Ventricular Assist System . *J Am Coll Cardiol*, 1375-82.

Sutcliffe P, Connock M, Pulikottil-Jacob R, Kandala NB et al. (2013). Clinical effectiveness and cost-effectiveness of second- and third-generation left ventricular assist devices as either bridge to transplant or alternative to transplant for adults eligible for heart transplantation: systematic review and cost-effectivene. *Health Technol Assess., 17*(53), 1-499.

Tarricone R, Boscolo PR, Armeni P. (2016). What type of clinical evidence is needed to assess medical devices? *Eur Respir Rev*, 259-265.

Weinstein MC, Torrance G, McGuire A. (2009). QALYs: The Basics. *Value in Health, 12*(1), S5-S9.