Outcome assessment and economic evaluation of short-stay intensive care for coronary artery bypass patients

Outcome assessment and

economic evaluation of

short-stay intensive care for

coronary artery bypass patients

Outcome assessment and economic evaluation of short-stay intensive care for coronary artery bypass patients

Proefschrift

Ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen,
op vrijdag 25 september 2009 om 12.00 uur

door

Ghislaine Anna Pauline Gemma van Mastrigt

UM
UNIVERSITAIRE
PERS MAASTRICHT

Promotores

Prof. dr. J.G. Maessen

Prof. dr. J.L. Severens

Prof. dr. M.H. Prins

Beoordelingscommissie

Prof. dr. J. H. Zwaveling (Voorzitter)

Prof. dr. E. Buskens (Universitair Medisch Centrum Groningen)

Prof. dr. J. W. M. Greve

Dr. J. P. A. M. Schönberger (Catherina ziekenhuis)

Voor Stephan, Jip en Julie

Contents

Abbreviations

95% CI	95% Confidence Intervals
AMI	Acute Myocardial Infarction
BDI	Beck Depression Inventory
BMI	Body Mass Index
CABG	Coronary Artery Bypass Grafting
CAD	Coronary Artery Disease
CBA	Cost Benefit Analysis
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CE plane	Cost-Effectiveness plane
CHD	Coronary Heart Disease
CHEC-list	Consensus Health Economic Criteria-list
CMA	Cost Minimization Analysis
CT	Cardio Thoracic
CTC	Cardio Thoracic Care
CUA	Cost Utility Analysis
CVA	Cerebro Vasculair Accident
CVD	Cardio Vascular Disease
DBC	Diagnosis Treatment Combination
DF	Degrees of Freedom
DH	District Hospital
EE	Economic Evaluation
EQ-5D	EuroQoL
HUI	Health Utilities Index
Hrs	Hours
IC	Intensive Care
ICC	Intraclass Correlation Coefficient
ICER	Incremental Cost-Effectiveness Ratio

IHD	Ischemic Heart Disease
ITT	Intention-To-Treat
K-S test	Kolmogornov-Smirnov test
LC	Low Care
MacNew	MacNew Heart Disease Questionnaire
MC	Medium Care
MI	Myocardial Infarction
Mins	Minutes
MILQ	Multidimensional Index of Life Quality
MMPI 2	Multiphasic Personality inventory 2
NYHA	New York Heart Association functional classification
PCI	Percutaneous Coronary Intervention
QALM	Quality Adjusted Life Month
QALY	Quality Adjusted Life Year
QLMI	Quality of Life after Myocardial Infarction
QoL	Health-Related Quality of Life
RR	Relative Risk
SAQ	Seattle Angina Questionnaire
SD	Standard Deviation
SES	Standardized Effect Size
SF-36	Short Form-36
SF-6D	Short Form-6D
SRM	Standardized Response Mean
SSIC	Short-stay Intensive Care
STAI-S	State-Trait Anxiety Inventory
UH	University Hospital
VAS	Visual Analogue Scale
Δ QALM	Delta Quality Adjusted Life Month
Δ QALY	Delta Quality Adjusted Life Year

CHAPTER

1

General Introduction

Background

This first chapter of the thesis outlines the background of both economic evaluation (EE) and coronary heart disease (CHD). In the first (1) part the general aim of performing EEs and elementary concepts of EEs are described. In the second (2) part CHD is introduced and its treatment options are described. In parts 3 and 4 the epidemiology and the economic burden of CHD are described. Finally, the aim and the thesis outline are summarized in part 5.

1 Economic evaluations in healthcare

Over the past decades a rapid increase of medical care costs was observed in western countries due to technological developments and ageing populations. At the same time there was increasing pressure of national governments on budgets and an awareness that limits must be set to the growth of healthcare costs. EEs were introduced as a tool to inform decision-makers (e.g., Minister of Public Health, insurance companies, management, professionals and individual healthcare providers) on the efficiency or cost-effectiveness of a specific healthcare program. EEs aims to answer the following two questions; 'Is the new treatment worth doing compared to what we could do with the same resources?', and 'Are we satisfied with the way healthcare resources (required to make the treatment available to those who benefit from it) are spent'? (Drummond, 1996).

An EE can be defined as the comparative analysis of alternative courses of action in terms of both costs and consequences. The basic framework of any EE is to identify, measure, value and compare costs and consequences of alternative treatments considered (Drummond, 1996). The decision making process of the use of specific healthcare technology is not only based on its cost-effectiveness, but also on its efficacy (can the new treatment work?), effectiveness (does it work?), and availability (is it

reaching those who need it?). The latter three aspects are not the primary subject of EEs (Drummond, 1996) and need to be assessed in additional evaluations.

Framing is a theoretical framework, which can be used to develop an EE. Framing can be defined as a series of decisions that collectively define and describe the EE study to be undertaken (Gold, 1996). In other words, moving from a general idea to concrete details it is necessary to design an EE study leading to the calculation of a cost-effectiveness ratio that is informative to the decision making context. The important aspects of framing will be briefly discussed below.

The first important aspect of framing an EE is defining the study objective and the target audience. In other words, which particular healthcare program will be the objective of the study, and who are the persons that make the decisions about the particular healthcare programs (an example of the objective can be: what is the cost-effectiveness of a specific type of fast-track treatment compared to care as usual; an example of the audience can be: general managers of the intensive care units).

The second aspect is choosing the correct type of analysis for the EE. Three types of full EEs and five types of partial EEs can be distinguished. In a full EE both the costs and consequences of alternative treatments or services are evaluated. The first type of full EE is the cost-effectiveness analysis (CEA). All relevant costs and health gains (for instance reduction of blood pressure, decrease in the number of myocardial infarctions) are measured. The second type of full EE is the cost-utility analysis (CUA), a special form of CEA analysis. The effects are expressed in the preference for the health state reached (like Quality Adjusted Life Years (QALYs). QALYs are a combined measure of quality of life and duration of life. The third type full EE is the cost benefit analysis (CBA). Both costs and effects are expressed in monetary terms.

15

In addition there are five types of partial EEs (Drummond, 1996). In three types of partial EEs no comparison of alternatives is being made (that is, a single service or program is being evaluated). They are named: outcome description (only the consequences of program or service are examined), cost description (only the costs are investigated), and the combination of the two previous ones, a cost-outcome description (Drummond, 1996). The two other types of partial EEs are efficacy or effectiveness evaluation and cost analysis. In both evaluation situations two or more alternatives are compared, but the costs and outcomes of each alternative treatment are not examined simultaneously. In efficacy or effectiveness evaluation only the outcomes and in cost analysis only the costs are examined.

A third aspect of framing an EE is choosing the relevant type of perspective. There are several perspectives, which can be chosen. First a societal perspective, which is recommended by the Dutch and the international pharmaco-economic guideline (ISPOR, 2008). In this type of perspective all relevant costs are taken into account: direct medical costs, direct non-medical costs, and indirect costs. Direct medical costs are all costs that are part of the healthcare sector, and are related to detection, treatment and rehabilitation of health impairment or its effect (for instance, costs of blood samples, drugs, general practitioner). Direct non-medical care costs are associated with provision of medical services outside the healthcare sector (for instance, out-of-pocket expenses, costs of informal care). Indirect costs are productivity costs: costs related to absence of work due to illness and treatment. Another perspective that can be used is the decision makers' perspective. For instance, the hospital perspective (costs only related to the hospital are accounted for and the healthcare insurer's perspectives (costs of care which the insurer pays). Furthermore, the patient perspective can be chosen. All relevant costs for the patient are calculated (e.g. travel and waiting times, out-of-pocket expenses of drugs).

The fourth aspect is defining the intervention, the comparison treatment and the target population. All three should be described in detail. The target population is the population for whom the treatment is intended and is mostly defined by eligibility criteria.

The fifth aspect is defining the boundaries and the time horizon. The boundaries are those aspects that are not primary related to the scope of the study. These aspects should not be incorporated in the analysis, like the influences on quality of life of family members and costs made by those relatives due to illness of the patients under study. The time horizon for the EE is the period of time over which costs and benefits are likely to differ between the alternative options being compared (Sculpher, 2006). This can be, for instance, the duration of patients' lives. If the appropriate time horizon extends beyond the availability of data, a mathematical modelling of costs and consequences might be unavoidable (Buxton, 1997).

2 Coronary heart disease

The term cardiovascular disease (CVD) covers a number of different diseases, which affect the heart. Coronary heart disease (CHD) (or coronary artery disease (CAD) or ischemic heart diseases (IHD) (ICD-10 I20-I25) are the type of CVD caused by atherosclerosis causing a shortage of oxygen in the heart muscle due to a reduction in blood supply. CHD can be divided into an acute (heart attack or myocardial infarction) and a chronic form (angina pectoris). Acute myocardial Infarction (AMI or MI), more commonly known as a heart attack, is a medical condition that occurs when the blood supply to a part of the heart is interrupted. The resulting ischemia or oxygen shortage causes damage and potential death of heart tissue. Angina pectoris, commonly known as angina, is chest pain due to ischemia (a lack of blood and hence oxygen supply) of the heart muscle, generally due to obstruction or spasm of the coronary arteries.

17

For both MI and angina pectoris the most common invasive treatment intervention to restore blood supply in the heart are the percutaneous coronary interventions (PCI) and coronary artery bypass grafting surgery (CABG) (Scholte op Reimer, 2006). Patients who undergo CABG-surgery can be stratified by risk of the development of postoperative complications. By means of a risk stratification system based on Parsonnet (Parsonnet, 1989) and Society of Thoracic Surgeons criteria (STS, 2003), low-risk CABG patients can be identified as those patients who fulfill the following criteria: stable cardiological hemodynamic condition (not having a cardiogenic shock, no need for inotropic therapy, no need for intraaortic balloon pump, and no ongoing or recent myocardial infarction), not older than 78 years, an ejection fraction of more than 30% and not requiring hemodialysis or having pulmonary hypertension.

High-risk CABG patients are all CABG patients who do not fulfill the above-mentioned criteria or CABG patients who also undergo valve replacements or valve repair surgery. In 2001 approximately 80% of the patients with CHD undergoing CABG-surgery in the Netherlands could be classified as low-risk (based on an estimate of the Summit database of the department of Cardio Thoracic Surgery of the University Hospital of Maastricht). The average length of hospital stay for Dutch CABG patients between 2005 and 2007 was 11.6 days (Prismant, 2008). After surgery patients are admitted to the intensive care unit for monitoring the essential organ functions (like cardiac output, fluid balance).

3 Epidemiology of coronary heart disease

Quantification of the burden of illness of CHD is important for healthcare planning, policy decisions and resource allocation for research. CHD is the most common heart disease in the Netherlands. In 2003 more than 675,000 people were diagnosed CHD. The prevalence of CHD in the United States of America (US) was 13.2 million in 2003. Data from the

Health Survey in the United Kingdom (UK) in 2006 show a prevalence of CHD of 6.5% in men and 4.0% in women (BHF, 2008).

The annual incidence of CHD in the Netherlands in 2003 was 51.0 per 1,000 men and 32.5 per 1,000 women (RIVM, 2008b). The estimated incidence in the US of MI is 600,000 new attacks and 320,000 recurrent attacks annually (AHA, 2008). The incidence of MI varies in the United Kingdom, but on average the incidence rate for men aged between 30 and 69 is about 600 per 100,000, and for women about 200 per 100,000 (BHF, 2008).

In the Netherlands, in the last three years (2005-2007), on a yearly basis almost 9,000 patients with CHD underwent CABG-surgery and more than 35,000 CHD patients underwent a PCI (Prismant, 2008). In the US in 2005 469,000 CABG-surgeries were performed (AHA, 2008).

CVD was for both men and women the most important cause of death in the Netherlands in 2004. More than 45,000 people died due to CVD, which is 33% of all causes of death. One third of these deaths was due to CHDs (NHS, 2006). CHD caused 451,326 deaths in the United States of America in 2004 (AHA, 2008). CHD causes over 94,000 deaths a year in the UK: this is approximately one in five deaths in men and one in seven deaths in women (BHF, 2008). These figures indicate that CHD is not only a health problem that affects both men and women, but it is also the number one cause of death in the Netherlands, the US and the UK.

4 Costs of illness

The costs of illness illustrate the economic burden of a disease in a specific country. The total healthcare costs due to CHD were estimated in 2005 at €1.251 million in the Netherlands (RIVM, 2008a). In the US the total direct and indirect costs of CVD are estimated at $448.1 billion (AHA, 2008). Overall CHD is estimated to cost the UK economy over £9.0 billion a year (in 2006). Of the total cost of CHD to the UK, around 36% is due to direct healthcare costs, 43% to productivity losses, and 21% to the infor-

19

mal care of people with CHD (BHF, 2008). CHD is estimated to cost the European Union over €49 billion a year. Around 48% of the total costs of CHD in the European Union are due to direct healthcare costs, 34% to productivity losses and 18% to the informal care of people with CHD (BHF, 2008). These figures show that the total healthcare costs of CHD are substantial for both the European Union and the US.

5 Aim and outline of the thesis

The aim of this thesis is twofold. The first part of the thesis focuses on the evaluation of short-stay intensive care treatment (SSIC). The second part deals with methodological aspects of measuring the outcomes as part of an EE of SSIC, more specifically, the measurement and analysis in quality of life of CABG patients.

Part 1: Evaluation of short-stay intensive care treatment

The first objective of this thesis is to evaluate the clinical effectiveness, quality of life and cost-effectiveness of the short-stay intensive care treatment for low-risk CABG patients.

The second chapter comprises a systematic review of all randomized clinical trials published from 1966 to 2004 in low-risk coronary artery by-pass patients who evaluated fast-track treatments. Furthermore, a meta-regression analysis will be performed to evaluate the impact of these types of intensive care treatments on the number of hours of intensive care and hospital stay. In addition, the methodological quality of trials, including the economic evaluation performed alongside, will be assessed.

Chapter 3 contains the clinical effectiveness and cost-effectiveness of the randomised controlled clinical equivalence trial evaluating the short-stay intensive care treatment protocol in low-risk CABG patients. In this study overnight intensive care stay (care as usual) will be compared with discharge from intensive care within 8 hours (SSIC).

In the fourth chapter the disease specific, generic and domain specific quality of life of the CABG patients receiving SSIC treatment will be evaluated until one year after surgery. The first four hundred patients who also participate in the trial (chapter 3) participate in this quality of life study.

Part 2: Methodological aspects of measurement and analysis of quality of life

The second objective of this thesis is the improvement of measurement and analysis of quality of life when evaluating interventions in CABG patients.

In chapter 5 the psychometric proportions of a generic (EQ-5D) and a disease specific (MILQ) quality of life measure will be assessed. Therefore, the item completion rate, the discriminative validity, the predictive validity, the criterion validity and the test-retest reliability of both questionnaires will be examined.

In the sixth chapter, the impact of imbalances in baseline utility in QALY calculations will be examined. Five different methods for the calculation of QALYs will be discussed. The data of the economic evaluation of the SSIC treatment will be used to demonstrate the impact of the various correction methods on the study outcomes.

Chapter 7 gives an overview of the thesis. The main findings will be summarized, discussed and recommendations for further research will be given.

Part

1

Evaluation of the short-stay intensive care treatment

CHAPTER

2

Does fast-track treatment lead to a decrease of intensive care unit and hospital length of stay in coronary artery bypass patients?

A meta-regression of randomised clinical trials

Ghislaine van Mastrigt
Jos Maessen
John Heijmans
Hans Severens
Martin Prins

Abstract

Objective: Evaluation of randomised controlled clinical trials studying fast-track treatment in low-risk coronary artery bypass grafting patients.

Design: Meta-regression.

Patients: Low-risk coronary artery bypass grafting patients.

Interventions: Fast-track treatments including (high or low) anesthetic dose, normothermia – vs. hypothermia and extubation protocol (within or after 8 hours).

Measurements: Number of hours intensive care unit stay, number of days of hospital stay, prevalence of myocardial infarction, and death. Furthermore, quality of life and cost evaluations were evaluated. The epidemiological and economic qualities of the different trials were also assessed.

Main results: A total of 27 studies evaluating fast-track treatment were identified, of which 12 studies were with major and 15 were without major differences in extubation protocol or anesthetic treatment or both. The use of an early extubation protocol (P-value =.000) but not the use of low anesthetic dose (P-value =.394) or normothermic temperature management (P-value =.552) resulted in a decrease of the total intensive care unit stay of low-risk coronary artery bypass grafting patients. Early extubation was found to be an important determinant of the total hospital stay for these patients. An influence of the type of fast-track treatment on mortality or the prevalence of postoperative myocardial infarction was not observed. In general, the epidemiological and economic qualities of included studies were moderate.

Conclusions: Although fast-track anesthetics and normothermic temperature management facilitate early extubation, the introduction of an early extubation protocol seems essential to decrease intensive care unit and hospital stay in low-risk coronary artery bypass grafting patients.

Introduction

Rising medical costs mandate a close look at our practices to assess the possibility of costs savings without comprising the quality of care (vander Salm & Blair, 1984). As coronary artery bypass grafting (CABG) is a costly and common procedure, this treatment merits attention. The intensive care unit (ICU) stay is an important cost component of the total cost of recovery after CABG (Cheng, 1996b). Hence, earlier discharge to lower-care facility would decrease the cost of hospitalization of CABG patients. CABG patients with a low Parsonnet score (Parsonnet, 1989) who stay in ICU facilities have low risk for postoperative complications and are there-fore candidates for a decrease of ICU length of stay and, hence, are eligible for fast-track treatment. This includes not only fast-track anesthet-ics, but also normothermic temperature management and the use of weaning protocols that make early extubation possible. Several reviews performed so far have demonstrated the safety for elements of fast-track treatment, for instance early vs. late extubation (Hawkes, 2003; Meade, 2001) or low vs. high use of opioids (Myles, 2003; Myles, 2002).

Due to the heterogeneity of the interventions, we sought to investigate the combined effect of the above-mentioned elements after CABG sur-gery on the decrease of ICU and hospital stay, and for this purpose, we used a meta-regression technique (Thompson & Higgins, 2002). Further-more, we systematically reviewed the studies on reported quality of life and costs.

Materials and Methods

Selection criteria

All randomised clinical trials of low-risk, adult (age ≥18 years) CABG patients that evaluated fast-track treatments and reporting ICU stay were eligible for inclusion. Studies were excluded if the patient sample included a majority of non-CABG patients or if a majority of patients were considered to have a high risk for complications. High-risk was defined as: age > 78 years, ejection fraction of <30%, stage 3 obesity, hemodialysis, pulmonary hypertension, recent cardiovascular accident (<1 month), recent myocardial infarction (MI; <24 hrs before CABG), cardiogenic shock, need for inotropic therapy, ongoing infarction or the need for intraaortic balloon pump (Priestley, 2002). Abstracts and reviews were also excluded.

Search strategy

A systematic search for all relevant trials published from 1966 until through October 2004 was performed. Studies were identified through MEDLINE and Cochrane databases using, for instance the following terms: coronary artery bypass surgery, randomised controlled clinical trial, fast-track, early extubation and ICU Stay. The complete search strategy is provided in Appendix 1. References lists were screened for additional relevant publications.

Classification of interventions

The interventions of the different studies were independently assessed by two reviewers (GvM, JH) and disagreement was resolved by discussion. The types of interventions were categorized first, by type and dose of anesthetics (high dose: fentanyl >15μg/kg, sufentanil >3μg/kg, alfentanil >60μg/kg (Ahonen, 2000), low dose: fentanyl <15μg/kg, sufentanil <2-3μg/kg, alfentanil<60μg/kg (Ahonen, 2000) or remifentanyl), second by

28

temperature management during cardiopulmonary bypass (normothermia (>35°C) or hypothermia (<35°C)) and third, by extubation protocol (the intention to extubate a patient within or after 8 hours postoperatively).

Outcome measures

The primary outcome measure was length of ICU stay (in hours). The other outcomes were hospital stay (days), MI, death, quality of life (QoL) and costs.

For this study, the MI definition of the trials needed to contain the following aspects: a typical rise and gradual fall (troponin) or more rapid rise and fall (CK-MB) with at least one of the following criteria: ischemic symptoms, development of pathologic Q waves on the echocardiogram or echocardiographic changes of ischemia ST depression/elevation or t-wave inversion (Alpert, 2000).

Methodological quality of the studies

The articles were not blinded for authors, institution or journal. The three assessed methods of methodological qualities were: allocation concealment (unclear and adequate), blinding of person or persons deciding on the length of ICU stay (adequate, unclear) and the use of intention-to-treat analysis for the primary outcome ICU stay. Uncertainties were assessed in consensus meetings, eligibility of the studies was discussed by two of the authors (GvM, JH), and concerns of the methodological quality were discussed by two of the authors (GvM, MP).

The quality of the economic evaluations was assessed using the Consensus Health Economic Criteria list (CHEC-list), including 19 items (Donaldson C, 2002). Two items, concerning 'future costs' and 'ethical and distributional issues' do not apply to these studies and were omitted. Items scored as 'yes' received 1 point. Items scored as 'unclear' or 'no' received no points. A total score was calculated by summing up the

score of the 17 items (range, 0-17). Uncertainties about quality of the economic evaluations were assessed in a consensus meeting (GvM, HS).

Statistics

When median and range were reported for continuous outcomes, the mean and standard deviation were estimated by assuming that the mean was equivalent to the median and the standard deviation was one quarter of the range (O'Rourke, 2002).

To evaluate the relation between length of stay at the ICU and the type of treatment, a regression analysis was performed. ICU stay was the dependent and anesthetic treatment (low, unknown or high dose; coding: -1,0,1), temperature management (normothermia, unknown or hypothermia, coding: -1,0,1) and extubation protocol (<8 hrs, unknown or >8 hrs, coding: -1,0,1) were the independent variables. The number of patients in the study divided by squared standard deviation is used as a regression weight in the model. Separated models were analyzed for studies categorized with or without major difference in anesthetic dose and/or extubation protocol between both studies arms. To correct for year of publication and the methodological quality variables (score of 1, if the score was adequate or yes and 0 if the score was unclear or no) were as co-variables included in the regression models. In two Forrest plots, the effects of the different studies with a major difference in anesthetic dose or extubation protocol on the ICU stay were presented. The CHEC-list score was not included in the models.

Results

Study inclusion

Figure 1 summarizes the process of inclusion of the studies for review. Of the 643 studies that were selected by the performed search, we ex-

cluded 367 studies as they did not evaluated fast-track treatments. Another 117 studies were excluded because the study population contained a majority of non-CABG patients (n=106), or consisted of high-risk patients (n=2), children (n=6), or animals (n=3). A total of 83 studies were excluded due to study-related characteristics: no randomised clinical study (n=77), abstract (n=1), reviews (n=2), Japanese language (n=1), not traceable (n=2). The remaining 76 studies were randomised, controlled trials reporting fast-track treatment in low-risk CABG patients. Thirty of them also reported IC stay and fulfilled all inclusion criteria. However, two studies, one performed by Cheng et al. (Cheng, 1996a) and the other performed by Howie et al. (Howie, 2001) were published in three additional reports (Cheng, 1996b; Cheng, 2001; Cheng, 2003). The data of these three additional reports were only used once for analysis.

Figure 1: Process of inclusion of studies for review and analysis
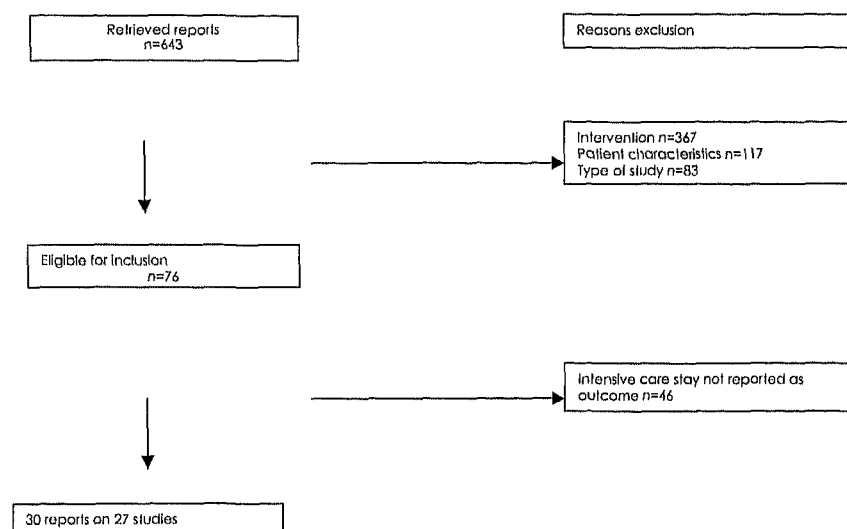
Study characteristics and study quality

The 27 included studies contained a total of 2.821 patients; most of them were published after the year 2000. Due to classification of the intervention as described in 'Materials and Methods', 12 studies (Berry, 1998; Cheng, 1996a; Engoren, 2001; Hall, 1991; Kadoi, 2003; Michalopoulos, 1998; Myles, 2002; Quasha, 1980; Reyes, 1997; Royse, 2003; Sherry, 1996; Shroff, 1997) showed a contrast between both study arms (table 1). Nine of these studies (Berry, 1998; Cheng, 1996a; Engoren, 2001; Hall, 1991; Kadoi, 2003; Myles, 2002; Royse, 2003; Sherry, 1996; Shroff, 1997) were categorized as with as having a major difference in anesthetic dose. In two of them (Berry, 1998; Cheng, 1996a) the extubation protocol was also categorized as different. In three studies (Michalopoulos, 1998; Quasha, 1980; Reyes, 1997) the extubation protocol was the only contrast. The other 15 studies (Bowler, 2002; Engoren, 1998; Fillinger, 2002; Howie, 2001; Latham, 2000; Mangano, 1992; Mollhoff, 2001; Mora, 1995; Myles, 1997; Petter, 2003; Samuelson, 1986; Simeone, 2002; Sulzer, 2001; Tritapepe, 2002; Zarate, 2000) did not show any differences between either trial arms when using our categorization criteria (table 2). Only one study fulfilled the three methodological quality criteria we evaluated. Ten of the 27 studies did not adequately fulfill any of these criteria. In table 3, a summary of the pooled data of all reviewed studies is given. The CHEClist score of the seven studies that included data on costs varied between 5 to 12 (table 4). In five of the studies (Engoren, 2001; Engoren, 1998; Mora, 1995; Myles, 2002; Sherry, 1996), the economic study design was not appropriate to the stated objective. Furthermore, in most of the studies not all relevant costs were identified and the identified costs were not measured (Engoren, 2001; Engoren, 1998; Fillinger, 2002; Mora, 1995; Sherry, 1996) and valued appropriately (Engoren, 2001; Engoren, 1998; Fillinger, 2002; Mora, 1995; Myles, 2002). None of the studies performed a sensitivity analysis. No statements on potential conflicts of interest were presented (scoring on separate items not shown).

Intensive care stay and hospital stay

Table 5 shows the results of the six regression models, in the first three models ICU stay was the dependent variable, and in the last three models, hospital stay was the dependent variable. In all models, anesthetic dose, temperature management, and extubation protocol are the independent variables. Models 1 and 4 show the results of the studies with a major difference in anesthetic dose and/or extubation protocol, models 2 and 5 the results of the studies without a major difference in anesthetic dose and/or extubation protocol, models 3 and 6 show the results of all 27 studies. Due to coding (1, 0, -1), the height of the B is not correspond with the number of hours or days reduction in ICU or hospital stay. All models except model 2 showed that extubation time has a significant effect on the ICU and hospital stay. If the quality variables and the year of publication are included in the models, the explained variance of all models improved. However, if these variables were eliminated the variable 'extubation protocol' stayed the important predictor of ICU and hospital stay. The Forrest plot (figure 2) with the studies (n=7) reporting a major difference in anesthetic dose shows that the type of anesthetic dose (low vs. high) has no effect on ICU stay. For the studies (n=5) who were categorized reporting a major difference in extubation protocol, a small but significant difference in ICU stay between extubation within or after 8 hours after surgery was found (figure 3).

The significance of the other variables in some of the models (for instance temperature management in model 5 and anesthetic dose in model 3) may be the result of multiple testing.

33

Table 1: Summary of studies with a major difference in anesthetic dose and/or extubation protocol

Author , year	no. of patients	anesthetic dose	Temp. Manage.	Extub. Prot. Hrs	ITT	allocation concealment	blinding	ICU stay Hrs N	mean	Hospital stay, days N	mean	MI N	n	Death N	n
Studies (n=7) with a major difference in anesthetic dose															
Kadoi, 2003	180	low	normo	-	unclear	unclear	unclear	90	120	-	-	-	-	-	-
		high	normo	-				90	144	-	-	-	-	-	-
Royse, 2003	76	low	normo	< 8	no	unclear	unclear	37	45.6	37	6.9	-	-	37	0
		high	normo	< 8				39	48.1	39	7.2	-	-	39	1
Myles, 2002		low	-	< 8				24	23	24	7.8	24	3	-	-
	77	high	-	< 8	no	adequate	adequate	24	22	24	5	24	0	-	-
		low	-	< 8				29	23	29	5.9	29	0	-	-
Engoren, 2001		low	normo	< 8				30	18.8	30	5	-	-	30	1
	90	high	normo	< 8	unclear	unclear	unclear	30	19.8	30	5	-	-	30	0
		low	normo	< 8				30	21.5	30	5	-	-	30	0
Shroff, 1997	21	low	-	< 8	yes	adequate	adequate	12	55	-	-	-	-	-	-
		high	-	< 8				9	65	-	-	-	-	-	-
Sherry, 1996	70	low	normo	-	no	unclear	unclear	37	22.4	-	-	-	-	-	-
		high	normo	-				33	23.6	-	-	-	-	-	-
Hall, 1991	47	low	hypo	-	yes	unclear	adequate	24	50	-	-	24	1	24	0
		high	hypo	-				23	59.8	-	-	23	1	23	0
Studies (n= 2) with a major difference in anesthetic dose and extubation protocol															
Berry, 1998	85	low	normo	> 8	yes	unclear	unclear	43	48	43	8	43	2	43	2
		high	normo	< 8				42	48	42	8	42	2	42	0
Cheng, 1996a*	102	low	normo	< 8	no	unclear	adequate	51	29.2	51	7	60	0	60	1
		high	normo	> 8				51	30.9	51	8.2	60	4	60	3
Studies (n= 3) with a major difference in extubation protocol															
Michalopoulos, 1998	144	high	-	< 8	unclear	unclear	adequate	72	16	72	7.3	72	0	72	0
		high	-	> 8				72	23	72	8.4	72	0	72	0
Reyes, 1997	404	high	normo	< 8	yes	unclear	unclear	201	27	-	-	121	4	201	15
		high	normo	> 8				203	44	-	-	151	5	203	18
Quasha, 1980	36	-	-	> 8	no	unclear	adequate	20	57	-	-	20	3	-	-
		-	-	< 8				16	46	-	-	18	0	-	-

- = not reported, IC = Intensive Care Unit, MI = myocardial infarction, Temp. Manage. = temperature management,
Extub. Prot., = extubation protocol, ITT= Intention-to-treat analysis, Normo = normothermia, Hypo = Hypothermia,
*also Cheng 2003 and Cheng 1996b

Table 2: Summary of studies without a major difference in anesthetic dose and/or extubation protocol

Author, year	no. of patients	anesthetic dose	Temp. Manage.	Extub. Prot. Hrs	ITT	allocation concealment	blinding	ICU stay Hrs		Hospital stay Days		MI		Death	
								N	mean	N	mean	N	n	N	n
Studies (n=15) without a major difference in anesthetic dose or extubation protocol															
Petter, 2003	34	-	hypo	< 8	no	unclear	unclear	18	23	-	-	-	-	-	-
		-	hypo	< 8				16	23.2	-	-	-	-	-	-
Simeone, 2002	49	low	-	< 8	unclear	unclear	unclear	24	29	-	-	-	-	-	-
		low	-	< 8				25	46.1	-	-	-	-	-	-
Bowler, 2002	24	low	normo	< 8	unclear	unclear	adequate	12	12	12	7	-	-	12	0
		low	normo	< 8				12	27	12	13	-	-	12	0
Tritapepe, 2002	195	low	-	< 8	unclear	unclear	adequate	65	28	65	7.4	65	2	65	4
		low	-	< 8				65	26	65	7.8	65	1	65	3
		low	-	< 8				65	20	65	7.5	65	1	65	3
Sulzer, 2001	36	low	normo	< 8	no	unclear	unclear	16	21.5	-	-	-	-	-	-
		low	normo	< 8				20	21.2	-	-	-	-	-	-
Howie, 2001*	304	low	normo	< 8	unclear	unclear	adequate	154	24	154	4.9	-	-	154	2
		low	normo	< 8				150	26.4	150	5	-	-	150	3
Mollhoff, 2001	293	low	-	< 8	no	adequate	adequate	148	21.6	148	6.1	148	3	-	-
		low	-	< 8				145	26.4	145	6.9	146	4	-	-
Latham, 2000	40	low	normo	< 8	no	unclear	unclear	20	31.2	20	5.6	20	0	-	-
		low	normo	< 8				20	26.4	20	5.2	20	1	-	-
Zarate, 2000	40	low	-	< 8	yes	unclear	unclear	20	31.2	20	5.6	-	-	-	-
		low	-	< 8				20	26.4	20	5.2	-	-	-	-
Engoren, 1998	70	low	-	< 8	unclear	unclear	unclear	35	23	35	6	-	-	35	0
		low	-	< 8				35	21	35	5.1	-	-	35	0
Myles, 1997	124	low	-	< 8	yes	unclear	unclear	66	41	-	-	66	2	66	0
		low	-	< 8				58	39	-	-	58	0	58	0
Fillinger, 2002	60	high	normo	< 8	no	unclear	unclear	30	30	-	-	-	-	30	0
		high	normo	< 8				30	31.7	-	-	-	-	30	1
Mora, 1995	90	high	-	-	unclear	unclear	adequate	24	82	-	-	-	-	24	1
		high	-	-				22	66	-	-	-	-	22	1
		high	-	-				23	66	-	-	-	-	23	0
		high	-	-				21	71	-	-	-	-	21	0
Mangano, 1992	106	high	hypo	> 8	unclear	unclear	adequate	54	84	54	11.2	54	1	54	2
		high	hypo	> 8				52	103.2	52	13.2	52	2	52	0
Samuelson, 1986	24	high	normo	-	yes	unclear	unclear	10	29.9	-	-	-	-	-	-
		high	normo	-				14	28.1	-	-	-	-	-	-

see table 1 for abbreviations, *also Cheng 2001.

Table 3: Pooled data of all studies categorized by type of intervention

Intervention		IC stay, Hrs		Hospital stay, Days		MI		Death	
Studies (n=12) with a major difference in anesthetic dose and/or extubation protocol									
Overall anesthetic dose		N	Mean	N	Mean	N	n	N	n
	high	889	44.38	330	7.38	565	16	742	37
	low	407	51.98	244	6.62	180	6	224	4
	-	36	52.11	-	-	38	3	-	-
Overall temperature management									
	hypothermia	47	54.80	-	-	47	2	47	0
	normothermia	1007	51.87	383	6.37	477	17	805	41
	-	278	27.61	221	7.28	259	6	144	0
Overall extubation protocol									
	> 8 hrs	389	39.51	166	8.23	47	2	378	23
	< 8 hrs	676	28.37	438	6.47	390	9	571	18
	-	297	94.08	-	-	47	2	47	0
Studies (n=15) without a major difference in anesthetic dose and/or extubation protocol									
Overall anesthetic dose		N	Mean	N	Mean	N	n	N	n
	high	280	67.41	106	12.18	106	3	717	15
	low	1175	26.71	966	6.15	653	14	256	5
	-	34	23.09	-	-	-	-	-	-
Overall temperature management									
	hypothermia	861	32.26	598	6.66	613	13	479	12
	normothermia	488	25.79	368	5.33	40	1	388	6
	-	140	76.34	106	12.18	106	3	106	2
Overall extubation protocol									
	> 8 hrs	106	93.42	106	12.18	106	3	106	2
	< 8 hrs	1269	26.81	966	6.15	653	14	777	16
	-	114	26.81	-	-	-	-	90	2
All studies (n=27)									
Overall anesthetic dose		N	Mean	N	Mean	N	n	N	n
	high	1169	49.90	436	8.55	671	19	1459	52
	low	1582	33.21	1210	6.24	833	20	480	9
	-	70	38.01	-	-	38	3	-	-
Overall temperature management									
	hypothermia	908	33.43	598	6.66	660	15	526	12
	normothermia	1495	43.36	751	5.86	517	18	1193	47
	-	418	43.93	327	8.87	365	9	250	2
Overall extubation protocol									
	> 8 hrs	495	51.05	272	9.77	153	5	484	25
	< 8 hrs	1945	27.35	1404	6.25	1043	23	1348	34
	-	411	75.42	-	-	47	2	137	2

Table 4: Data of studies reporting costs

Author, year	CHEC-list	Design	Perspective	Year	Currency	Time Horizon	Cost Identified (categories)	Cost Valuation	Conclusion Concerning Total Costs
Studies (n=3) with a major difference in anesthetic dose									
Myles, 2002	11	CMA	hospital	1999	AUS $	operating theater ICU stay	healthcare (surgery, anesthetic, ICU)	unclear	no significant difference between the two groups
Engoren, 2001	8	CMA	hospital	unclear	US $	hospital stay	healthcare (anesthetic, hospital)	unclear	no significant difference between the two groups
Sherry, 1996	8	CMA	hospital	1993	UK pounds	until ICU discharge	healthcare (nursing, drug)	salary scales and acquisition cost	no significant difference between the two groups
Studies (n=1) with a major difference in anesthetic dose and extubation protocol									
Cheng, 1996a*	12	CMA	insurer	unclear	CAN $	until 12 months	healthcare (hospital, rehabilitation days, physician visits, cardiac drugs	fees, expenditures	significant reduction of costs in low dose fentanyl and early extubation group
Studies (n=3) without a major difference in anesthetic dose									
Engoren, 1998	9	cost-analysis	-	unclear	US $	unclear	healthcare (drugs)	acquisition cost	significant lower in one group
Fillinger, 2002	9	CMA	hospital	unclear	US $	hospital stay	healthcare (unclear)	charges	no significant difference between the two groups
Mora, 1995	5	cost-analysis	-	unclear	US $	unclear	healthcare (drugs)	acquisition cost	unclear

- = not reported, CHEC-list = Consensus Health Economic Criteria list, CMA= Cost minimization analysis, ICU= Intensive Care Unit, *also Cheng 2003 and Cheng 1996a

Figure 2: Forrest plot of studies (n=9) with a major difference in anesthetic dose

	Standardis⁻d Me⁻n Difference (95% CI)	% W⁻ight
Kadoi 2003	-0.27 (-0.57,0.02)	24.6
Royse 2003	-0.17 (-0.62,0.28)	10.5
Myles 2002	0.11 (-0.37,0.60)	9.1
Engoren 2001	0.06 (-0.38,0.49)	11.1
Shroff 1997	-0.28 (-1.15,0.59)	2.8
Sherry 1996	-0.01 (-0.48,0.46)	9.6
Hall 1991	-0.25 (-0.83,0.32)	6.4
Berry 1998	0.00 (-0.43,0.43)	11.7
Cheng 1996	-0.14 (-0.53,0.25)	14.1
Overall (95% CI)	-0.11 (-0.26,0.03)	

-1.14574   0   1.14574

Standardised Mean difference
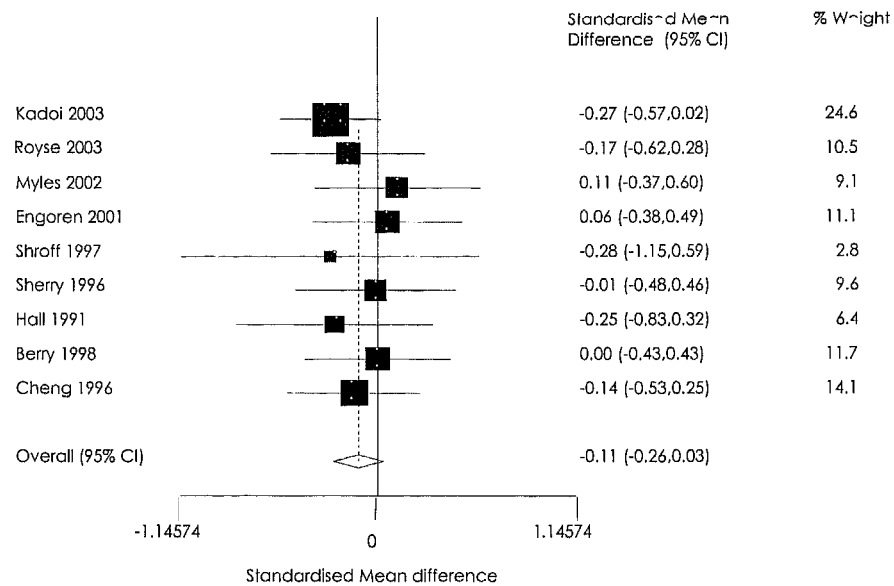
Figure 3: Forrest plot of studies (n=5) with a major difference in extubation protocol

	Standardised Mean Difference (95% CI)	% Weight
Berry 1998	0.00 (-0.43,0.43)	13.0
Cheng 1996	-0.14 (-0.53,0.25)	15.6
Michalopoulos 1998	-5.60 (-6.32,-4.87)	4.4
Reyes 1997	-0.04 (-0.24,0.15)	61.7
Quasha 1980	-0.48 (-1.14,0.19)	5.3
Overall (95% CI)	-0.32 (-0.48,-0.17)	

-6.32368   0   6.32368

Standardised Mean difference

38

Table 5: Regression models (1-4) of the different studies with or without a major difference in anesthetic dose and/or extubation protocol

Model 1 dependent ICU [a] stay: Studies (n=12) with a major difference in anesthetic dose and/or extubation protocol

R-Square = .714	B	P-value
B_0	2019.328	.320
Anesthetic dose	-2.552	.394
Temperature management	-5.143	.552
Extubation protocol	-3.497	.000
Intention-to-treat	13.524	.294
Allocation Concealment	6.932	.448
Blinding	-4.329	.678
Year of publication	-.998	.325

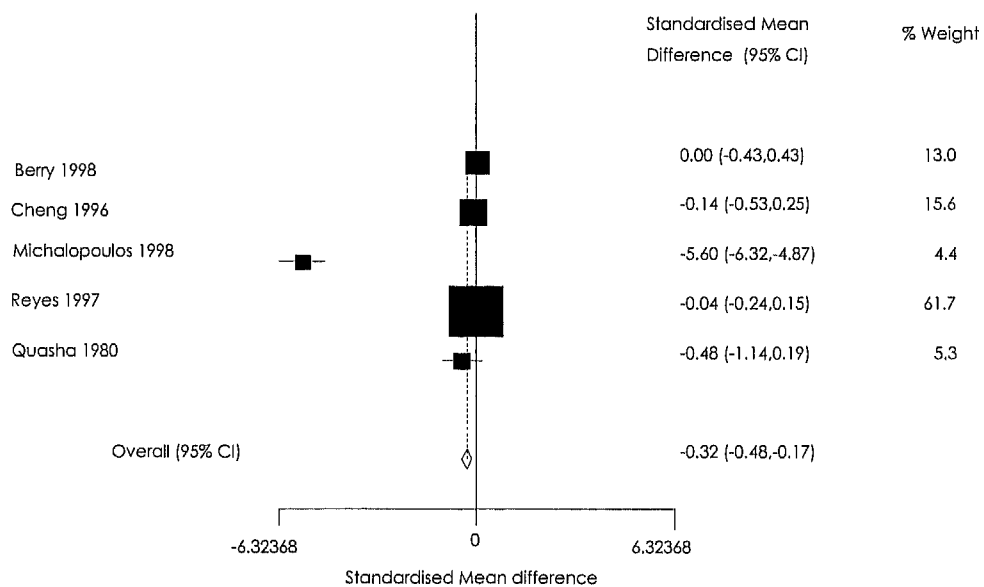Model 2 dependent ICU stay: Studies (n=15) without a major difference in anesthetic dose and/or extubation protocol

R-Square = .514	B	P-value
B_0	-408.534	.802
Anesthetic dose	1.757	.781
Temperature management	6.857	.013
Extubation protocol	-16.979	.282
Intention-to-treat	-10.358	.201
Allocation Concealment	-13.997	.064
Blinding	3.310	.059
Year of publication	.228	.781

Model 3 dependent ICU stay: All studies (n=27)

R-Square = .619	B	P-value
B_0	681.038	.453
Anesthetic dose	-4.826	.001
Temperature management	-.180	.938
Extubation protocol	-3.423	.000
Intention-to-treat	5.020	.431
Allocation Concealment	2.180	.628
Blinding	.155	.926
Year of publication	-.329	.469

Model 4 dependent hospital stay: Studies (n=6) with a major difference in anesthetic dose and/or extubation protocol

R-Square = .933	B	P-value
B_0	-1947.845	.010
Anesthetic dose	.160	.446
Temperature management	-1.878	.073
Extubation protocol	-.459	.019
Intention-to-treat	5.185	.003
Allocation Concealment	-4.568	.013
Blinding	6.923	.006
Year of publication	.975	.010

Table 5: (continued) regression models (5-6) of the different studies with or without a major difference in anesthetic dose and/or extubation protocol

Model 5 dependent hospital stay: Studies (n=8) without a major difference in anesthetic dose and/or extubation protocol

R-Square = .550	B	P-value
B_o	-1500.374	.011
Temperature management	1.377	.005
Extubation protocol	-5.091	.001
Intention-to-treat	-1.455	.024
Allocation Concealment	-4.479×10^{-2}	.946
Blinding	-.739	.395
Year of publication	.756	.011

Model 6 dependent hospital stay: All studies (n=27)

R-Square = .708	B	P-value
B_o	-108.354	.613
Anesthetic dose	.274	.156
Temperature management	-3.287×10^{-2}	.943
Extubation protocol	-.617	.002
Intention-to-treat	1.325	.019
Allocation Concealment	-.586	.477
Blinding	1.450	.016
Year of publication	5.735×10^{-2}	.593

[a] ICU= Intensive Care Unit

Myocardial infarction and death

In six additional regression models, the influence of the different aspects of fast-track treatment on the percentage MI and death was evaluated. Due to a limited number of studies reporting MI and death, the three quality variables and the year of publication could not be taken into account. In none of the models were the anesthetic dose, temperature management or extubation protocol significant predictors for the percentage MI and death (data not shown).

Quality of life and cost-effectiveness

Ten of the 27 reviewed studies reported QoL as study outcome (table 6). Most of these studies reported pain (Bowler, 2002; Engoren, 2001; Fillinger, 2002; Howie, 2001; Mollhoff, 2001; Royse, 2003; Shroff, 1997; Zarate, 2000) and/ or cognitive function (Cheng, 1996a; Howie, 2001; Kadoi, 2003).

As the instruments used and the time point measured varied across studies, it was impossible to pool the data.

Only one of the studies reporting pain categorized in the group -with a major difference in anesthetic dose or/and extubation protocol- was able to reduce the pain in the low-dose group significantly more compared with the high-dose group (Royse, 2003). The two studies in this category of studies reporting cognitive function did not find a difference between both treatment groups.

Seven studies reported data on costs (table 4). Five included studies (Cheng, 1996a; Engoren, 2001; Fillinger, 2002; Myles, 2002; Sherry, 1996) refer to cost-minimization analysis and two studies (Engoren, 1998; Mora, 1995) to cost analysis. All cost-minimization analyses were from a hospital perspective except the study performed by Cheng et al. (Cheng, 1996a), which used an insurer perspective. Because the costs categories used and the time horizon of the measured costs varied, it was impossible to pool the cost data. Although the conclusions of the majority of economic evaluations do not report a decrease of costs, the study (Cheng, 1996b) with the highest CHEC-list score reports a significant decrease in costs.

Discussion

This meta-analysis evaluated the effect of three aspects of fast-track treatment (anesthetic dose, extubation protocol and temperature management). Results indicate that use of a protocol, directed toward early extubation (within 8 hrs after surgery) results in a substantial shortening of the total ICU and hospital stay of low-risk CABG patients. This effect was observed not only in studies that were categorized into different extubation protocols but also in all other studies and persisted when quality parameters were included in the analysis. An influence of the type of fast-track treatment on mortality or the prevalence of postoperative MI was

not observed. In general, the methodological quality of the included studies was moderate, and only one study fulfilled the three evaluated criteria. Similarly, only one of the seven studies that reported data on costs had an acceptable score on the CHEC-list (Donaldson C, 2002). However, the highest-quality study demonstrated a cost reduction.

Fast-track treatment

For decades, the most common anesthetic technique for CABG was anesthesia with a high dose of opioids, such as fentanyl in the range of 50-100 µg/kg. High-dose-opioids anesthesia provides excellent intra - operative hemodynamic stability and absence of adverse events; how-ever, prolonged ventilatory support was needed (Stanley & Webster, 1978). A shift from high- to low-dose opioids (e.g., <15µg/kg fentanyl) was made in the late 1990s. The use of low-dose opioids theoretically results in extubation within a few hours postoperatively. Recently, Remifentanyl (Rosow, 1993), an anesthetic with a half-life time of about 5-8 mins, was introduced to practice, it would permit an even more prompt tracheal extubation compared to low dose fentanyl.

Temperature management is the second important aspect of fast-track treatment. Patients need to be normothermic (>35°C) before weaning can be started. As hypothermic patients (<35°C) need extra time before adequately warmed, hypothermia interferes with early recovery after surgery (Karski, 1995).

Early extubation is also essential for fast-track treatment and is mostly de-fined as extubation within 1 to 8 hours after surgery (Karski, 1998). These early extubation protocols are mainly use the following criteria: patient alert, hemodynamically stable, temperature 35.5-36.5°C, chest tube drainage of <100 ml for 2 hours, no intraaortic balloon pump, and ab-sence of new or uncontrolled arrhythmias (Cheng, 1996a; Cheng, 1996b; Cheng, 2003; Michalopoulos, 1998; Mollhoff, 2001; Petter, 2003; Quasha, 1980; Sulzer, 2001)

42

Intensive care stay and hospital stay

Numerous non-randomised studies showed that early extubation or fast-track treatments (Arom, 1995; Chong, 1993; Chong, 1992; Engelman, 1996; Engelman, 1994; Foster, 1984; Klineberg, 1977; Konstantakos & Lee, 2000; Lazar, 2001; Prakash, 1977) resulted in a decrease of ICU and hospital stay. As stated correctly by others, the above-described improvements in the perioperative and postoperative treatments do not automatically reduce ICU and hospital stay (Bowler, 2002). However, to accomplish a reduction in hospital and ICU stay, it is important to identify the important predictors. Our review shows that a protocol directed on early extubation could be regarded as a crucial step to reduce ICU and hospital stay. Although normothermic temperature management and fast-track anesthesia facilitate early extubation, it seems that extubation based on systematic application of criteria for stability is an essential step. Indeed, only extubated patients can be transferred to lower care facilities. Interestingly, the anesthetic dose was not an important predictor of ICU stay. On one hand, this may be related to the categorization of the different types of anesthesia, in only two categories in our study. On the other hand, many of the studies were directed on the evaluation of safety and did not adapt ICU discharge policies (Myles, 1997; Myles, 2002; Tritapepe, 2002). Indeed, only four studies reported their IC discharge criteria (Cheng, 1996a; Howie, 2001; Myles, 2002; Tritapepe, 2002). It is also possible that, as stated by others, the anesthetic dose is not the rate-limiting step in the fast-track process but the patient's preoperative physical status (age, sex) and postoperative complications (atrial arrhythmias and bleeding) (Arom, 1995; Bando, 1997; Doering, 1998; Habib, 1996).

Table 6: Data of studies reporting quality of life

author, year	Dimension of QoL	Instrument	Measurement moments	Conclusion
Studies (n=4) with a major difference in anesthetic dose				
Kadoi, 2003	Cognitive function	Mini-Mental State Test	preop, 6 months postop	no significant difference between the two groups
		Rey auditory Verbal learning test		
		Trail-Making Test (part A and B)		
		Digit span forward and grooved pegboard		
Royse, 2003	Pain	visual analogue scale (0 to10)	1-3 days postop	significant lower pain (1-2 days)
	Posttraumatic stress	subscales for posttraumatic stress and depression of Minnesota	2,6 weeks postop	less posttraumatic stress and depression in low dose group compared to high dose group
	Depression	Multiphasic Personality inventory 2 (MMPI 2)	preop, 3 days after cessation of analgesia	
	Somatic sensation	von Frey hairs		no significant difference between the two groups
Engoren, 2001	Pain	101-point rating scale (0 to 100)	30 mins postop and 6:30 AM	no significant difference between the two groups
Shroff, 1997	Pain	pain score (0 to 10)	1-24 hrs after surgery	no significant difference between the two groups
Studies (n=1) with a major difference in anesthetic dose and extubation protocol				
Cheng, 1996a*	Cognitive function	Mini-Mental State Test	4 and 48 hrs after surgery	no significant difference between the two groups
Studies (n=5) without a major difference in anesthetic dose				
Bowler, 2002	Pain	visual analogue scale (0 to100)	1-6, 12 and 24 hrs after extubation	significant lower in one group until 2 hrs after extubation
Howie, 2001**	Pain	pain score (0 to 10)	baseline, 1-3 days postop	significant more muscle ache (pain) in one group
	Cognitive function	Mini-Mental State Test		No significant difference between the two groups
Mollhoff, 2001	Pain	self-rated scale (0 to 4)	every 10 min during down titration opioids	significant more postoperative ache(s) in one group
			15-30-60-90-120 min after discontinuation of opioids	
Zarate, 2000	Pain	visual analogue scale (0 to100)	1-2-4-8-12-24-48 hrs after extubation	significant more postoperative pain in one group in all points in time
Fillinger, 2002	Pain	visual analogue scale (0 to100)	first day postop	no significant difference between the two groups

QoL= quality of life, preop= preoperative, postop= postoperative, mins= minutes , hrs= hours, *also Cheng 2003 and Cheng, 1996b, ** also Cheng 2001

The factor temperature management was also not significantly related to ICU and hospital stay. The explanation for this can be that in none of the 27 studies we evaluated was the intervention itself directed toward temperature management. Furthermore, only two of the studies reported hypothermia treatment, so the effect of this variable in the regression analysis is very limited (Hall, 1991; Petter, 2003).

ICU and hospital stay are not exclusively dependent of the clinical readiness for discharge. Few studies tried to overcome this problem by reporting besides the actual ICU discharge time, also the eligibility for ICU discharge time (Cheng, 1996a; Cheng, 1996b; Cheng, 2003; Mollhoff, 2001). However, from a cost perspective, it is not important to know when a patient is clinical stable enough to be transferred from the ICU to a lower-care facility but when she/he actually is transferred. The eligibility data however, give an indication for further reduction of stay if all other circumstances are optimal.

Quality of life and cost-effectiveness

The instruments and measurement moments used for the evaluation of QoL varied over the reviewed trials. When future research is planned in this field its advisable to standardize this; in this way, comparability of study outcomes is possible, and it enables reviewers to pool data. Therefore, we recommend the use the visual analogue scale, (range from 0-100) at 12 and 24 hours after surgery and at hospital discharge. The visual analogue scale was found to be methodologically sound, conceptually simple, and easy to administer (Coll, 2004). Furthermore, we recommend to measure different domains of QoL with instruments with established reproducibility, validity and responsiveness (Guyatt, 1989). Ideally, a generic instrument (like, EuroQoL) and disease-specific (for instance, angina questionnaire) or domain-specific questionnaire (e.g., Beck Depression Inventory) should be used. These questionnaires should be filled out preoperatively and 1 month postoperatively. An additional follow-up as-

sessment can be added, depending on the purpose of the study. Furthermore, guidelines for reporting the QoL results should be used (Staquet, 1996).

Although it seems obvious that fast-track treatments reduce the total costs of the treatment of CABG patients, the evidence is mainly based on non-randomised studies (Chong, 1993; Engelman, 1994). Only one study (Cheng, 1996a) with a acceptable CHEC-list score (Donaldson C, 2002) concluded that fast-track treatment is a cost saving approach. In the other studies (cost-effectiveness was not the main study outcome subject) that did not find a significant decrease in costs had poor quality. In our opinion for future research in this field, the quality economic evaluations need to be improved. This can be achieved by incorporation of the following aspects:

1. Appropriate design should be used. A cost-minimization analysis is only suitable if the effects in both treatment arms of the trials are equal. However, more appropriate would be, even if effects are not significantly different, to use a cost-effectiveness design (Briggs & O'Brien, 2001).

2. More detailed costing should be applied and reported. This means, identifying and reporting all relevant cost-items. Furthermore, measure and value all resources changes that can occur as a certain healthcare interventions are performed, not only reporting and calculating drug costs, but also these of all other relevant cost categories (such as blood samples, personnel and hospital stay). In addition, the real costs instead of fees or charges should be used to value the resources. The time horizon should be until no influence of the intervention on the outcome can be expected. For instance, for anesthetic interventions the time frame should be until 1 month after surgery instead of until ICU or hospital discharge.

3. Up-to-date analysis techniques and presentation of cost-effectiveness data should be used -for instance, reporting uncertainty around estimated mean outcome variables (e.g., in percentiles of standard devia-

tions), reporting confidence intervals of cost differences, and performing sensitivity analysis of key variables.

For further reading, see the two handbooks (Drummond, 1996; Gold, 1996).

Methodological aspects of this review

1. Although the included studies were randomised trials, a meta-regression across trials does not have the benefit of randomisation to underpin a causal interpretation (Thompson & Higgins, 2002). Therefore, the interpretation of the results found incidentally most taken with precaution. However, the main finding of this review, that the use of early extubation was the important predictor of ICU and hospital stay, was a constant outcome.

2. The use of ICU and hospital stay as outcome. On one hand, ICU and hospital stay are not only dependent on the patient's clinical situation but also other factors such as availability of beds in step-down units, common practice at ICU, and preferences of medical personnel (Hwang, 1999; Velasco, 1995). Furthermore, transfer of patients during nighttime may not be warranted because it may result in sleep disturbances that can lead to psychosis and delirium (Gelling, 1999). On the other hand, not all studies reported the ICU and hospital stay as mean and standard deviation values as needed for analysis but in mean or median and (interquartile) range or percentiles (O'Rourke, 2002). The real variance of the study outcomes may therefore be different.

3. The moderate methodological quality of the included studies. However, the results of this meta-regression were constant.

4. As we focused on studies that evaluated fast-track treatments and reported ICU stay, we excluded studies that only reported safety in terms of morbidity and mortality. However, reviews which focused on safety also found no difference on MI and death between different treatments (Hawkes, 2003; Meade, 2001; Myles, 2003).

47

Implications for further research and conclusion

Although fast-track anesthetics and normothermic temperature management facilitate early extubation, the introduction of an early extubation protocol seems essential to accomplish a decrease of ICU and hospital stay in low-risk CABG patients. Further trials should evaluate the effect of adding up ICU discharge criteria to fast-track protocols. Furthermore, randomised clinical trials of a good methodological quality are to be preferred. Ideally, cost-effectiveness evaluations should be up to the current standard (Drummond, 1996; Gold, 1996)

Appendix 1: Search strategy

((random* in AB) or (random* in TI) or (((singl* or double* or trebl* or tripl*) near (blind* or (mask* in TI))) or ((singl* or double* or trebl* or tripl*) near (blind* or (mask* in AB))))) or ((singl* or double* or trebl* or tripl*) near (blind* or mask*)) or (explode CLINICAL-TRIALS/ all subheadings) or ((clin* near trial*) in AB) or ((RANDOMIZED-CONTROLLED-TRIALS) or (CON-TROLLED-CLINICAL-TRIAL in PT) or (RANDOMIZED-CONTROLLED-TRIAL in PT) or (CLINICAL-TRIAL in PT) or (SINGLE-BLIND-METHOD) or (DOUBLE-BLIND-METHOD) or (RANDOM-ALLOCATION)) or ((clin* near trial*) in TI)) and ((conventional near extubation) or (early near extubation) or (fast-track cardiac anesthesia) or (fast-track anesthesia) or (fast near track) or (recovery) or ("Anesthesia-Recovery-Period" in MIME,MJME) or (removal near endotracheal) or (fast extubation) or (mechanical near ventilation) or (late near extubation) or (routine near extubation) or (remifentanyl or remifentanil) or (sufentanil or sufentanyl) or (alfentanil or alfentanyl) or (fentanyl or fentanil or fentanyl*)) and ((Myocardial Revascularization) or (cardiac surgery) or (heart surgery) or ("Thoracic-Surgery"/ all subheadings) or (explode "Cardiac-Surgical-Procedures"/ all subheadings) or (explode "Coronary-Artery-Bypass"/ all subheadings) or (Myocardial Revascularisation)) and (intensive care or intensive or intensive care units or intensive care unit or IC or IC or recovery or length of stay) and (((AGE:MEDS = ADOLESCENT) or (AGE:MEDS = ADULT) or (AGE:MEDS = AGED) or (AGE:MEDS = MIDDLE-AGED)) and (TG:MEDS = HUMAN))

CHAPTER

3

Short-stay intensive care after coronary artery bypass surgery: randomised clinical trial on safety and cost-effectiveness

Ghislaine van Mastrigt
John Heijmans
Hans Severens
Erik Fransen
Paul Roekaerts
Gemma Voss
Jos Maessen

51

Abstract

Objective: To evaluate the safety and cost-effectiveness of short-stay intensive care treatment for low-risk coronary artery bypass patients.

Design: Randomised clinical equivalence trial.

Setting: University Hospital Maastricht, the Netherlands.

Patients: low-risk coronary artery bypass patients.

Interventions: A total of 600 patients were randomly assigned to undergo either short-stay intensive care treatment (8 hrs intensive care treatment) or control treatment (care as usual, overnight intensive care treatment).

Measurements: The primary outcome measures were intensive care re-admissions and total hospital stay. The secondary outcome measures were total hospital costs, quality of life, postoperative morbidity and mortality. Hospital costs consisted of the cost of hospital admission or admissions and outpatient costs.

Main results: The difference in intensive care readmission between the two groups of 1.13% was very small and not significantly different (P-value =.241; 95% confidence interval: -0.9% to 2.9%). The total hospital stay (P-value=.807; 95% confidence interval: 1.2 to -0.4) and postoperative morbidity were comparable between both groups. The SSIC group's quality of life improved more compared to the control group (P-value .0238, 95% confidence interval: .0012 to .0464). The total hospital costs for SSIC were significantly lower (95% confidence interval: €-1,581 to €-174) compared to those for the control group (€4,625 and €5,441, respectively). The estimated incremental cost-effectiveness ratio (cost/ delta quality adjusted life months) thus showed the dominance of SSIC. Bootstrap and sensitivity analyses confirm the robustness of the study findings.

Conclusions: Compared with usual care, SSIC is a safe and cost-effective approach. SSIC can be a considered as an alternative for conventional postoperative intensive care treatment for low-risk coronary artery bypass patients.

Introduction

The reintroduction of off-pump coronary artery bypass grafting (CABG) has resulted in important changes in on-pump CABG as well. Improved extra corporal circulation technology for on-pump CABG and the use of low-dose opioid-based anesthetic techniques enabling early extubation have freed postoperative care from conventions that had held up for decades. Fast-track treatment after cardiac surgery is becoming very popular (Berry, 1998; Cheng, 1996a; Dumas, 1999; Engoren, 1998; Michalopoulos, 1998; Quasha, 1980; Reyes, 1997; Silbert, 1998; Sirio & Martich, 1999), although evidence-based research on safety and cost-effectiveness is limited. For that reason, not all cardiac centers apply this treatment.

A number of trials that evaluated early vs. late extubation suffered from methodological flaws in conducting intention-to-treat analyses (Meade, 2001). Furthermore, they did not focus on the safety of early intensive care (IC) discharge. To the best of our knowledge, only one trial with a reasonable quality (n=100) has been performed to evaluate the cost-effectiveness of early vs. late extubation (Cheng, 1996b). Although fast-track treatments offer the possibility of discharging CABG patients from IC within a few hours, more quantitative evidence on safety and cost-effectiveness is needed on whether or not to take the risk of early IC discharge.

The focus of this randomised clinical equivalence trial was to evaluate the safety and cost-effectiveness of early discharge -within 8 hours- of post-surgical CABG patients from IC to medium care (MC).

Materials and methods

This single blinded, randomised, clinical equivalence trial was designed to evaluate the safety of patients who received short-stay intensive care (SSIC < 8 hrs) and its cost-effectiveness compared with control group patients who stayed overnight in IC (care as usual) (box, figure and table 1 provide details on patient management).

To investigate this, the following two hypotheses were formulated to evaluate the primary outcomes: 1) SSIC patients have a comparable number of IC readmissions in the first postoperative month compared with patients in the control group and 2) there is no difference between either group's total hospital stay.

Another three hypotheses were formulated to assess the secondary outcomes: 1) postoperative morbidity and mortality is comparable in both groups, 2) total hospital costs in the SSIC group are lower compared with costs in the control group as a consequence of lower IC costs and 3) quality of life, measured by means of EQ-5D, is not different between the two groups.

Randomization and blinding

Before starting the study, a random list of permutated blocks (with a length of 10) was generated for either control ($n=300$) or SSIC ($n=300$) by a computer, and the results were entered sequentially into numbered, opaque envelopes, sealed with tape. Research nurses not involved in the treatment assigned consecutive patients once written consent and baseline measurement had been obtained. The envelopes were opened by the IC nurses in IC immediately postoperatively. Patients and the surgical team were blinded for the group assignments. It was impossible to blind IC-unit physicians who were responsible for discharge and readmission. However, the decision for readmission was made on the basis of objective criteria. Furthermore, the physicians were not aware

that IC readmission was one of the two primary endpoints of the study. Whenever possible, data entry and analysis were performed without knowledge of treatment allocation. The local ethical and research council approved the study.

Patient recruitment and eligibility criteria

All CABG patients who were admitted to the University Hospital of Maastricht in the period between February 2001 and March 2003 were considered for inclusion. High-risk CABG patients were separated from the low-risk patients by means of a risk stratification system, based on the criteria of Parsonnet et al. (Parsonnet, 1989) and Society of Thoracic Surgeons (STS, 2003) (box).

Clinical effectiveness

The medical data, obtained from different databases, case record forms, medical records, and discharge letters from both university (n=1) and district hospitals (n=5), were registered in a standardized manner. To include adverse events (morbidity and mortality), the follow-up time was ≤30 days after surgery. The power calculation was based on the primary outcome probability of IC readmission. The data on IC readmission rate was obtained from the University Hospital of Maastricht Cardio Thoracic Surgery department data registration system (SUMMIT). The expected probability of readmission for both groups was 5%. With a power of 80%, an alpha of 5% (one-sided), and an accepted IC readmission difference of 5% between SSIC and control groups, 300 patients were needed in each group. Mortality was defined as any death occurring within 30 days after surgery. The definitions of the Society of Thoracic Surgeons were used to describe postoperative morbidity (STS, 2003).

Costs

The cost analysis was performed from a hospital perspective with a time horizon from admission to IC until 30 days postoperative. Costs were expressed in euros (2001) and identified for all patients, including the cost of clinical and outpatient procedures and the cost of inpatient hospital days.

The resource use of clinical and outpatient procedures (e.g., electrocardiography, chest radiography, cardiologist visit) in the university hospital was measured by means of the hospital billing system. Data on outpatient procedures in district hospitals were obtained by questionnaires. The number of hours or days of inpatient hospital stay in the different departments (IC, MC, Low Care (LC)) in the university hospital were obtained from medical records, and data on LC inpatient stays in the district hospitals were obtained from discharge letters.

The calculation of unit costs of inpatient days was performed according to the Dutch guidelines for costing (Oostenbrink, 2003; Oostenbrink, 2000). Guideline prices were used for the cost categories: nutrition, laundry, accommodation, cleaning, overheads and equipment. As these prices were not available for inpatient stay on the MC department, the mean of IC and LC prices was applied. Guideline prices were converted into euros and price-indexed (CBS, 2003). Hospital-specific cost prices were calculated to estimate other cost categories (nursing, specialists and residents, material and medication) (Oostenbrink, 2003). Nursing costs were calculated on the basis of actual salary and occupation figures for the different departments. As specialists and residents have various tasks it was unachievable to calculate cost prices, so guideline prices were used (Oostenbrink, 2000). The cost of material and laundry was based on average use by a CABG patient. Medication costs were based on the cost calculation of a random sample of 50 patients in each group. The unit costs for one day's inpatient hospital stay were

€1,082, €502, €249, and €246, respectively, for IC, MC and LC in the university hospital and for LC in a district hospital.

All cost calculations were based on the actual situation, except for the SSIC patients who received their first hours of MC treatment in the IC department. The valuation of these hours of MC stay was based on a shadow price. Due to logistics, it was not possible to transfer these patients immediately from the IC department to MC (box). Therefore, because MC was provided, the unit prices of MC in the MC department were used as a shadow price for these hours of MC treatment in IC.

BOX 1: Patient management

Exclusion criteria

Exclusion criteria were: age > 78 yrs, ejection fraction < 30%, stage 3 obesity (body mass index >40 kg/m²), hemodialysis (kidney-replacing therapy), pulmonary hypertension (systolic pressure of < 40 mm Hg), recent cardiovascular accident (<1 month), recent myocardial infarction (<24 hrs), cardiogenic shock (systolic blood pressure of < 80 mm Hg, central filling pressure of > 20 mm Hg, cardiac index of <1.8 Lmin⁻¹m⁻²), need for inotropic therapy (>5mgμg⁻¹min⁻¹ dopamine or dobutamine), ongoing infarction (a significant increase of myocardial muscle creatine kinase isoenzyme within 4 hrs before surgery), or the need for intraaortic balloon pump. Other reasons for exclusion were, for instance, inability to give informed consent, inability to speak/read/understand the Dutch language, emergency surgery. For details on reasons of refusal and exclusion criteria, see figure 1.

Preoperative and Perioperative treatment

The preoperative and perioperative treatment protocol was the same for all CABG patients. The day before surgery, patients underwent a general screening (e.g., physical examination, electrocardiography, chest radiography, laboratory testing). During surgery, all patients were anesthetized with total intravenous infusions of propofol and a short acting opioid.

Postoperative treatment

Postoperative treatment for control group and SSIC group was different according to the study design. The flow chart shows the time lines for both groups (figure).

Intensive Care treatment protocol

Both control and SSIC group patients received standard IC monitoring, which consisted of: a five-lead electrocardiograph, a pulmonary artery catheter for continuous blood pressure management, and temperature probes for central and peripheral temperatures. Continuous cardiac output and mixed venous oxygen saturation were measured (Vigilance). At arrival in intensive care, electrocardiography and chest radiography were performed. Blood sampling was done at arrival, 2, 4, 8, 12 and 16 hrs after admission. Heart rate, blood pressure, pulmonary artery pressure, central venous pressure, occlusion pressure, oxygen saturation, and temperature were checked every 15 mins during the first 4 hrs. After 4 hrs, the checks were performed once every 30 mins. Chest tube drainage and urine production were checked every 30 mins. After extubation, measurements were performed every hour unless the postoperative course necessitated more frequent checks.

Intensive care discharge policy

To allow the transfer of SSIC patients from intensive care to medium care within 8 hrs, the discharge criteria described in table 1 had to be fulfilled. The SSIC patients who did not fulfill the discharge criteria within 8 hrs and all control group patients were treated according to the usual care.

Protocol adjustments

The discharge criteria were adapted at two points in time. First, the following criterion: Pao_2 of > 10 kPa and $Paco_2$ of < 6.5 kPa" was adapted to "Pao_2 of > 10 kPa and the $Paco_2$ < 6.0 kPa, depending on observation of the patient's clinical situation". This protocol adjustment was made on February 7, 2001.

The other protocol adjustment was done as follows to the criterion: "the blood gas sampling is done". To this criterion, the following part is added: "but the results do not need to have come back". This adjustment was made November 27, 2001.

Medium care treatment protocol

For logistical reasons, it was not possible to transfer SSIC patients who fulfilled the criteria for early discharge from intensive care to the medium care situated on another floor in the hospital. Consequently, the nurse/patient ratio of one nurse to one patient was maintained. However, the care provided by the nurse was adapted to the medium care guidelines. Thus, these patients received medium care treatment in the intensive care unit according to the following protocol:

1. One full blood sample was taken on arrival to medium care, and one arterial and venous blood sample were taken in the morning.
2. Heart rate, blood pressure, mixed venous oxygen saturation, temperature, respiratory rate, chest tube drainage, urine production, blood and fluid balance were measured every hour.
3. If stable for 4 hrs, checks were carried out after every 2 hrs.
4. If emergencies occurred, adherence to the protocol was abandoned, and all measures necessary for the safety of the patient were taken.
5. The pulmonary artery catheter stayed *in situ* for safety reasons because of the investigational conditions, after a patient was transferred to the medium care, but the monitor for the registration of the cardiac index was switched off.
6. The arterial catheter was left *in situ*, which is common use in our medium care.

Transfer of patients one day postoperatively

The morning after surgery, all patients were transferred to a medium care ward connected to the general ward. The nurse/patient ratio in this medium care was one nurse to two patients. Patients stayed there for another 24 hrs. The treatment protocol consisted of the following:

1. One full blood sample was taken on arrival to medium care.
2. Heart rate, blood pressure, mixed venous oxygen saturation, temperature, respiratory rate, chest tube drainage, urine production, blood and fluid balance were measured every hour.
3. If stable for 4 hrs, checks were carried out after every 2 hrs.
4. If emergencies occurred, adherence to the protocol was abandoned, and all measures necessary for the safety of the patient were taken.

On the second postoperative day patients were admitted to the general ward (low care), where they stayed until discharge or transfer to the district hospital.

Table 1: Description of criteria for discharge from ICU for SSIC and control group

Pulmonary

Patient needs to be exitubated for at least 30 mins before he or she can be transferred to the medium care unit
Last blood gas sampling is done, but the results do not need to have come back

Patient doesn't need > 5 L oxygen nasally

Breathing frequency $\geq$ 10 to $\leq$ 25 breaths/min

Pao_2 of > 10 kPa and $Paco_2$ < 6.0 kPa, depending on clinical observation, can be deviated from this value (in consultation with ward doctor)

In case of chronic obstructive pulmonary disease or with pulmonary-challenged patients, the arterial blood gas may not deviate > 20-25% from preoperative blood gas (in consultation with ward doctor)

Cardiac

Patient has no ongoing infarction

Patient has no ischemic changes on the electrocardiogram

Patient has no arrhythmia with hemodynamic consequences

Fluid balance

Patient has $\leq$ 100 mL chest tube drainage per hour

Patient has diuresis > 0.5 mL/kg^{-1}/hr^{-1}

Neurologic

Patient has no signs/symptoms of neurologic complications

Hemodynamic and supporting therapy

Patient doesn't need inotropic support or sodium niproprusside (nitroglycerin and dopamine (2μg/kg^{-1}/min^{-1}) are allowed)

Patient does not have an intraaortic balloon pump

Patient with pulmonary artery catheter: the last measured cardiac index of $\geq$ 2 L/min^{-1} /m^{-2} mixed venous saturation (55-60%)

Figure: Flow chart of the treatment for the two study groups
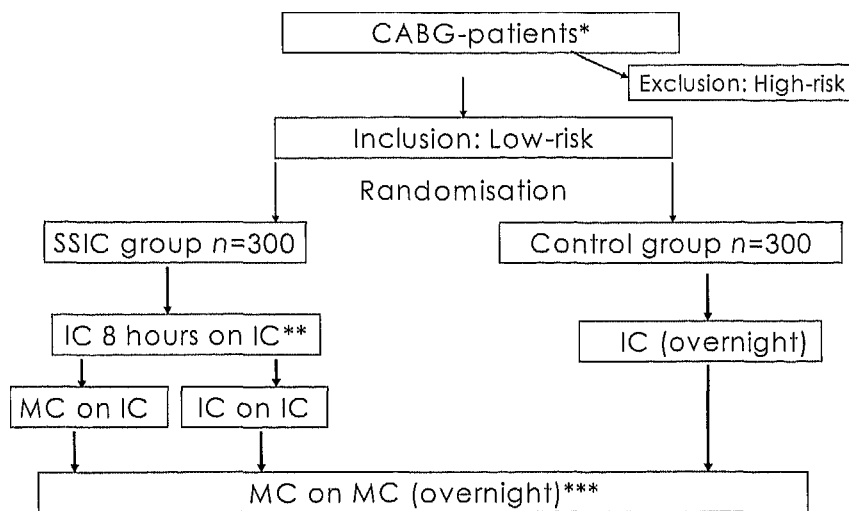
*coronary artery bypass grafting; **intensive care; ***medium care

59

Quality of life

Quality of life (QoL) was measured 1 day before and 1 month postoperatively using EQ-5D (EuroQoL-Group, 1990). The EQ-5D includes five 3-level dimensions: morbidity, self care, usual activities, pain, and anxiety/depression. The patient's answers to the five EQ-5D-items were used to express a single utility value, ranging from 0 to 1. In the absence of a set of Dutch population-based utility weights, we used the most robust valuation set (Dolan, 1997).

The utility values were used to calculate the delta quality adjusted life month (Δ QALM) for follow-up. This is calculated by subtracting the utility measured at the end of follow-up from the baseline utility score divided by two. By dividing the delta by two, the average gain or loss in utility is estimated over the course of the month, assuming that the change was linear over that time period.

Statistical analyses

All analyses were conducted according to the intention-to-treat principle. A P-value <.05 was considered to be statistically significant. All data were analyzed with SPSS, version 10.1 (SPSS, Chicago, IL). Missing values were imputed by the mean group scores.

The differences in clinical outcome measures between the two groups were analyzed by means of Chi-square or Mann Whitney test, depending on level and type of variables.

As cost and QoL data are not normally distributed and skewed, a nonparametric bootstrap sampling method with 1,000 replications was used to assess the statistical significance of cost or QoL differences between the SSIC and the control group (Briggs, 1997). The mean costs (total hospital costs) and effects (Δ QALM) of the control group were subtracted from the mean costs and effects of the SSIC group, resulting in an incremental cost-effectiveness ratio (ICER). The non parametric bootstrap sampling method was used to estimate the probability that the esti-

mated ICERs dominate. Figure 3 shows the cost-effectiveness plane, which simultaneously represents the difference in mean costs (on the Y-axis) and differences in Δ QALM (on the X-axis) between the SSIC and the control group.

Sensitivity analysis

In six 1-way analyses, the unit costs of hospital inpatient days were varied with a minimum and maximum cost estimate. The estimates of IC stay were based on the results of a study performed by Oostenbrink (Oosten-brink, 2003). The same proportion of the costs is used to determine the range for MC and LC costs (table 6). In an additional sensitivity analysis, the shadow price of MC stay for SSIC patients who received their first hours of MC treatment in the IC department, was replaced by the IC price.

The worst-case scenario for the SSIC group was shown in a multivariate analysis. The lowest cost estimate for IC (€36.18), the highest estimate for MC (€16.74) and the lowest for LC (€11.83) were used to calculate the costs of inpatient stay.

Results

Study population

Between February 2001 and March 2003, 1062 patients were admitted to the University Hospital Maastricht for CABG surgery. Of these, 380 (35.1%) patients were excluded and 181 were identified as high-risk patients. A total of 199 patients were excluded for logistical reasons and research-related factors (figure 1). A total of 702 of the patients (66.1%) were eligible for inclusion and were asked to give informed consent. Of these eligible patients, 102 patients (14.6%) refused to participate in this study (figure 1). A total of 600 patients (55.5%) gave informed consent. Three of

the patients included initially were excluded from the study before treatment allocation was revealed (figure 2). The preoperative and intra-operative patient characteristics are described in table 2. The study population consisted of a group of CABG patients with a mean age of 62 years; 80% of the patients were men. All measured data - demographic, previous cardiac, medical, and operative history- were comparable between the two groups.

Table 2: Comparison of demographic data, previous medical and cardiac history and operative data between control and SSIC patients

	Control			SSIC		
	N	n	%	N	n	%
Demographic data						
Age (Years)[a]	297	62.6	9.42	296	62.7	9.30
Body Mass Index (Kg/M²)[a]	276	27.4	3.92	274	28.1	3.70
Male (sex)	299	240	80.3	298	236	79.2
Previous cardiac and medical history data						
Ventricle ejection fraction	276			281		
> 35 %		174	58.2		180	60.4
35 - 40 %		37	12.4		39	13.1
40 - 50%		65	21.7		62	20.8
NYHA classification angina	278			281		
I		9	3.0		7	2.3
II		58	19.4		46	15.4
III		129	43.1		143	48.0
IV		82	27.4		85	28.5
Hypertension	299	132	44.1	298	130	43.6
Hypercholesterolemia	299	190	63.5	298	194	65.1
Smoking	299	101	33.8	298	98	32.9
Family history	299	188	62.9	298	174	58.4
Diabetes mellitus	299	53	17.7	298	52	17.4
Operative data						
Number of grafts[a]	291	3.2	1.2	296	3.1	1.2
Cardiopulmonary Bypass time in mins[a]	299	78.4	32.9	298	77.9	32.5
Aorta cross-clamp time in mins[a]	299	50.4	23.3	298	48.6	24.5
Without aorta cross-clamping	299	66	22.1	298	56	18.8
Re-do CABG	299	18	6.0	298	16	5.4

[a] In Mean (standard deviation)

Figure 1: Flow chart of patient inclusion

1062 (100%) patients were admitted to the University Hospital for CABG surgery in the period from March 2001 until February 2003

360 (33.9%) patients were excluded

181 (50.3%) risk stratification

702 (66.1%) patients fulfilled the inclusion criteria

102 (14.5%) refused

Reasons for non-informed consent.

Categories	n	%
Patient related	39	38.2
Was anxious, emotional or nervous	18	
Was too tired, ill	5	
Thought to be too old for study	2	
Has gone through too much during last period	3	
Has too much to think of	11	
Previous experiences	12	11.8
Previous surgery	6	
Previous research	5	
Displeasure about waiting list	1	
Research related	27	26.5
Family didn't want the patient to participate in the study	9	
Assurance Appendix	5	
Not willing to fill in the questionnaires	13	
No specific reason	24	
Total	102	100

Reasons and number of excluded patients based on risk- stratification criteria.

Categories	n	
Older than 78 years	77	42.5
Having intra aortic balloon pump	50	27.6
Ejection fraction of less than 30 %	36	19.8
Stage 3 obesity (BMI > 40kg/m)	6	3.3
Haemodialysis	6	3.3
Recent myocardial infarction	3	1.7
Pulmonary hypertension	2	1.1
Need for inotro ic thera	1	0.6
Recent CVA	0	0
Cardiac shock	0	0
Ongoing infarction	0	0
Total	181	100

179 (47.7%) other criteria

Number of excluded patients for logistical reasons and research related factors.

Categories	n	%
Participation in other research project	62	31.2
Late surgery schedule change or urgent surgery	59	29.6
Not able to speak / read / Dutch	22	11.1
Not able to give informed consent	15	7.5
Second CABG surgery in study period	7	3.5
Other (e.g. missed patient)	14	7.0
Total	179	100

600 (85.5%) informed consent

Figure 2: Trial Profile

Randomized (N=600)

Allocation

Allocated to control group (N=300)

Received allocated control intervention (N=297)

Did not receive allocated control intervention(n=3)

Reasons

- Exclusion (n=1)
1 patient appeared to have < 30% left ventricle ejection fraction intra operatively

- Not treated according to protocol (n=2)*
  treated as SSIC patient (n=2)

*100 of the 300 control group patients checked

Allocated to SSIC group (N=300)

Received allocated SSIC intervention (N=283)

Did not receive allocated SSIC intervention (n=19)

Reasons

- Exclusion (n=2)
1 patient underwent CABG and additional surgery
1 patient died during surgery

- Not treated according to protocol (n=17)

9 patients were treated as control patient
2 patients received additional blood gas measures after 8 hours
3 patients had their clinical situation wrongly interpreted
1 patient was exlubated after 8 hours without reason
2 patients reason unknown

Completeness of data

Missing data

Maximum number of missing medical data
Preoperative (n=23) (table 2)
Postoperative (n=3) (table 3-4)

Cost data
Clinical postoperative (n=0) (table 5)
Outpatient postoperative (n=31)

Quality of Life data
Preoperative (n=1)
Postoperative (n=28)

Missing data

Maximum number of missing medical data
Preoperative (n=24) (table 2)
Postoperative (n=0) (table 3-4)

Cost data
Clinical postoperative (n=0) (table 5)
Outpatient postoperative (n=19)

Quality of Life data
Preoperative (n=1)
Postoperative (n=16)

Analysis

Analyzed (N=299)

Excluded from analysis (n=1)

Reason
1 patient appeared to have < 30% left ventricle ejection fraction perioperatively

Analysed (N=298)

Excluded from analysis (n=2)

Reasons
1 patient underwent CABG and additional surgery
1 patient died during surgery

Missing data and protocol deviations

Mean imputation was used because the number of missing data were very limited (details provided in figure 2). As the analysis before and after mean imputation showed the same results, the method of imputation of the missing data had no influence on the outcomes.

The protocol deviations in both groups were minimal (SSIC group, $n=17$ and, control group, $n=2$ (figure 2). Data were obtained from all SSIC patients and a random sample of control patients ($n=100$) on actual MC and IC treatment by evaluating bed lists. This was performed by checking the registration of actual extubation time, the registration of pulmonary artery occlusion pressures and the number of laboratory tests performed.

Clinical effectiveness

The primary study outcome revealed that IC readmission was lower in both groups than the expected 5%. IC readmission was 1.34% ($n=4$) and 2.68% ($n=8$) in the control group and the SSIC group, respectively. The difference of 1.13% was not significantly different (P-value = .241 95% confidence interval (95% CI): -.9% to 2.9%). In other words, if patients were treated according to the SSIC protocol, a maximum of 3 out of 100 patients needed IC readmission.

The reasons for the four readmissions in the control group were infection, respiratory insufficiency, reexploration for bleeding, and lack of beds in MC. In the SSIC group, readmission was necessary eight times for the following reasons: cardiologic, hemodynamic (three times), re-exploration for bleeding (twice), respiratory insufficiency, and order of physician.

No statistically significant differences were found between the two groups with respect to postoperative morbidity and 30 day-mortality (table 3). The patients in the SSIC group stayed in the IC unit for significantly fewer hours compared to control group patients (P-value = .000, 95% CI: 13.1 to 9.5) (table 4). In the SSIC group, 161 patients (55%) were trans-

ferred from IC to MC within 8 hours. The different reasons for not transfer-ring the 137 patients in the SSIC group were: surgical (n=18), anesthetic (n=25), respiratory (n=19), hemodynamic (n=18), patient-dependent (n=27), cardiologic (n=30), protocol (n=30), and other (n=8) (patients could be grouped in more then one category). The mean postoperative total hospital stay was comparable between the two groups (P-value of .807, 95% CI: 1.2 to -0.4).

Table 3: Comparison of postoperative morbidity and mortality between control group and SSIC patients, data presented in number (%)

	Control			SSIC			Relative Risk	95% CI [a]	P-value [b]
	N	n	%	N	n	%			
Postoperative morbidity									
Re-exploration for bleeding	299	14	4.7	298	11	3.7	0.78	.36 to 1.71	.546
Re-exploration for occlusion graft	299	3	1.0	298	1	0.3	0.33	.03 to 3.19	.317
Tamponade	299	1	0.3	298	2	0.7	2.00	.03 to 5.26	.561
Postoperative MI [c]	299	16	5.4	298	17	5.7	1.06	.18 to 21.94	.850
Atrial arrhythmias	299	84	28.1	298	85	28.5	1.01	.52 to 1.96	.979
Pleura / Pericard effusion	296	3	1.0	298	3	1.0	1.00	.77 to 1.28	.997
Sternum superficial wound infect. [d]	298	4	1.3	298	3	1.0	0.75	16 to 3.32	.479
Sternum deep-wound infect.	298	1	0.3	298	0	0.0	0.99	[e]	.318
Other infections	298	19	6.4	298	19	6.4	1.18	.61 to 2.25	.997
Stroke permanent	299	2	0.7	298	2	0.7	1.00	.14 to 7.07	.997
Increased creatinine	299	7	2.3	298	7	2.3	0.99	.35 to 2.8	.995
Dialysis	297	2	0.7	298	0	0.0	0.05	[e]	.157
30 day-mortality	299	3	1.0	298	1	0.3	0.33	.03 to 3.20	.317

[a] 95% CI= 95% confidence interval, [b] Chi-square test, [c] MI= myocardial infarction, [d] infect.= infection, [e] not appli-cable

Quality of life

The 1-month utility score was equal in both groups (mean (standard error), 0.71 (0.01)). However, the mean and standard error for baseline util-ity score were 0.71 (0.01) and 0.66 (0.01) for the control group and SSIC group, respectively. This was significantly different between the two groups (mean 95% CI: -0.05 (-0.01 to -0.09). Thus, the Δ QALM difference between both groups was small (.0238) but significantly different (95% CI: .0012 to .0464) between the two groups. Assuming that Δ QALMs are

equal on the other days of the month, patients in the SSIC group would be in full health for 1.4 days compared to 0 days in the control group. The Δ QALM was .0015 ± .1388 and .0253 ± .1424 for the control group and SSIC group respectively, indicating that the control group improved less in overall quality of life compared with the SSIC group.

Table 4: Comparison of extubation and hospital stay data between control and SSIC patients

	Control					SSIC						
	N	Mean	Median	2.5[a]	97.5[b]	N	Mean	Median	2.5	97.5	95% CI	P-value[c]
Extubation time in (hrs)	299	9.4	7.0	.0	39.9	298	7.7	6.5	2.0	19.8	3.5 to .1	.514
Total IC stay in UH [d] (hrs)	299	31.0	22.3	16.5	117.0	298	19.7	8.0	5.0	89.6	13.1 to 9.5	.000
Total MC stay in UH (hrs)	299	24.8	20.5	.0	66.8	298	36.9	32.7	16.0	84.3	16.5 to 7.7	.000
Total LC stay in UH (days)	299	4.7	4.0	1.5	15.5	298	4.5	4.0	1.0	11.0	4.6 to 4.2	.984
LC Stay in DH [e] (days)	299	1.4	0.0	.0	13.5	298	1.2	0	.0	11.5	.7 to .3	.598
Total hospital stay (days)	299	8.5	6.7	4.6	24.6	298	8.1	6.7	4.5	26.7	1.2 to .4	.807

[a] 2.5 th Percentile, [b] 97.5 th Percentile, [c] Mann-Whitney U test, [d] University Hospital (n=1), [e] District Hospital (n=5)

Costs

The mean total costs were €5,441 in the control group and €4,625 in the SSIC group. The mean difference was €-816. The costs were significantly lower (95% CI: €-1,581 to €-174) in the intervention group, due to lower costs for clinical procedures and for IC stay at the university hospital (table 5).

However, costs of the MC stay were higher in the SSIC group, indicating that part of the cost savings from the IC stay were compensated by higher costs of the MC stay. However, overall, it turned out that SSIC would lead to substantial cost savings per patient for the hospital. The costs of LC, outpatient procedures, and costs in district hospitals were comparable in both groups.

Table 5: Summary of the total clinical costs of the SSIC patients and the control patients

Cost Category	Control Mean	% of total costs	SSIC Mean	% of total costs	Summary Information bootstrapped cost-differences Mean	Median	2.5 [a]	97.5 [b]
Procedures								
Total Clinical procedures CTC IC [c]	1,073.89	19.74	845.74	18.29	-260	235	-507	-50
Total Clinical procedures MC and LC	540.16	9.93	464.60	10.04	-74	-73	-177	26
Total Clinical procedures other departments	306.55	5.63	185.94	4.02	-118	-104	-432	126
Total outpatient procedures UH [d] and DH [e]	75.95	1.40	73.93	1.60	-2	-2	-20	14
Total procedures [f]	1,996.55	36.70	1,540.93	33.32	-454	-444	-945	-21
Hospital admission days								
Total IC stay UH	1,397.07	25.68	890.45	19.25	-506	-509	-703	-308
Total MC stay UH	518.77	9.54	771.74	16.69	250	246	167	349
Total LC stay UH	1,177.28	21.64	1,128.56	24.40	-50	-48	-178	72
Total Care stay UH [f]	3,093.12	56.85	2,790.75	60.34	-294	-294	-561	-44
Total LC stay DH	350.98	6.45	294.42	6.37	-60	-63	-180	75
Total Inpatient Stay [f]	3,444.10	63.30	3,085.18	66.70	-358	-357	-671	-56
Total costs [f]	5,440.64	100.00	4,625.29	100.00	-823	-816	-1,581	-174

[a] 2.5 [th] Percentile, [b] 97.5 [th] Percentile, [c] These figures also include medium care procedures for SSIC patients who were discharged from IC to MC within 8 hrs, [d] University Hospital (n=1), [e] District Hospital (n=5), [f] Costs do not add up correctly due to rounding errors

Cost-effectiveness

Figure 3 shows the bootstrap results of the ICERs (cost/ Δ QALM). The ICERs were situated in the Southeast quadrant, indicating that the SSIC group is dominant over the control group. Ninety-eight percent of the bootstrapped ICERs showed that SSIC patients' QoL improved more and their costs were lower compared to control patients. Only 0.016% of the bootstrapped ICERs was situated in the Southwest and 0.004% in the Northeast quadrant.

Figure 3: Cost-effectiveness plane showing the results of difference between cost and effect between both treatments arms in 1,000 bootstrap replications
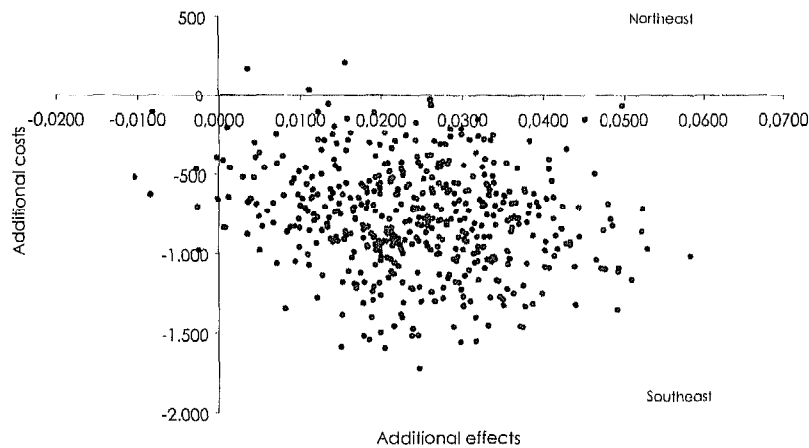
Sensitivity analyses

The results of the one-way sensitivity analyses on the minimum and maximum estimate of the unit cost of inpatient hospital days showed that these results were very robust. At least 97% of the bootstrapped ICERs lay in the same quadrant as in the base case analyses. Zero percent of the bootstrapped ICERs are situated in the Northwest quadrant, where the SSIC group would be inferior (less effective and more costly). As explained previously, the hypothetical cost for first hours of MC stay were calculated for SSIC patients. The unit price for these hours of MC treatment was replaced by the IC treatment unit price. This sensitivity analysis showed that calculation of the actual costs did not influence the base case findings. Even in a worst-case scenario, when the costs of IC and LC were set at the lowest cost estimate (SSIC patients stay less time in IC and LC compared with the control group) and the MC costs were set at the highest cost estimate (SSIC patients stay longer in MC compared with the control group), the results of the primary analyses still held up (table 6).

Table 6: Results of the different univariate and one multivariate sensitivity analyses (percentage of bootstrap ICERs).

	Costs	NE [a]	NW [b]	SW [c]	SE [d]
Base case		0	0	2	98
Hospital admission days: univariate					
LC Costs (minimum)	8.30 per hr	0	0	2	98
LC Costs (maximum)	11.83 per hr	0	0	2	98
MC Costs (minimum)	16.74 per hr	0	0	2	98
MC Costs (maximum)	23.85 per hr	1	0	1	98
IC Costs (minimum)	36.18 per hr	1	0	2	97
IC Costs (maximum)	51.51 per hr	0	0	3	97
Replacing MC unit cost by IC unit cost in		5	0	1	94
SSIC patients who received MC treatment on IC					
Hospital admission days: multivariate	**worst-case scenario**				
IC Costs (minimum)	36.18 per hr				
MC Costs (maximum)	23.85 per hr	1	0	2	97
LC Costs (minimum)	8.30 per hr				

[a] NE= Northeast quadrant: SSIC more effective and more costly compared to control group, [b] NW= Northwest quadrant: SSIC less effective and more costly compared to control group (inferior), [c] SW= Southwest quadrant: SSIC less effective and less costly compared to control group, [d] SE= Southeast quadrant: SSIC more effective and less costly compared to control group (dominant)

Discussion

Summary of the results

SSIC discharge for low-risk CABG patients from IC to MC after 8 hours is safe and feasible. There is no increase in IC readmissions compared to postoperative overnight IC stay. An influence of SSIC treatment on the total hospital stay, 30 day-mortality and postoperative morbidity was not observed. Furthermore, SSIC is a cost-effective approach as it lowers the total hospital costs and has a positive effect on QoL. However, it is questionable whether this QoL improvement is clinically relevant.

Clinical effectiveness

Several trials have examined strategies to reduce the duration of mechanical ventilation in patients after CABG surgery. Methodologic problems occurred in these trials: for instance a lack of information concerning concealment, a lack of intention-to-treat analyses, and flaws in power calculations, dealing with missing data and loss to follow up

(Berry, 1998; Michalopoulos, 1998; Quasha, 1980; Reyes, 1997; Silbert, 1998). As stated before by Meade et al, a trial of thousands of patients - or at least many hundreds of patients with a combined endpoint- would be required to study the safety of early extubation due to low adverse event rates (Meade, 2001). Our study, with safety as its primary endpoint, defined as the probability of being readmitted to IC, has tried to over- come this problem. Our trial and three other studies (Arom, 1995; Michalopoulos, 1998; Reyes, 1997) reporting the number of IC readmis- sions found no significant difference between the two study groups. Al- though the true IC admission rate of 1.13% was lower than the expected 5%, the results were clinically acceptable. In the power calculation of our study, it was presumed that a maximum of 8% of the patients treated in SSIC needed IC readmission. The results of this trial showed that a maxi- mum of 3% of these patients had to be readmitted on the IC. The most plausible reason in our opinion for the discrepancy between real and expected readmission rate is the fact that the power calculation was performed in 1999 and the study started two years later. In addition to this, the power calculation was based on retrospective data, whereas the trial was based on prospective data.

Successful discharge of the majority of patients in the SSIC group to MC resulted in a decrease of 11 hours IC stay. This is more than the weighted mean difference of -7.02 (95% CI: -7.42 to -6.61) reported in an recently published review (Hawkes, 2003). In contrast to two previous studies (Michalopoulos, 1998; Silbert, 1998) we did not notice a decrease in the number of hospital admission days. Our design meant that we deliber- ately focused on IC discharge only. After IC discharge, both SSIC and control patients received routine conventional care, including one over- night stay in MC. If the MC stay had been restricted to 24 hours instead of an overnight stay, it might have been possible to decrease MC treat- ment to only 24 hours. A reduction of 1 day's hospital admission stay is feasible, as demonstrated by the pooled results of five studies (relative

71

ratio: -1.06, 95% CI: -1.32 to -0.08) (Meade, 2001). The ability to transfer a patient may be affected by various factors unrelated to the patient's clinical situation. These factors include the availability of beds and nursing staff and, perhaps more importantly, the reluctance of medical personal to transfer patients in the middle of the night (Velasco, 1995).

Cost-effectiveness

Besides evaluating medical effects, it is also important to evaluate whether SSIC influences the QoL. To our knowledge, to date, no study has investigated the differences in overall QoL for low-risk CABG patients after fast-track treatment. The SSIC group improved significantly more in QoL compared with the control group. However, the effect is very small, indicating that its clinical relevance is doubtful.

The conventional effect measure in cost-effectiveness analyses is the quality adjusted life year (QALY). We chose to use QALM instead of QALY, as the follow-up in this study was 1 month. Besides this, it is difficult to interpret a 1-month follow-up QALY (maximum QALY in one month is 0.083). The area under the curve is estimated in such a calculation, in other words the sum of the utility score for both measurement moments (at baseline and 1 month) divided by 2 and multiplied by 1/12 (Gold, 1996). We used a Δ QALM (Guthrie, 1999), because baseline differences in the utility score for the QALM calculation were not suitable. Only the increment of baseline utility score and utility score at 1 month divided by 2 was used for the Δ QALM calculation.

Gold et al. recommended employing cost-effectiveness analyses from a societal perspective (Gold, 1996). However, we chose to perform the study from a hospital perspective as we assumed that, given the relatively short time horizon, there would be no difference in costs outside the hospital between both groups. Furthermore, as the recovery of both groups was comparable, other medical costs like costs of primary care were not expected to be different between both groups.

It is interesting to see that the driving force of fast-track recovery is cost containment (Higgins, 1992), but as stated by Velasco, current literature has produced inconsistent evidence that only early extubation reduces the costs of cardiac surgery (Velasco, 1995), not transferring the patients to MC. Grade-1 evidence has only been produced by one randomised clinical trial (Cheng, 1996b). Other studies (Arom, 1995; Engelman, 1994; Lee, 1996; Westaby, 1993) that claimed to study cost reductions or even cost-effectiveness gave incomplete or no information concerning cost calculation, uncertainty of outcomes and unit prices, which are essential for interpreting the results of an economic evaluation. Furthermore, most cost studies were not randomised clinical trials, and thus, evidence of causal relations cannot be supported.

The decrease of total hospital costs in our study of the SSIC group was due to fewer laboratory blood tests being performed in IC and by a shorter stay in the high-cost IC setting. A reduction in laboratory blood tests was also reported by others (Cheng, 1996b; Foster, 1984; Lee, 1996). As in our study, the early transfer to a lower level of care resulted in cost-savings (Cheng, 1996b), as 1 hr of IC stay is twice as expensive as 1 hr of MC stay (IC, €45; MC, €21).

Limitations of the study

The SSIC intervention was restricted to the first 24 hours of postoperative care (8 hrs IC treatment, IC discharge and 16 hrs MC treatment). The conventional overnight MC treatment was maintained for both patient groups in accordance with local protocol. From a financial point of view, a reduction of up to 24 hours of MC treatment would have resulted in additional cost-savings, but for purity of study findings, it was necessary to separate these factors.

We chose only one decision moment to transfer SSIC patients from IC to MC to minimize protocol deviations. This moment was after 8 hours of IC

73

treatment. It might have been more effective if the decision to transfer the SSIC patients had also been taken 10 or 12 hours after IC admission.

It is not possible to extrapolate our findings to more high-risk CABG patients.

We used data obtained from an English population for the valuation of the EQ-5D items. One may argue that these valuations may differ from the Dutch population; however, no Dutch value set is currently available, and we used the one that was recommended as being the most robust (Books & de Charro, 2003).

The general applicability of the results may be limited because not all patients who fulfilled the inclusion criteria were included. Patients who refused to participate in the study were on average older, a higher proportion of them were women, and generally had a lower ventricle ejection fraction. Furthermore, the ability to generalize may be limited due to the fact that this study was a single-center study. Other hospitals may use other treatment protocols and have different patient/nurse ratios in IC and MC, which may influence the actual cost savings.

Conclusions

This is the first randomised clinical trial that shows that short-stay in IC is safe for low-risk CABG patients. The IC readmission rate was very low in SSIC patients, and postoperative complications and 30 day-mortality in SSIC patients were comparable with patients who stayed overnight on the IC. SSIC is also a cost-effective approach. It lowers the total hospital costs, and it has a positive effect on QoL. SSIC can be a considered as an alternative for conventional postoperative IC treatment for low-risk CABG patients.

CHAPTER

4

Health-related quality of life after fast-track treatment: results from a randomised controlled clinical equivalence trial

Ghislaine van Mastrigt
Manuela Joore
Fred Nieman
Hans Severens
Jos Maessen

77

Abstract

Purpose: This randomised clinical equivalence trial was designed to evaluate health-related quality of life (QoL) after a fast-track treatment for low-risk coronary artery bypass patients (CABG).

Methods: Four hundred and ten CABG patients were randomly assigned to undergo either short-stay intensive care treatment (SSIC, 8 hrs of intensive care stay) or control treatment (care as usual, overnight intensive care stay). QoL was measured until one year postoperatively by means of the Multidimensional Index of Life Quality (MILQ), the EQ-5D, the Beck Depression Inventory and the State-Trait Anxiety Inventory.

Results: At one month after surgery no statistically significant differences in the summary scores were measured (MILQ-score P-value=.508, Overall MILQ-index P-value=.543, EQ-5D VAS P-value=.593). The domains, physical and social functioning of the MILQ, were significantly higher at one month postoperatively in the SSIC group compared with care as usual (respectively, P-value= .049; 95% CI: 0.01 to 2.50 and P-value = .014; 95% CI: 0.24 to 2.06). However, these differences were no longer observed at one-year follow-up.

Conclusions: According to our definition of clinical equivalence, the QoL of SSIC patients is similar to patients receiving care as usual. As safety and cost-effectiveness of this intervention was demonstrated in a previously reported analysis, SSIC can be considered as an adequate fast-track intensive care treatment option for low-risk CABG patients.

Introduction

Rapid technological advancements changed the practice of intensive care medicine in the last few decades. These improvements facilitated the introduction of fast-track protocols for the treatment of cardiac surgical patients, which are very popular as they contribute to a more efficient utilization of the existing care facilities. Various reviews, based on randomised controlled studies, focused on the evaluation of the clinical effectiveness of different types of fast-track treatments (Hawkes, 2003; Meade, 2001; Myles, 2003). However, for the evaluation of these intensive care interventions it is essential to also assess its impact on QoL(Suter, 1994). These patient-reported outcomes contain valuable information for intensive care physicians, other healthcare providers, policymakers, patients and their relatives.

Several 'fast-track' trials incorporated QoL as an outcome measure in their evaluations, but mostly only one specific domain with a short follow-up time has been assessed (Bowler, 2002; Cheng, 1996a; Cheng, 1996b; Cheng, 2001; Cheng, 2003; Engoren, 2001; Fillinger, 2002; Howie, 2001; Kadoi, 2003; Mollhoff, 2001; Royse, 2003; Shroff, 1997; Zarate, 2000). These limited assessments are helpful in describing the domains of patient outcome in a specific area of QoL. However, they are not designed to evaluate either the overall health state (Heyland, 1998) or the specific domains of QoL which may contain important aspects of health. For a comprehensive evaluation of QoL it is important to asses both overall or generic QoL and disease specific QoL (Mayou & Bryant, 1993). We chose to use the EQ-5D as a generic measure, because it's brief, simple and easy to administer. The reliability and validity of the EQ-5D has been studied successfully for cardiac patients (Nowels, 2005; Schweikert, 2006) and critical care patients (Badia, 2001). In a previous paper we also reported generic QoL at one month after surgery as an outcome measure in an economic evaluation of the short-stay intensive care treatment (van

79

Mastrigt, 2006). The patient's answers to the five EQ-5D items were used to calculate an utility value (a score of 0 to 1), based on population weights (Dolan, 1997). In the current study this outcome was not reported because the aim of this paper was to report data from a patient point of view and not from a societal perspective.

As disease specific measure, the Multidimensional Index of Life Quality was chosen. Although this questionnaire is not frequently used its reliability and validity has been tested by Avis et al. for cardio vascular patients (Avis, 1996). In addition, we assessed the impact of the fast-track treatment on anxiety and depression, as these two psychological aspects of QoL may be affected after cardiac surgery (Duits, 1997).

As several studies after intensive care admission have demonstrated the importance of long-term follow-up (Badia, 2001; Cuthbertson, 2005; Hurel, 1997; Konopad, 1995), we assessed QoL at one year after surgery.

The focus of this randomised clinical equivalence trial is to evaluate the impact of a fast-track protocol, i.e. early discharge -within 8 hours- of post-surgical coronary artery bypass patients from intensive to medium care on disease specific, generic and domain specific (anxious and depressive feelings) QoL up to one year postoperatively.

Materials and methods

Study design and hypotheses

This single blinded randomised clinical equivalence trial was designed to evaluate the QoL of low-risk CABG patients at one year postoperatively who were randomised for either short-stay intensive care (SSIC < 8 hrs) or overnight intensive care stay (control group patients).

To investigate this, the following hypothesis was formulated to evaluate the primary outcome: 1) up till one year postoperatively the disease specific QoL (measured by means of MILQ) will be similar in both groups.

Two additional hypotheses were formulated to assess the secondary outcome measures: 1) in the first year after surgery both generic (measured by means of EQ-5D) and domain specific QoL (measured by means of Beck Depression Inventory and State-Trait Anxiety Inventory) will be comparable in both groups, 2) at one month postoperatively disease specific, generic and domain specific QoL will also be comparable in both groups.

For details on patient recruitment, eligibility criteria, treatment, randomization and blinding, see chapter 3.

Procedures quality of life measurement

After obtaining informed consent, QoL questionnaires were self-administered by the patients the day before surgery at the hospital, one month and one year postoperatively at home. If a patient was unable to complete the questionnaire individually, the research nurse performed an interviewed administration. The application of the four QoL instruments took almost 20 minutes and was done in a fixed order: first the EQ-5D, secondly the MILQ, followed by State-Trait Anxiety Inventory and Beck Depression Inventory.

Patients who did not respond to the follow-up questionnaires received a written reminder. If they still did not respond, they were contacted by telephone to find out the reasons for non-response.

Quality of life questionnaires

The Multidimensional Index of Life Quality (MILQ), a QoL measure specially developed for cardiovascular disease, was used to measure disease specific QoL (Avis, 1996). In the first item (which is referred to as the 'MILQ-score') the respondent is asked to rate his/her QoL on a seven

81

point-Likert scale (from 1= worst to 7= best possible QoL score). In the other 35 items the respondent's degree of satisfaction with various aspects of QoL is scored on a seven point-Likert scale (from 1= very dissatisfied to 7= very satisfied). These items cover the following nine domains: mental health, physical health, physical functioning, cognitive functioning, social functioning, intimacy, financial status, relationship with healthcare professionals and productivity. As each domain contains 4 items, the total theoretical domain score ranges from 4 to 28 (except for the domain of productivity which contains 3 items, the total score ranges from 3 to 21). The Dutch translation of the MILQ has been used to assess the QoL of CABG patients (Falger, 2000). An unweighted index was calculated by adding up overall nine-domain scores, theoretically ranging from 35 to 245 (which is referred to as the 'overall MILQ-index').

Furthermore, generic QoL was measured by the EQ-5D (EuroQoL-Group, 1990). The EQ-5D consists of two components: an EQ-5D descriptive system and an EQ-5D Visual Analogue Scale (EQ-5D VAS). In the descriptive system the respondent is asked to rate his or her health by checking one of three levels of severity - 'no problems' (coded as 1), 'some or moderate problems' (coded as 2), 'severe problems or unable to perform' (coded as 3)- in each of the following five EQ-5D dimensions; mobility, self care and pain/discomfort, usual activities and anxiety/ depression. For the EQ VAS, participants draw a line from a box to the point on the thermometer-like scale corresponding to their health state (with a range from 0 'the worst imaginable health state' to 100 'the best imaginable health state').

Depression was evaluated by means of the Beck Depression Inventory (BDI) (Beck & Steer, 1987). The State-Trait Anxiety Inventory (STAI-S) was used to measure temporal and transient aspects of anxiety (van der Ploeg, 1986). Both STAI-S (20 items) and BDI (21 items) contain items expressed on a four point-Likert scales. Both BDI and STAI-S have been used in studies with CABG patients (Boudrez & De Backer, 2001; Eisenberg,

2001; Langeluddecke, 1989), and are considered as valid and reliable instruments (Beck & Steer, 1987; van der Ploeg, 1986).

One year quality of life equivalence and power analysis

Equivalence in QoL was defined a priori. It was expected that the disease specific QoL (MILQ) measure would be the most sensitive to detect differences between the two groups. The overall MILQ-index varies between 1 (worst of health) and 7 (best of health). Equivalence on the overall MILQ-index at one year was assumed if the SSIC group and the control group would not differ within a tolerance range of 5%, i.e. within 0.4 on a 7 points scale. It was assumed that the standard deviation of this difference would be 1.15, so with a two-sided test, alpha of 0.05, a power of 90% and an expected drop-out percentage of 10% the number of patients needed amounted to 194 for each study group. For the MILQ-score and EQ-5D VAS the power calculations resulted in a lower number of patients needed to be included in each trial arm. In the line with the assumptions of the power analysis, a difference of approximately one third of the standard deviations of the one-year QoL differences with baseline score between SSIC group and control group was defined as a clinical meaningful difference.

Statistical analysis

The reasons for non-response were categorized as follows: 'Informative censoring' (if the reason was non-random e.g. related to the health status) (Staquet, 1996), 'non-informative censoring' (if the reason was random) (Staquet, 1996) and 'unclear' (if the reason was unknown).
All analyses were conducted following the intention-to-treat principle. Missing values were imputed by the separate mean group scores of non-missing patients for the SSIC and the control group. This way, chances of the alternative hypothesis that possible differences in QoL are statistically significant were enhanced to the detriment of our equivalence hypothe-

sis of no clinically relevant differences. QoL scores at baseline were re-ported in means and standard deviations and at follow-up as mean change scores (baseline minus one month or one year minus one month) with a 95% confidence interval for both groups. The Kolmogorov-Smirnov test was used to test for normality of the differences between postopera-tive one month and one year scores on QoL scales at one hand and preoperative scores at the other. Student t-tests were performed to ex-amine the mean differences in QoL change between both treatment groups at one month after surgery.

Repeated measures ANCOVA analysis using the preoperative scores as a covariate was performed to test equivalence of QoL scores at one year. Interaction effects between preoperative QoL and treatment modality group on one-year differences in QoL were also tested for statistical sig-nificance. Next, in regression analyses with the MILQ-score, the overall MILQ-index and the EQ-VAS as 'dependent' variables the relation with other factors was tested. Selection of the potential predictors in the re-gression models was based on results of previous research (see appen-dix). Regression modeling was done with list wise deletion of missing cases using both forward selection and backward elimination tech-niques. Predictors statistically significant in both techniques were selected in the model, except for treatment modality group, which was always kept within the model, irrespective of its statistical significance. Finally, eventual regression model results are presented in table 4 with unstan-dardized effects, P-values and variance explained (R-square) by each eventual model.

A P-value of less than 0.05 was assumed to be statistically significant. All data were analyzed with SPSS, version 12.0.1 (SPSS, Chicago, IL).

Results

Study population

Between February 2001 and March 2003, out of the 1062 available CABG patients, 702 were eligible for inclusion for the clinical part of the study (figure 1). They were asked to give informed consent. A total of 600 patients gave informed consent and participated in the SSIC-trial. For this QoL-study only the first 410 of the included CABG patients were needed. The study population consisted of a group of low-risk CABG patients with a mean age of 62 years, 80% of them were men (table 1).

The non-response rates were low and appear to be similar for control and SSIC group in all three categories (informative, non-informative and unclear) and at the three measurement moments (table 2). Four patients died (control n=1 and SSIC n=3) and one patient was readmitted to the hospital. These events were unrelated to the IC treatment received. 39 follow-up questionnaires (4.8% of the total) were not returned for reasons that were unclear, however, mortality as reason was ruled out.

Figure 1: Flow chart of patient inclusion

*for reasons see table 2
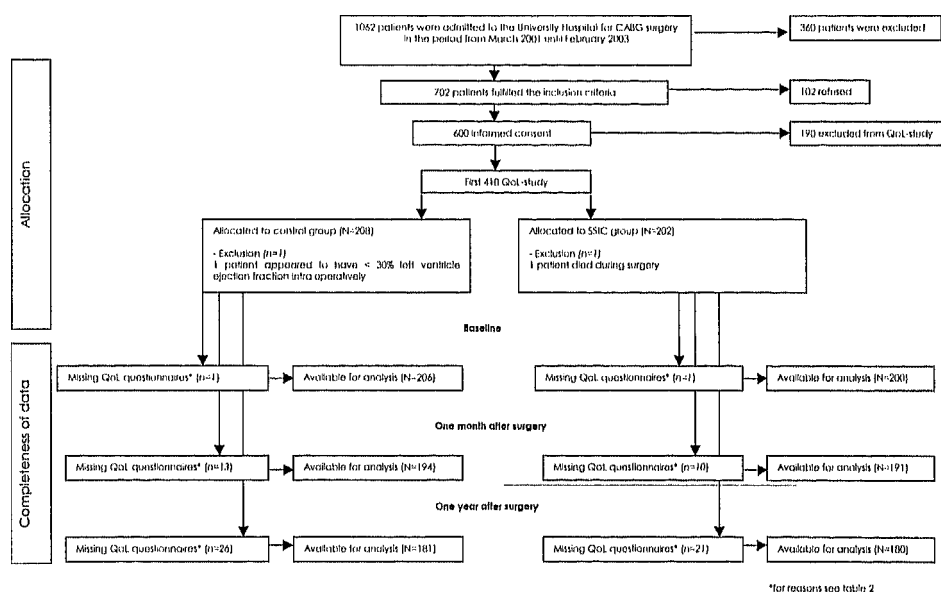
The non-responders were more often females with no partner, had slightly longer surgery times and hospital stay compared to the responders. However, there was no difference in patient characteristics between the non-responders in the control and the non-responders in the SSIC group (data not shown). Baseline QoL outcomes appeared to be not significantly different between SSIC group and control group (P-values not shown).

Disease specific quality of life

At one month after surgery, the change scores (one month postoperative minus baseline) of both the MILQ-score and overall MILQ-index were not significantly different between the two treatment groups (MILQ-score P-value=.270; 95% CI: -0.11 to 0.40, overall MILQ-index P-value=.154; 95% CI: -1.49 to 9.41) (table 3). However, the domains physical and social functioning of the MILQ were significantly better in the SSIC group compared to the control group (respectively, P-value = .049; 95% CI: 0.01 to 2.50 and P-value = .014; 95% CI: 0.24 to 2.06) (table 3).

Results of the repeated measures ANCOVA analysis with the baseline QoL scores as covariates show no statistically significant differences between the SSIC group and the control group in the linear trend in the year after surgery on the MILQ-score (P-value= .508), nor on the overall MILQ-index (P-value = .543) (table 3).

In the regression models the linear trend of the MILQ-score and the linear trend of the overall MILQ-index were used as "dependent" variables. Controlled for potential confounding factors, the type of intensive care treatment received was again no significant predictor of QoL (MILQ-score, P-value= .736; 95% CI: -.115 to .162, nor for the overall MILQ-index P-value = .934; 95% CI: -3.64 to 3.35) (table 4). Apart from this, the baseline scores of the corresponding QoL measures were significant predictors of linear trend of QoL. All tests for first-order interactions and all tests on

the other predictors like gender, age and disease severity (appendix) were not statistically significant (table 4).

Table 1: Comparison of demographic data, previous medical, cardiac history and operative data between control and SSIC patients

	Control			SSIC		
	N	n	%	N	n	%
Demographic data						
Age (Years) [a]	206	62.6	13.7	201	61.8	10.67
Body Mass Index (Kg/M²) [a]	193	27.4	3.84	185	27.8	3.55
Male (sex)	207	169	81.6	201	161	80.1
Civil state	201			198		
Married		150	72.5		158	78.6
Unmarried no partner		40	19.3		31	15.4
Unmarried with partner		14	6.8		9	4.5
Education	195			196		
Primary school not finished		2	1.0		5	2.5
Primary school		53	25.6		40	19.9
Lower education		47	22.7		57	28.4
Intermediate education		63	30.4		64	31.8
High education		30	14.5		30	14.9
Previous cardiac and medical history data						
Ventricle ejection fraction	193			189		
> 35 %		115	55.6		120	59.7
35 - 40 %		30	14.5		27	13.4
40 - 50%		48	23.2		42	20.9
NYHA classification angina [b]	195			191		
I		8	3.9		5	2.5
II		41	19.8		26	12.9
III		82	39.6		96	47.8
IV		64	30.9		64	31.8
Hypertension	204	88	42.5	201	86	42.8
Hypercholesterolemia	196	124	59.9	194	136	67.7
Smoking	199	69	33.3	198	73	36.3
Family history	195	126	60.9	191	117	58.2
Diabetes mellitus	205	38	18.4	201	33	16.4
Operative data						
Number of grafts [a]	204	3.2	1.3	200	3.1	1.2
Cardiopulmonary Bypass time in mins [a]	207	81.6	35.3	201	77.4	34.1
Aorta cross-clamp time in mins [a]	207	52.5	25.9	201	47.8	27.3
Without aorta cross-clamping	207	56	27.1	201	47	23.4
Re-do CABG	207	14	6.8	201	10	5.0

[a] in Mean (Standard deviation), [b] NYHA= New York Heart Association

Table 2: Reasons of missing data in baseline, one month and one year questionnaire

Reason		Control		SSIC	
		N	%	N	%
Baseline questionnaire					
Informative censoring					
Psychological problems		1	0.3	0	0.0
Non-informative censoring					
Lost questionnaire		0	0.0	1	0.3
	Total	1	0.3	1	0.3
One month questionnaire					
Informative censoring					
Deceased		1	0.5[a]	0	0.0
Admitted to a hospital		1	0.5	0	0.0
Non-informative censoring					
Not willing to fill in		3	1.4	2	1.0
Private or family circumstances		1	0.5	0	0.0
No telephone number / address available		2	1.0	1	0.5
Reason censoring unclear					
Admitted to a nursing home		2	1.0	0	0.0
Not reported		3	1.4	7	3.5
	Total	13	6.3	10	5.0
One year questionnaire					
Informative censoring					
Deceased		1	0.5[a]	3	1.5
Non-informative censoring					
Not willing to fill in		2	1.0	4	2.0
Private or family circumstances		1	0.5	0	0.0
No telephone number / address available		3	1.4	3	1.5
Reason censoring unclear					
Admitted to a nursing home		1	0.5	0	0.0
Not reported		18	8.7	11	5.5
	Total	26	12.6	21	10.5

[a] Same patient

Table 3: The short term and long term results of the quality of life for both the control group and the short-stay intensive care group

| | Baseline | | | | One month post-operative | | | | | | | Twelve months post-operative | | | | ANCOVA[e] | |
| | Control | | SSIC | | Control | | SSIC | | T-Test | | | Control | | SSIC | | | |
	worse-best	Mean	SD[a]	Mean	SD	Mean change[b]	SD	Mean change[b]	SD	Mean change difference in groups	95% CI[c]	P-Value	Mean change[d]	SD	Mean change[d]	SD	Time P-Value	Group X Time P-Value
Disease specific: MILQ																		
Mental health	0-28	20.4	4.29	20.6	4.21	1.09	4.62	1.10	4.30	0.01	-0.86 - 0.87	.985	-0.26	3.79	-0.02	3.63	<.001	.620
Physical health	0-28	17.3	5.42	17.2	4.96	2.08	5.50	2.54	5.07	0.46	-0.58 - 1.48	.388	1.49	4.32	1.00	4.63	<.001	.445
Physical functioning	0-28	16.3	6.63	15.5	6.03	-0.33	6.73	0.93	6.03	1.26	0.01 - 2.50	.049	4.25	5.52	3.69	5.57	<.001	.295
Social functioning	0-28	21.4	4.43	20.9	4.18	-1.15	4.75	0.00	4.61	1.15	0.24 - 2.06	.014	1.96	4.29	1.43	3.80	<.001	.070
Intimacy [f]	0-28	20.9	4.29	23.1	4.55	-0.19	4.45	0.66	4.35	0.85	-0.02 - 1.71	.054	0.27	3.78	-0.14	3.78	<.001	.527
Cognitive functioning	0-28	22.9	4.17	23.0	3.59	-0.06	3.79	-0.16	3.69	-0.10	-0.83 - 0.62	.775	0.06	3.24	0.22	3.20	<.001	.841
Financial status	0-28	23.6	4.27	23.4	4.17	-0.13	4.47	-0.40	3.69	-0.27	-1.07 - 0.53	.508	-0.52	4.03	-0.22	3.80	<.001	.558
Healthcare professionals	0-28	24.1	2.78	24.2	2.80	-0.74	3.70	-0.99	3.77	-0.25	-0.98 - 0.47	.491	-0.42	3.88	-0.14	4.26	<.001	.725
Productivity	0-21	12.9	5.08	12.2	4.82	-0.65	5.11	0.28	5.43	0.93	-0.10 - 1.95	.077	2.54	4.72	2.21	4.36	<.001	.579
Overall MILQ-Index [g]	35-245	180.2	28.8	177.8	26.9	-.10	28.1	3.86	27.3	3.96	-1.49 - 9.41	.154	9.50	24.99	7.98	25.19	<.001	.543
MILQ score	1-7	4.5	1.2	4.4	1.1	0.35	1.33	0.50	1.33	0.15	-0.11 - 0.40	.270	0.39	1.12	0.36	1.15	<.001	.508
Generic: EQ-5D																		
VAS [h]	0-100	60.4	18.2	58.04	18.9	7.00	19.0	9.99	21.8	2.99	-0.98 - 6.96	.140	6.45	16.19	5.20	17.23	<.001	.593
Mobility [i]	3-0	1.5	.53	1.5	.51	-0.23	.56	-0.18	.59	0.05	-0.07 - 0.16	.433	0.08	0.48	0.00	0.55	<.001	.217
Self care [i]	3-0	1.1	.31	1.1	.29	0.04	.39	0.07	.40	0.03	-0.04 - 0.11	.386	-0.05	0.34	-0.10	0.37	<.001	.325
Usual activities [i]	3-0	1.6	.62	1.7	.65	0.16	.79	0.06	.80	-0.10	-0.25 - 0.06	.248	-0.37	0.64	-0.36	0.64	<.001	.924
Pain/discomfort [i]	3-0	1.5	.55	1.6	.57	0.14	.70	0.11	.64	-0.03	-0.16 - 0.10	.669	-0.21	0.70	-0.28	0.56	<.001	.311
Anxiety/Depression [i]	3-0	1.5	.64	1.5	.62	-0.22	.64	-0.30	.64	-0.08	-0.21 - 0.04	.175	-0.46	0.53	-0.04	0.49	<.001	.569
Domain specific: BDI[j] and STAI-S [k]																		
Total BDI	21-0	8.0	5.55	8.4	5.27	-0.18	4.75	-0.52	5.46	-0.34	-0.13 - 0.65	.496	-0.14	5.23	-0.44	5.53	<.001	.621
Total STAI-S	80-20	44.3	12.0	45.0	12.1	-7.19	12.1	-8.58	12.8	-1.39	-3.8 - 1.03	.259	-0.38	10.13	0.14	8.87	<.001	.653

a SD= Standard Deviation, b one month – baseline score, c CI= Confidence Interval, d one year – one month score, e ANCOVA= Analysis of covariance, f n=164 control and n=167 SSIC group,
g Overall MILQ-index= Multidimensional Index of Life Quality; clinically meaningful defined as difference of 0.4 points between SSIC and control group at one year is, h VAS= Visual Analogue Scale, difference of 8 points at one year, i difference of 0.3 points at one year, j BDI= Beck Depression Inventory, k STAI-S= State-Trait Anxiety Inventory

Table 4: Results of linear regression analyses for the prediction models of quality of life at one year after surgery

	B	P-value
Model 1: Linear trend of the MILQ-score [b]		
R-square =.334		
Bo	2.43	<.001
Group	.024	.736
MILQ-score at baseline	-.435	<.001
Model 2: Linear trend of the Overall MILQ-index [b]		
R-square = .208		
Bo	77.29	<.001
Group	-.148	.934
Depression at baseline	-.406	.050
Overall MILQ-index at baseline	-.368	<.001
Model 3: Linear trend of the EQ-5D VAS [c]		
R-square = .357		
Bo	30.33	<.001
Group	-.648	.556
Gender	-5.00	<.001
Anxiety at baseline	.106	.026
EQ-5D VAS at baseline	-.424	<.001

[a] CI= Confidence Interval, [b] MILQ= Multidimensional Index of Life Quality, [c] VAS= Visual Analogue Scale

Generic quality of life

The linear change score of the EQ-5D VAS was not significantly different for both groups at one month (P-value= .140; 95% CI: -0.98 to 6.96) (table 3). The same accounts for the change scores of the five EQ-5D domains. QoL increased similarly in both groups in the first year after surgery on EQ-5D VAS (P-value=.593) and on all five EQ-5D domains (table 3).

In the regression models, the type of intensive care treatment had no statistically significant effect on the linear trend of the EQ-5D VAS (P-value=.556; 95% CI: -2.81 to 1.51; table 4). The baseline score of the EQ-5D VAS, the baseline anxiety score and gender were significant predictors of the linear trend of the EQ-5D VAS. All tests for first-order interactions and all tests on the other predictors (appendix) were not statistically significant (table 4).

Domain specific quality of life

The change score of the BDI and STAI-S did not differ in both groups at one month after surgery (P-value= .496; 95% CI: -0.13 to 0.65, P-value= .259; 95% CI: -3.8 to 1.03) (table 3).

The results of the repeated measures ANCOVA analysis show no statistically significant differences in the linear trend at one year after surgery for both groups on BDI (P-value=.621) or STAI-S (P-value=.653) (table 3).

Discussion

Summary of study findings

This large randomised controlled equivalence trial is to our knowledge the first study that evaluates the impact of a short-stay intensive care treatment (a type of fast-track treatment) on disease specific, generic and domain specific QoL in the year after surgery, in low-risk CABG patients.

After controlling for baseline differences and potential confounding factors like gender and age, the changes of disease specific, generic and domain specific QoL in the first year after surgery appear to be similar for patients who received the short-stay intensive care treatment (discharge within 8 hrs after surgery), and those who stayed overnight on the intensive care. The scores on the domains physical and social functioning of the MILQ were significantly higher in the SSIC group compared to the control group at one month after surgery. However, these differences were no longer observed at one-year follow-up.

Fast-track treatments and quality of life

The desirability of decreasing post-cardiac surgery intensive care utilization was recognized two decades ago (Johnson, 2006) with as main driving force, cost containment (Higgins, 1992). Examples of fast-track treat-

91

ments which make IC discharge within a few hours after surgery possible are for instance: the use of fast-track anesthesia, normothermic temperature management and early extubation at the intensive care. Due to the heterogeneity of the interventions, differences in case-mix, various quality of life questionnaires used and variable measurement moments used, a proper comparison between study findings is difficult. Nevertheless, our results confirm the findings of others that fast-track treatments do not have an impact on cognitive function (Cheng, 1996a; Howie, 2001; Kadoi, 2003). Royse et al. found less posttraumatic stress and depression at two and six weeks postoperative in the fast-track group receiving low-dose anesthetics compared to the group that received a high dose (Royse, 2003). We did not find any evidence of a difference in the number of depressive and anxious feelings between the two groups in the first postoperative year. No comparison with other studies on the dimension pain can be made as these trials only reported on pain in the first days after the IC-treatment (Bowler, 2002; Engoren, 2001; Fillinger, 2002; Howie, 2001; Kadoi, 2003; Mollhoff, 2001; Shroff, 1997; Zarate, 2000).

Clinical relevance of study findings

A strong point of our study is its sample size. In the current study we were able to detect the possibility of a clinically meaningful QoL difference at one year (of 0.4 points on the MILQ, of 0.3 points on the EQ-5D domains and of 8 points on the EQ-5D VAS) between the SSIC and the control group. As none of these three scales at one year showed a statistically significant difference between the two groups, the reported differences between both groups cannot be defined as clinically meaningful. This indicates that the short-stay intensive care intervention does not have a relevant adverse impact on quality of life at one year after surgery. In another study (chapter 5) we investigated the validity of both the MILQ and the EQ-5D. One of the findings was that the baseline ceiling effects (a high percentage of patients with a maximum score) in all EQ-5D di-

mensions were substantial, but these were not found on the MILQ-score and EQ-5D VAS. For the current study this indicates that there is a possibility that not all relevant changes in QoL measured by means of the EQ-5D domains were identified, however as outcomes reported by means of the EQ-5D domains are consistent with those of the EQ-5D VAS and the MILQ, the impact of this on the study conclusions is probably negligible.

The differences at one month postoperative on the physical and social functioning of the MILQ were statistically significant different between the two groups, however these differences were no longer observed at one year follow-up. As the power calculation was based on one-year differences of QoL between the two groups it is not possible to make a proper judgment whether the reported small positive effects in the SSIC group at short term are of clinical relevance.

In a previous paper we also reported generic QoL at one month after surgery as an outcome measure in the EE of the short-stay intensive care treatment. The patient's answers to the five EQ-5D items were used to calculate an utility value based on population weights (Dolan, 1997) (ranging from 0 to 1) and subsequently the Δ Quality Adjusted Life Months (Δ QALM's) was calculated. The generic QoL significantly improved more in the SSIC group compared to the control group. In the current study this outcome was not reported because the aim of this paper was to report data from a patient point of view and not from a societal point of view. The observed difference measured at one month was very small (0.0238) and lower than the score of 0.03 which in the current study is considered as being clinically relevant (Marra, 2005).

If we compare the change scores of the EQ-5D at 12 months after surgery in our study with the results in patients receiving planned surgery reported by Badia et al. (Badia, 2001), the results of the domains self care, usual activities, pain/discomfort of the EQ-5D and the EQ VAS are very similar, but the scores on the other domains, mobility and anxiety/depression, are different. The mean score on mobility improved clini-

93

cally meaningful were as in our study no such an improvement was observed. In contrast, patients in the SSIC study reported a decrease on the psychological dimension of the EQ-5D whereas patients planned for various types of surgeries this score did not change. An explanation for these discrepancies between the two studies can be related to differences in the patient characteristics. For instance, the mean age was lower in the study performed by Badia et al. compared to the mean age of our patients. Furthermore, not only CABG patients but also liver transplant patients and other types of cardio surgical patients were included in the Spanish study. Another explanation can be related to the differences in measuring the baseline quality of life. The retrospective rating of the baseline QoL in the Spanish study can have resulted in an under- or overestimating of the scores and hence, in an under- or overestimation of the QoL change scores.

Study limitations

This study has the following limitations. Ideally, preoperative administration of the QoL questionnaires should be done before emotional responses could confound the QoL measures, but this was not possible due to logistic reasons. However, as this study was comparative this cannot have influenced study findings much.

A measurement of pain within a few days after intensive care treatment was not performed, as it was practically impossible to perform this.

The generalizability of this study may be somewhat limited because in the group of the primary non-responders, older patients and females were over represented compared with the group of the responders. However, as all baseline characteristics for the responders and the non-responders of the two treatment modalities were comparable this cannot have had a large impact on our study results.

Finally, fatigue of the responders may have biased the study findings.

The limitations concerning the study design and patient eligibility are described elsewhere (van Mastrigt, 2006).

Conclusions and recommendations

This randomised controlled equivalence trial evaluated thoroughly the impact of a short-stay intensive care protocol on quality of life in low-risk CABG patients. The short-stay intensive care treatment appears to have no effect on disease and generic QoL in the first year after surgery. As none of the observed QoL differences between the two groups can be defined as clinically meaningful, and safety and cost savings of the SSIC intervention already have been demonstrated (van Mastrigt, 2006), short-stay intensive care can also from a patient perspective be considered as a good alternative for postoperative IC treatment in low-risk CABG patients.

Appendix: Potential predictors of quality of life used in the regression models

Independent	Source	Coding	Description
Group	Case Record Form	care as usual=0 SSIC=1	
Demographic data			
Age (Years)	database	continuous	high age (Chocron, 1996; Herlitz, 2005; Welke, 2003)
Sex (female)	idem	female=1 male=0	female (Chocron, 1996; Herlitz, 2005; Herlitz, 1999; Phillips Bute, 2003; Simchen, 2001; Yun, 1999)
Civil state	self reported	with partner=1 without partner=0	no partner (Phillips Bute, 2003)
Education	idem	primary school not finished yes=1, no =0	low education (Duits, 1997)
Previous cardiac and medical history			
Left ventricular ejection fraction >50%, <50%	database	<50%=1 >50% =0	(Caine, 1999; Chocron, 1996)
NYHA classification angina [a] I- II, III-IV	idem	III-IV=1, I-II=0	NYHA classification III-IV (Chocron, 1996; Simchen, 2001)
Hypertension	idem	yes=1 no =0	history of hypertension (Herlitz, 2005; Herlitz, 1999; Yun, 1999)
COPD [b]	idem	yes=1 no =0	history of COPD (Herlitz, 2005; Herlitz, 1999; Welke, 2003)
CVA [c]	idem	yes=1 no =0	history of CVA (Al-Ruzzeh, 2005; Herlitz, 2005; Herlitz, 1999)
Smoking	idem	yes=1 no =0	history of smoking (Schrader, 2004)
Diabetes mellites	idem	yes=1 no =0	diabetes (Herlitz, 2005; Herlitz, 1999; Yun, 1999)
Quality of life data			
Baseline depression	self reported		depressive feelings at baseline (Duits, 1997; Schrader, 2004)
Baseline anxiety	idem	continuous	anxious feelings at baseline (Duits, 1997; Schrader, 2004)
Baseline EQ-5D-VAS	idem	continuous	baseline QoL (Herlitz, 2005; Herlitz, 1999; Rumsfeld, 2001)
Baseline MILQ index	idem	continuous	baseline QoL (Herlitz, 2005; Herlitz, 1999; Rumsfeld, 2001)
Other data			
Postoperative (life) events	idem	yes=1 no =0	'stressful life event' (Chocron, 1996; Pirraglia, 1999)

[a] NYHA=New York Heart Association, [b] COPD= chronic obstructive pulmonary disease, [c] CVA= cardiovascular accident.

Part

2

Methodological aspects of measurement and analysis of quality of life

CHAPTER
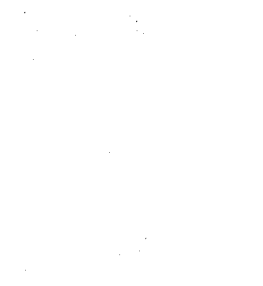
Testing validity and reliability of the EuroQoL and the disease specific Multidimensional Index of Life Quality in patients with coronary heart disease

Submitted

Ghislaine van Mastrigt
Manuela Joore
Fred Nieman
Jos Maessen
Hans Severens

Abstract

Objective: To assess the validity and the reliability of the EuroQoL (EQ-5D) and the Multidimensional Index of Life Quality (MILQ) in coronary artery bypass surgery patients.

Study design: Prospective comparative study.

Setting: University Hospital.

Patients: 410 CABG patients completed both EQ-5D and MILQ preoperatively, and at one and twelve months postoperatively.

Results: At baseline all five EQ-5D dimensions (35.7% to 90.0%) showed substantial ceiling effects, which were also observed in the MILQ dimensions financial status and healthcare professionals. The effect sizes (or the predictive validity) at one month of the EQ-5D domain anxiety/depression, EQ-5D Visual Analogue Scale (EQ-5D VAS), the MILQ domain physical health and the MILQ-score were moderate (effect sizes: -5.0 to .67). At 12 months postoperatively, the EQ-5D VAS, the MILQ-score and the two physical dimensions of the MILQ had a good predictive ability (effect sizes: .89 to 1.0). Only the EQ-5D domain usual activities and the EQ-5D VAS distinguished patients who were classified in different New York Heart Association functional classification categories. The test-retest reliability of both questionnaires can be defined as not adequate.

Conclusions: The discriminative validity of the EQ-5D domains and the external criterion, the predictive validity and reliability of both the EQ-5D and the MILQ were limited. Therefore, these two questionnaires are not the first choice when evaluating quality of life in CABG patients.

Introduction

Quality of life (QoL) is an important outcome measure in randomised controlled clinical trials. To measure this outcome it is recommended to use two types of instruments (Dempster & Donnelly, 2000). Firstly, disease specific QoL questionnaires are mostly developed from a patient perspective, for a specific disease and are in general more responsive compared to generic ones (Wiebe, 2003). Secondly, generic instruments, as they are developed from community perspective, measuring general health and therefore make comparisons between various populations possible. As policy decisions regarding the reimbursement of healthcare interventions are based on general health outcomes, generic QoL instruments are essentially from a 'decisional perspective' (Dowie, 2002). On the other hand, patients, healthcare professionals and policy makers value a deep understanding of the benefits and harms an intervention produces (Guyatt, 2002). Therefore, from a 'knowledge perspective' disease specific QoL measures render important outcomes.

When selecting a QoL instrument, it is important to know how well it will perform in providing the most appropriate and required information (Thompson, 1998). Therefore, both the reliability and validity of a questionnaire should be explored. These are not one-time-only attributes: they need to be re-established when the instrument is used in different populations or cultures (Thompson & Yu, 2003). Various investigators investigated the psychometric proportions of both generic and disease specific measures used in cardiac patients, like for instance Short Form-36, Quality of Life after Myocardial Infarction (QLMI) and Seattle Angina Questionnaire (SAQ) (Dempster & Donnelly, 2000; Smith, 2000). They conclude that the available instruments have a lack of sensitivity to change and further research is needed on this topic. In this study the psychometric proportions of the EQ-5D and the Multidimensional Index of Life Quality (MILQ) are examined. The EQ-5D (EuroQoL-Group, 1990) developed

by a consortium of investigators in Western Europe is nowadays a well-accepted and broadly used generic QoL measure. This questionnaire can be used in two different ways, firstly as a generic utility measure (based on the health state valuations it is suitable for economic evaluations), and secondly as a measure to generate health profiles of patients. The psychometric proportions of the EQ-5D were previously assessed after myocardial infarction and cardiac rehabilitation (Nowels, 2005; Schweikert, 2006), however the validity and the reliability of this questionnaire in coronary artery bypass surgery patients was never assessed. The MILQ was developed by Avis et al. for the measurement of QoL in cardiovascular patients (Avis, 1996). Although the MILQ seems to be a valid and a reliable instrument for the measurement of QoL, this questionnaire has rarely been used in studies (Falger, 2000; Sleeper, 2005; Smith, 1999; Smith & Larson, 2003). The objective of this study was the evaluation of the psychometric proportions of both the EQ-5D and the MILQ in CABG patients regarding the response and the item completion rate, the discriminative, the predictive and the external criterion validity, and the test-retest reliability.

Materials and methods

Design

For this prospective comparative study data were used of CABG patients who were included in the short-stay intensive care trial between February 2001 and March 2003 at the University Hospital in Maastricht. In this study a stay of 8 hours intensive care was compared with the usual treatment of overnight intensive care stay. A detailed description of exclusion criteria, short-stay intensive care protocol, clinical effectiveness, cost-effectiveness and QoL data is available elsewhere (chapter 3 and 4).

The ethical and research council of the University Hospital Maastricht approved the study.

Procedures

After obtaining informed consent from the patients the QoL questionnaires were self-administered at three times: the first time at the day before surgery at the hospital, then twice at home, at one and twelve months after surgery. If a patient was unable to complete the baseline questionnaire by him-/her self, a research nurse assisted by filling in the form. At follow-up the patients were asked to respond without help from another person(s), but if this was not possible, to ask a relative or friend to assist. Patients who did not respond to the follow-up questionnaires received a reminder letter together with a new copy of the questionnaire. If again they did not return the questionnaire, they were contacted by telephone to find out the reasons of non-response.

Quality of life questionnaires

Generic QoL was measured by the EQ-5D (Brooks, 1996). The EQ-5D consists of two components: the EQ-5D descriptive system and the EQ Visual Analogue Scale (EQ-5D VAS). In the descriptive system the respondent is asked to rate his or her health by checking one of three levels of severity - 'no problems' (coded as 1), 'some or moderate problems' (coded as 2), 'severe problems or unable to perform' (coded as 3)- in each of the following five EQ-5D dimensions: mobility, self care and pain/discomfort, usual activities and anxiety/depression. For the EQ-5D VAS, participants draw a line from a box to the point on the thermometer-like scale (with a range from 0 'the worst imaginable health state' to 100 'the best imaginable health state') corresponding to their health state. At follow-up, patients were asked whether their general health had been improved, deteriorated or stayed the same, compared to the previous situation.

The MILQ is a QoL measure especially developed for patients with cardiovascular disease (Avis, 1996). In the first item (referred to as 'MILQ-score') the respondent is asked to value his/her general QoL on a seven point-Likert scale (from 1= worst to 7= best possible QoL). In the other 35 items the respondent's degree of satisfaction with various aspects of QoL on a seven point-Likert scale (from 1= very dissatisfied to 7= very satisfied) is scored. These items cover the following nine domains: mental health, physical health, physical functioning, cognitive functioning, social functioning, intimacy, financial status, relationship with healthcare professionals and productivity. The domains contain each 4 items, so the additive domain score can range from 4 to 28 (except for the domain productivity which contains 3 items, for which the maximum score lies theoretically between 3 and 21). The Dutch translation of the MILQ already has been successfully used to assess the QoL of CABG patients (Falger, 2000). An unweighted index was calculated by summarizing overall nine domain scores. The total overall score range from 35 to 245 is referred to as the 'overall MILQ-index'.

On the average, answering the EQ-5D takes five minutes, while answering the MILQ takes fifteen minutes.

Psychometric analysis methods

Firstly, the response and the item completion rates are assessed. These rates are investigated by calculating the proportion of the missing questionnaires and the proportion of missing items of the respondents within both questionnaires. It is assumed that both questionnaires achieve an acceptable response and item completion rate.

Secondly, the discriminative validity of both questionnaires is assessed. Discriminative validity is defined as the instrument's ability to detect changes in the observed variable without provoking 'floor' or 'ceiling' effects. For each domain the proportion of the respondents with a minimum score (referred to as 'floor effects') or a maximum score (referred

to as 'ceiling effects') is calculated. Floor and ceiling effects over 15% are considered critical (McHorney & Tarlov, 1995). At forehand it is expected that floor effects will be rarely (less than 10%) and ceiling effects will more frequently be observed in both questionnaires. Kolmogorov-Smirnov tests (K-S) for assessing normality are performed.

Thirdly, the predictive validity (or responsiveness) is tested. It is assumed that this type of validity is reflected by its ability to detect small but important changes in QoL. Only patients who reported a change (improvement or deterioration) in their health state were included for these analyses. Two statistics were calculated to assess the predictive validity: the Standardized Response Mean and the Standardized Effect Size. The Standardized Response Mean (Cohen, 1988) is calculated: (SRM= (M1-M2)/(SD1-SD2)), where as M1 is the mean pre-assessment and M2 is the mean post assessment and SD1 is the standard deviation of the pre-assessment and SD2 is the standard deviation of the post assessment. The Standardized Effect Size (SES) is calculated: ((M1-M2)/(SD1)). An SES or SRM of less than 0.2 can be described as a 'minimal' effect; an SES or SRM of 0.2 to 0.5 can be defined as a 'small' effect, 0.5-0.8 as a 'moderate' effect and an SES or SRM above 0.8 as a 'large' effect (Cohen, 1988). It is expected that the predictive validity of the MILQ will be better compared to the predictive validity of the EQ-5D.

Fourthly, external criterion validity is assessed for both questionnaires. It is assumed that this type of validity is reflected by its ability to discriminate at baseline between patients with mild and severe disease by showing higher QoL scores in mild compared to lower QoL scores in more severely diseased patients. The New York Heart Association functional classification (NYHA) classes (I-II, III and IV) are used as a criterion for cardiovascular disease severity (NYHA, 1994). To test the discriminative ability, the independent sample t-test was performed for the overall MILQ-index while all other scores were analyzed using Mann-Whitney U tests. It is expected

107

that the MILQ and the EQ-5D are able to discriminate between the different NYHA classes.

Finally, the one-month test-retest reliability of QoL of the consecutive measurements (before surgery and at one month postoperatively) of both questionnaires is assessed for the sub sample of patients who indicated that their QoL remained the same. The proportion of agreement and a kappa statistic (k) is calculated for both the five EQ-5D domains and the MILQ-score. For continuous data of the MILQ domains and the EQ-5D VAS the Intraclass Correlation Coefficients (ICC's) are calculated (Jenkinson & McGee, 1998). This is done by reliability analyses (two-way random effects model based on consistency agreement). The following four categories were used: k or ICC of >.20 indicating poor agreement; scores from .21 to .40 suggest a fair degree of agreement, values from .41 to .60 represent moderate and >.60 good or adequate agreement (Altman, 1999). It will be expected that both questionnaires have at least a kappa or ICC score of .60 (McDowell & Newell, 1996).

A P-value of less than 0.05 was assumed to be statistically significant. All data were analyzed with SPSS, version 12.0.1 (SPSS, Chicago, IL).

Results

Study population

Six hundred patients gave informed consent and participated in the short-stay intensive care trial performed at the University Hospital Maastricht. The data of this study are based on the first 410 patients. Two of the included patients had to be excluded from the study before treatment allocation was revealed.

At randomization, the mean age of the patients was 62 years. Only twenty percent of the patients were females. Eighty patients were cate-

gorized in NYHA class I-II (20.7%), 178(46.1%) in NYHA class III and 128 (33.2%) in NYHA class IV.

Response rate and item completion rate

The response rates were very high at baseline; only two (0.6%) question-naires were missing. At one and twelve months postoperative, the non-response rates were respectively 23 (11.3%) and 47 (23.1%). The percent-age of missing items in the EQ-5D was less then 1 %. For the MILQ it ranged from 2 to 5%. Patients living alone (n=71) did not complete the four questions of the MILQ dimension intimacy (data not shown).

Table 1: Discriminative validity, preoperative data of the EQ-5D and the MILQ

	Floor effects (%)	Ceiling effects (%)	Mean	SD ᵃ	Median	Quartile ranges	Skewness
EQ-5D							
Mobility	0.9	44.3	1.53	0.52	2	1-2	0.09
Self Care	0.9	90.0	1.09	0.30	1	1-1	3.52
Usual activities	10.9	35.7	1.65	0.44	2	1-2	0.44
Pain/discomfort	4.8	39.6	1.60	0.56	2	1-2	0.25
Anxiety/depression	7.8	53.4	1.50	0.63	1	1-2	.848
EQ-5D VAS	0.0	0.4	59.22	18.56	60	50-75	-0.23
MILQ							
Mental health	0.0	3.5	20.51	4.25	21	18-24	-0.73
Physical health	0.0	0.9	17.22	5.19	17	13-21	-0.19
Physical functioning	0.0	1.3	15.92	6.35	16	11-22	0.03
Social functioning	0.0	6.1	21.13	4.31	22	19-24	-0.83
Intimacy	0.0	2.6	21.05	4.33	22	18-24	-0.87
Cognitive functioning	0.0	7.4	22.93	3.89	23	22-25	-1.80
Financial status	0.0	21.3	23.48	4.21	24	22-27	-1.69
Healthcare professionals	0.0	15.2	24.17	2.79	24	23-26	-1.12
Productivity	0.0	2.2	12.55	4.96	13	9-17	-0.23
Overall MILQ-index	0.0	0.0	179.00	27.88	180	160-200	-0.27
MILQ-score	0.9	1.7	4.46	1.16	5	4-5	-0.16

ᵃ SD = standard deviation

Discriminative validity

At baseline ceiling effects on the EQ-5D dimensions were substantial, varying from the lowest in the dimension usual activities (35.7%) to the highest percentage in the dimension self care (90.0%) (table 1). Ceiling effects in the MILQ were observed in only two dimensions: financial status (21.3%) and healthcare professionals (15.2%). Floor effects were not observed in the EQ-5D and MILQ. Only the overall MILQ-index showed a normal distribution (K-S test, P-value =.189).

Table 2: Predictive validity of the EQ-5D and the MILQ of patients who reported a change in their health state (n=259)

	One month postoperative		Twelve months postoperative	
	SES[a]	SRM[b]	SES[a]	SRM[b]
EQ-5D				
Mobility	-.47	-.40	-.38	-.34
Self care	.13	.10	-.10	-.09
Usual activities	.05	.04	_-.51_	_-.44_
Pain/Discomfort	.16	.12	-.30	-.24
Anxiety/Depression	_-.50_	_-.51_	_-.54_	_-.54_
EQ-5D VAS	_.67_	_.54_	**1.0**	**.83**
MILQ				
Mental health	.35	.34	.29	.29
Physical health	_.65_	_.61_	**.90**	**.77**
Physical functioning	.26	.25	**.89**	**.74**
Social functioning	-.11	-.09	.33	.28
Intimacy	.13	.12	.14	.13
Cognitive functioning	-.03	.03	.10	.09
Financial status	-.01	-.02	-.14	-.15
Healthcare professionals	-.26	-.20	-.31	-.24
Productivity	-.16	-.15	_.60_	_.51_
Overall MILQ-index	-.25	.25	_.58_	.22
MILQ-score	_.59_	.49	**.92**	**.81**

Moderate effects are underlined, Large effects are indicated in bold, [a] SES= Standardized Effect Size, [b] SRM= Standardized Response Mean

Predictive validity

The data in table 2 give an indication of the predictive validity of the EQ-5D and MILQ questionnaire of the 259 patients who reported a **change** in their health state. A moderate effect size was observed for the following 'one-month' change scores: the EQ-5D dimension anxiety/depression,

the EQ-5D VAS, the MILQ dimension physical health and the MILQ-score. For the EQ-5D dimension mobility, the MILQ dimensions mental health, physical functioning, healthcare professionals and the overall MILQ-index small effects were observed. In all other dimensions the SES's and SRM's at one month can be categorized as minimal.

Large SES's and SRM's were observed for the 'one year' change scores of the EQ-5D VAS, the MILQ dimensions physical functioning and physical health, and the MILQ-score. Moderate effects were observed for the EQ-5D dimensions usual activities and anxiety/depression, the MILQ dimension productivity and the overall MILQ-index. The effects of all other scores can be categorized as minimal or small.

Table 3: External criterion validity, distribution of the MILQ and the EQ-5D over the New York Heart Association functional classification classes in median (quartile range)

	New York Heart Association functional classification		
	I-II	III	IV
	N=80	N=178	N=128
EQ-5D			
Mobility	1 (1-2)	2 (1-2)	2 (1-2)
Self care	1 (1-1)	1 (1-1)	1 (1-1)
Usual activities	<u>1 (1-2)*</u>	2 (1-2)	<u>2 (1-2)**</u>
Pain/Discomfort	**1 (1-2)**	**2 (1-2)**	2 (1-2)
Anxiety/Depression	1 (1-2)	1 (1-2)	1(1-2)
EQ-5D VAS	<u>**70 (50-80)**</u>	**60 (50-70)**	<u>58 (40-75)</u>
MILQ			
Mental health	22 (19-24)	21 (18-24)	20 (17-24)
Physical health	19 (14-22)	17 (12-21)	18 (13-22)
Physical functioning	18 (11-23)	15 (11-20)	15 (10-22)
Social functioning	22 (19-24)	21 (18-24)	23 (19-24)
Intimacy	22 (18-24)	21 (18-24)	23 (19-24)
Cognitive functioning	23 (20-25)	23 (22-25)	24 (22-26)
Financial status	24 (22-27)	24 (22-26)	24 (22-27)
Healthcare professionals	24 (23-26)	*24 (23-25)****	*24 (24-26)*
Productivity	15 (9-18)	12 (8-15)	13 (8-18)
Overall MILQ-index	188 (159-204)	172 (160-196)	180 (159-203)
MILQ-score	5 (4-5)	4 (4-5)	4 (3-5)

*Indicated in **Bold** significant difference (P-value<.005) between NYHA classes I-II and III; usual activities (P-value=.002), pain/discomfort (P-value=.003), EQ-5D VAS (P-value=.003), **<u>Underlined</u> significant difference between NYHA classes I-II and IV; usual activities (P-value=.009), EQ-5D VAS (P-value=.003), ***indicated in *Italic* significant difference between NYHA classes III and IV; relationship with healthcare professionals (P-value=.022)

External criterion validity

Table 3 presents the median and quartile range of the scores on both questionnaires of patients classified into the different New York Heart Association functional classification categories. The EQ-5D dimension usual activities and the EQ-5D VAS discriminate between patients in the NYHA category I-II versus III as well as in the category I-II versus IV. Furthermore, the EQ-5D dimension pain/discomfort discriminated between NYHA categories I-II versus III. The MILQ dimension healthcare professionals was able to differentiate between NYHA category III and IV.

Table 4: Test-retest reliability, agreement of the MILQ and the EQ-5D, of the patients who reported a stable health state (n =149)

	Proportion of agreement N (%)	Kappa [a]
EQ-5D		
Mobility	106 (72.1)	.413
Self care	129 (86.6)	.222
Usual activities	72 (48.2)	.115
Pain/Discomfort	89 (59.7)	.219
Anxiety/Depression	101 (67.8)	.297
MILQ		
MILQ-score	37.5 (54.0)	.147
	Mean difference (standard deviation) Test-retest	**ICC**
EQ-5D		
EQ-5D VAS	4.58 (18.1)	**.433**
MILQ		
Mental health	.399 (4.7)	.395
Physical health	.980 (5.1)	.404
Physical functioning	-1.57 (6.0)	.414
Social functioning	-.87 (4.4)	**.550**
Intimacy	-.130 (4.1)	**.550**
Cognitive functioning	-.618 (3.6)	**.550**
Financial status	-.622 (4.8)	**.465**
Healthcare professionals	-1.20 (3.9)	.229
Productivity	-1.57 (5.1)	.400
Overall MILQ-index	4.58 (27.6)	**.520**

[a] Moderate kappa and ICC scores are indicated **in bold**

Test-retest reliability

Table 4 shows the results among those CABG patients who reported no change in QoL in the first postoperative month compared with the baseline score. The agreement measured by means of the kappa for the EQ-5D dimensions can be classified as poor for usual activities, moderate for mobility and fair for the domains: self care, pain/discomfort and anxiety/depression. In the EQ-5D VAS, the MILQ dimensions mental health and physical health QoL improved, in all other MILQ dimensions a deterioration of QoL was observed. The EQ-5D VAS and six out of nine dimensions of the MILQ the ICC can be classified as moderate.

Discussion

The purpose of this study was to evaluate the psychometric proportions of a generic QoL measure (EQ-5D) and a disease specific measure (Multidimensional Index of Life Quality (MILQ)) in patients with coronary heart disease. The methods used to examine this will be discussed below.

Response rate and item completion rate

In terms of practicality, it was expected that the MILQ would be a more difficult instrument to complete than the EQ-5D. This appears to be supported by the higher proportion of missing items in the MILQ compared to the EQ-5D. However, the number of missing data in both questionnaires was lower compared to other studies (Holland, 2004; Schweikert, 2006). Although the questionnaires were applied in a trial setting and it is expected that the percentage item response would be lower in a real world setting, these findings indicate that the two questionnaires are well understood by CABG patients. It should be noted that the EQ-5D was completed prior to the MILQ at baseline, and there is a theoretical possibility that filling in the former will be influenced by filling in the latter.

Discriminative and predictive validity

At one month after surgery in both questionnaires no large effects were observed for those patients who reported a change in their health state. This indicates that at short term after surgery it can be difficult to measure change (improvement and deterioration) in QoL with both questionnaires.

A responsiveness of the EQ-5D was reported in a study in which the EQ-5D was administered at 12 months after intensive care unit discharge by patients who were received a planned liver transplant or cardiac surgery (Badia, 2001). These findings could not be confirmed in our study, because at 12 months after surgery only for the EQ-5D VAS large effects were found.

For the MILQ domains mental health, social functioning, intimacy, cognitive functioning, financial status and healthcare professionals, the responsiveness was lower compared to the more 'physical' MILQ domains physical health, physical functioning and productivity. An explanation for this finding can be that patients undergoing a CABG surgery experience the most benefit on the physical aspects of their quality of life. Another explanation can be that the MILQ questionnaire was not sensitive enough to detect a change in these 'non'-physical dimensions of quality of life.

The predictive validity of the MILQ was better compared to EQ-5D. Differences in the characteristics of both questionnaires can provide an explanation of these findings. Firstly, the 7 point-Likert items of the MILQ are likely to be more sensitive to changes compared to the 3 point-Likert items of the EQ-5D. Secondly, the EQ-5D contains for every dimension only one item, while the different MILQ domains contain 4 items each. Furthermore, a high proportion of patients reporting ´no problems´ (scoring 1 on the different items) probably will influence the responsiveness of the EQ-5D. This phenomenon known as ´ceiling effect´ was previously reported in EQ-5D in cardiac patients with acute coronary syndromes

114

and patients with acute myocardial infarction (Macran, 2003; Nowels, 2005; Schweikert, 2006). The consequences of significant ceiling effects for individual scores can be that measurement of an improvement from the baseline score is not possible; in other words there is a possibility that the questionnaire is unable to detect clinically significant changes at the higher spectrum of QoL.

External criterion validity

It was assumed that both the MILQ and the EQ-5D were able to detect cross-sectional differences between individuals by an independently external measured criterion. The NYHA functional classification was used as such a criterion (NYHA, 1994).

Dougherty et al. reported that the Seattle Angina Questionnaire (disease specific) was able to discriminate between coronary artery disease patients in the four Canadian Cardiovascular Society functional classes (comparable with the four NYHA classes) (Dougherty, 1998). Furthermore, they found that the Short Form-36, a generic QoL measure, could only distinguish for selected subscales between the Canadian Cardiovascular Society functional classes. This ability was also reported for the EQ-5D in patients with an acute myocardial infarction (Nowels, 2005). We were not able to confirm these findings. Only the EQ-VAS and the dimension usual activities of the EQ-5D were able to distinguish between different NYHA classes. As the dimension usual activities directly measures functional status, which is represented in the different NYHA classes, this is not surprising. Contrary to this, the domains physical functioning and productivity of the MILQ and the dimensions self care and mobility of the EQ-5D, which are also related to physical aspects of QoL, were not able to distinguish between severe or milder cardiac disease states.

Test-retest reliability

The good test-retest reliability scores of the MILQ reported by Avis et al. (Avis, 1996) could not be confirmed in our study. The reason for this can be the difference in test-retest period in both studies. We used a recall period of one month whereas they used a two-week interval period (10 to 21 days). In addition, our data were collected by means of self-administration, and their data were collected only by means of a tele-phone interview. Our EQ-5D reliability scores were lower compared to the reliability observed for patients with acute coronary syndrome (Schweikert, 2006). However, in the study performed by Schweikert et al. the number of patients who were assessed (reporting no change in QoL) was limited (n=11). The recall period in that study was relatively long (3 months) compared to the one-month period in our study.

Study limitations

A limitation of our study is that its not possible to extrapolate our findings to other patient groups like for instance high-risk CABG patients. Another drawback is the method of assessment of the reliability; a golden stan-dard for the assessment of test-retest reliability of questionnaires is the assessment of QoL in a patient population whose health status is stable. However, there is some evidence that health transition questionnaires are also valid (Fitzpatrick, 1993).

Conclusion

The item completion rate of both the EQ-5D and the MILQ are accept-able. Contrary to the EQ-5D, ceiling effects are rarely observed in the MILQ. Both questionnaires were not sensitive to change at short term. Only the EQ-5D VAS, the MILQ-score and the two physical dimensions of the MILQ had a good predictive ability at 12 months after surgery. In ad-dition, the external criterion validity for disease severity and the test-retest

reliability of both the EQ-5D and the MILQ can be classified as less adequate. Based on these findings, we conclude that for the assessment of QoL in low-risk CABG patients the MILQ and the EQ-5D are not to be the preferred questionnaires. Further research on the psychometric proportions of generic and disease specific QoL measures suitable for the evaluation of QoL in cardiac patients is needed to identify the most valid and reliable instruments.

CHAPTER

How to deal with imbalances in baseline utility between treatment groups in quality adjusted life year calculations?

Submitted

Ghislaine van Mastrigt
Thea van Asselt
Fons Kessels
Jos Maessen
Hans Severens

Abstract

Background: To our knowledge, a comparison of different adjustment methods for imbalances in baseline utility between treatment groups in Quality Adjusted Life Year (QALY) calculations has not been performed.

Methods: The data of a trial based cost-effectiveness analysis evaluating a short-stay intensive care treatment for coronary artery bypass patients was used as an example. In the current study five different methods suited for the analysis of quality of life data on a patient level were used for calculation of QALY differences: 1) Conventional QALY calculation method (no adjustment for imbalances in baseline utility); 2) Mean difference adjustment method: correction with the differences in the mean baseline utility of the treatment groups; 3) Delta adjustment: uses the change in utility over time; 4) Regression-based utility adjustment: adjusting the follow-up utility with a regression model with the follow-up utility as dependent variable, and baseline utility as independent variable; 5) Regression-based QALY adjustment: adjusting the QALY with a regression model with the QALY as dependent variable, and baseline utility as independent variable.

Results: The mean and standard error for baseline utility score for control and experimental group were respectively .71 (.01) and 0.66 (.01). The mean total hospital costs during follow-up were significantly lower in the experimental group: €-816 (95% CI: €-1,581 to €-174). Cost-effectiveness acceptability curves show that the different adjustment methods result in different study outcomes.

Conclusion: As imbalances between treatment groups in baseline utilities can have an impact on the QALY as outcome, we recommend reporting the outcomes of the regression-based QALY adjustment in the base case analysis and the outcomes of the other proposed adjustment methods in additional sensitivity analyses.

120

Introduction

Quality Adjusted Life Year (QALY) is a well established outcome in randomised clinical trial based cost-effectiveness analyses (Drummond, 2005). In trials, it is common to report the distributions of several baseline variables by treatment group and to identify any 'unlucky' imbalances between treatment groups that may have arisen by chance (Pocock, 2002). The importance of the impact of baseline differences depends on the strength of the association of the variable with the outcome (Pocock, 2002). This certainly holds for studies analysing QALY, because baseline quality of life (EuroQoL-Group) or the baseline utility is incorporated in QALY calculations and can therefore directly impact the study findings. Furthermore, it is known that a patient's baseline utility is likely a strong predictor for follow-up QoL measurements (Manca, 2005). However, as Richardson et al. demonstrated, only five of the 23 published trial based cost-effectiveness studies performed adjustments for baseline utility if baseline imbalances in utility occurred (Richardson & Manca, 2004).

In literature two different methods are described to adjust for imbalances in baseline utility in QALY calculations. The first one, known as 'delta QALY method' (Δ QALY), corrects by using only the change over time in QoL (increase or decrease in utility), instead of calculating the area under the curve, which is used in conventional QALY calculations. Several studies used the delta method to calculate the QALY as study outcome (Cheng, 2000; Guthrie, 1999; Hurskainen, 2001; Lee, 2002; Patterson, 1995). Manca et al. proposed the use of a regression method in the calculation of the QALY when baseline QoL imbalances do occur (Manca, 2005). However, this regression-based method only corrects for baseline imbalances on a group level and not on an individual level. In our opinion, the latter is preferred because patient data are needed for applying bootstrap analyses in order to identify the uncertainty around the estimated incremental cost-effectiveness ratios (ICERs) and constructing

121

cost-effectiveness acceptability curves (CEACs) (Briggs, 1997; Fenwick, 2004).

The purpose of the present study is to compare the conventional QALY calculation method with different adjustment methods for baseline utility imbalances between treatment groups which allow analyses of data on a patient level, and which can be applied in QALY calculations. The results of the different QALY calculation methods will be illustrated by means of the data of a randomised controlled trial. Finally, recommendations will be given regarding which adjustment method to apply when imbalances in baseline utility occur.

Materials and methods

The case study

The data of the short-stay intensive care study, a randomised controlled clinical equivalence trial designed to evaluate the safety and cost-effectiveness of two types of intensive care treatment, was used. Six hundred low-risk coronary artery bypass patients were randomised for either discharge from the intensive care within 8 hours after admission (experimental group, e), or overnight stay at the intensive care (control group, c).

The utilities were measured at baseline, before randomization (patients were blinded for group assignments) and at one month postoperative, using the EQ-5D descriptive system (EuroQoL-Group, 1990) and valued by means of the British valuation set (Dolan, 1997). For a simplification of interpretation of the results of the present study, the utilities measured at one month in the trial are presumed to be the utility scores at one year in order to make it possible to calculate a reasonable QALY.

The cost analyses were performed from a hospital perspective with a follow-up of one month postoperative. The included costs were inpatient

hospital days, clinical procedures and outpatient procedures. A detailed description of the short-stay intensive care treatment protocol, study and results is available elsewhere (chapter 3).

Different methods for calculating QALYs

In the following section five different methods are described that can be used to calculate the QALYs.

An overview of the different formulas is presented in table 1 to make comparison between the different QALY calculation methods.

1 Conventional method (no baseline adjustment)

The equations 1 and 2 give a description of the QALY calculation for the experimental or control group given one follow-up measurement.

(1) $Qe = \frac{1}{2} (Ue_0 + Ue_1)$

(2) $Qc = \frac{1}{2} (Uc_0 + Uc_1)$

Qe and Qc represent the QALY of the patient in the experimental and in the control group, Ue_0 and Uc_0 the baseline utility of the patient in the experimental and in the control group, and Ue_1 and Uc_1 the follow-up utility of the patient in the experimental and in the control group. The incremental QALY (IQ) between the experimental and the control treatment is calculated by subtracting the mean QALY of the experimental group from the mean QALY of the control group (table 1).

2 Mean difference adjustment

In this method the incremental QALY is adjusted according to the difference of the mean baseline utility between the two groups. The adjustment with the mean difference can be applied as follows: the baseline and follow-up utility in the group with the lowest mean baseline utility (in our case study the experimental group) is increased with half of the

123

mean baseline utility difference, and in the group with the highest mean baseline utility (in our case study the control group) the baseline and follow-up utility is decreased with half of the mean baseline utility difference. For the calculation of the QALY at patient level this means for the experimental group subtracting a quarter of the difference between mean baseline utilities of the experimental and the control group (Mean difference adjustment A, table 1). One could also decrease with the mean baseline utility difference between the experimental and the control group in the group with the highest mean baseline utility score (in our case study the control group) (Mean difference adjustment B, table 1). Another way of applying the mean difference adjustment is adding the baseline difference in the group with the lowest utility score. At patient level this means that individual values of the QALY in the experimental group are increased with the difference in baseline utilities between the experimental and the control group (table 1, Mean difference adjustment C). We only reported the first mean difference adjustment method (A), as the estimated QALY increments between the experimental and the control group of all the mean difference adjustment methods (2A-2C) are identical (table 1).

3 Delta adjustment

In this adjustment method not the area under the curve but the change in utilities between baseline and follow-up measures is used for the calculation of the QALY. For each individual patient, this so-called delta-QALY is defined as the change in utility and calculated by multiplying the difference between baseline utility and follow-up utility by one half. To enable comparison between the different adjustment strategies, the mean utility at baseline of the complete group is added to this score to calculate the individual QALY. The consequence for the QALY increment between the experimental and the control group using the delta adjustment method is shown in table 1.

124

4 Regression-based utility adjustment

The adjustment for baseline difference based on a regression model uses the follow-up utility as dependent variable (equation 3).

(3)   Regression Utility: $U_1 = \alpha_U + \beta_U U_0$

Where U_1 = follow-up utility

            U_0 = baseline utility

            α_U, β_U = regression coefficients

The regression coefficient (β_U) (equation 3) is used as a correction factor for the follow-up utilities. The individual QALYs and the increment between groups are calculated from the equations described in table 1.

5 Regression-based QALY adjustment

In this method the individual QALY is the dependent variable and baseline utility the independent variable.

(4)   Regression QALY: $QALY = \alpha_Q + \beta_Q U_0$

The QALYs are calculated using the beta (β_Q) (equation 4) from the regression as a correction factor. Equations for the QALY in the experimental and the control group as well as the increment are given in table 1.

125

Table 1: Overview of the formulas for the individual QALY calculations of patients in the experimental and in the control group, and the difference in QALYs between the two groups

Method	QALY of patient in experimental group	QALY of patient in control group	Difference in QALYs
No baseline adjustment (conventional method)	Q_e	Q_c	IQ_{e-c}
Mean difference adjustment A	$Q_e - \frac{1}{4}\Delta_0$	$Q_c + \frac{1}{4}\Delta_0$	$IQ_{e-c} - \frac{1}{2}\Delta_0$
Mean difference adjustment B	Q_e	$Q_c - \frac{1}{2}\Delta_0$	$IQ_{e-c} - \frac{1}{2}\Delta_0$
Mean difference adjustment C	$Q_e + \frac{1}{2}\Delta_0$	Q_c	$IQ_{e-c} - \frac{1}{2}\Delta_0$
Delta adjustment	$Q_e + \bar{U}_0 - \bar{U}_{e0}$	$Q_c + \bar{U}_0 - \bar{U}_{c0}$	$IQ_{e-c} - \Delta_0$
Regression-based utility adjustment	$Q_e + \frac{1}{2}\beta_u (\bar{U}_0 - \bar{U}_{e0})$	$Q_c + \frac{1}{2}\beta_u (\bar{U}_0 - \bar{U}_{c0})$	$IQ_{e-c} - \frac{1}{2}\beta_u \Delta_0$
Regression-based QALY adjustment	$Q_e + \beta_Q (\bar{U}_0 - \bar{U}_{e0})$	$Q_c + \beta_Q (\bar{U}_0 - \bar{U}_{c0})$	$IQ_{e-c} - \beta_Q \Delta_0$

Q_e and Q_e	= QALY in experimental and control group
IQ_{e-c}	= $\frac{1}{2}(U_{e0} + U_{e1}) - \frac{1}{2}(U_{c0} + U_{c1})$
$\bar{U}_0$	= mean of baseline utility of complete group
$\bar{U}_{e0}$ and $\bar{U}_{c0}$	= mean of baseline utility of experimental and control group
Δ_0	= $\bar{U}_{e0} - \bar{U}_{c0}$
β_u	= regression coefficients with utility as outcome
β_Q	= regression coefficients with QALY as outcome

Statistical analysis

Firstly, the mean costs (total hospital costs) and mean effects (using different adjustment methods) of the control group are subtracted from the mean costs and effects of the experimental group, resulting in the incremental cost-effectiveness ratios (ICERs). Secondly, to quantify uncertainty in the estimated ICERs, bootstrap sampling was applied using both the individual cost values and the values of the individual QALY for the different correction methods (Briggs, 1997). Cost-effectiveness planes (CE planes) were constructed after performing 1000 bootstrap replications (figure 1). Cost-effectiveness acceptability curves (CEACs) (figure 2) were used to show the probability of ICER acceptability given the various maximum levels that decision makers may be willing to pay/accept to gain/lose an additional QALY (van Hout, 1994).

126

Results

The mean and standard error for baseline utility score were .71 (.01) and .66 (.01) for respectively the control (n=299) and the experimental group (n=298). The difference was statistically significant

(mean difference = -.05; 95% confidence interval: -.01 to -.09). The follow-up utilities were equal in both groups (mean = .71; standard error = .01). The mean total hospital costs were €5,441 in the control group and €4,625 in the experimental group. The mean total hospital costs were significantly lower € -816 (95% confidence interval: €-1,581 to €-174) in the experimental group.

Table 2: The different ICERs calculated by the means of the different correction methods and percentages bootstrapped ICERs in the four quadrants of the cost-effectiveness planes

	ICER	ICER (Cost/effect)	ICER quadrant	Percentage bootstrapped ICERs			
				NE [a]	NW [b]	SW [c]	SE [d]
No baseline adjustment (conventional method)	€ 29,843	(€-815/ -.030)	Southwest	0	5	85	10
Mean difference adjustment A	€ 273,079	(€-815/ -.000)	Southwest	1	3	52	44
Delta adjustment	€ - 35,512	(€-815/ .020)	Southeast	2	0	3	95
Regression-based utility adjustment	€ 38,678	(€-815/ -.002)	Southwest	1	3	82	14
Regression-based QALY adjustment	€ - 181,173	(€-815/ . 000)	Southeast	1	3	33	63

[a] NE = Northeast quadrant: experimental group more effective and more costly compared to control group, [b] NW = Northwest quadrant: experimental group less effective and more costly compared to control group (inferior), [c] SW = Southwest quadrant: experimental group less effective and less costly compared to control group, [d] SE = Southeast quadrant: experimental group more effective and less costly compared to control group (dominant)

When the conventional QALY calculation method (no baseline adjustment for imbalances of baseline utility) is applied, the point estimate of the ICER (€ 29,843 (€-815/-.030) and most of the bootstrapped ICERs (85%) are situated in the Southwest quadrant of the CE plane, indicating that the experimental intervention is less effective and less costly (table 2, figure 1). In the mean difference method (mean difference adjustment A, table 2, figure 1) approximately 50% of the bootstrapped ICERs is situ-

ated in the Southwest quadrant. 82% of the bootstrapped ICERs of the regression-based utility method is situated in the Southwest quadrant (table 2, figure 1). However, the estimated ICERs of the delta adjustment and the regression-based QALY adjustment (table 2, figure 1) indicate dominance of the experimental treatment, which is confirmed by the bootstrap analyses.

Figure 1: Cost-effectiveness planes with 1000 bootstrap replications of the incremental cost-effectiveness of the experimental versus the control group for no baseline adjustment and the four different adjustment methods

No baseline adjustment (conventional method)
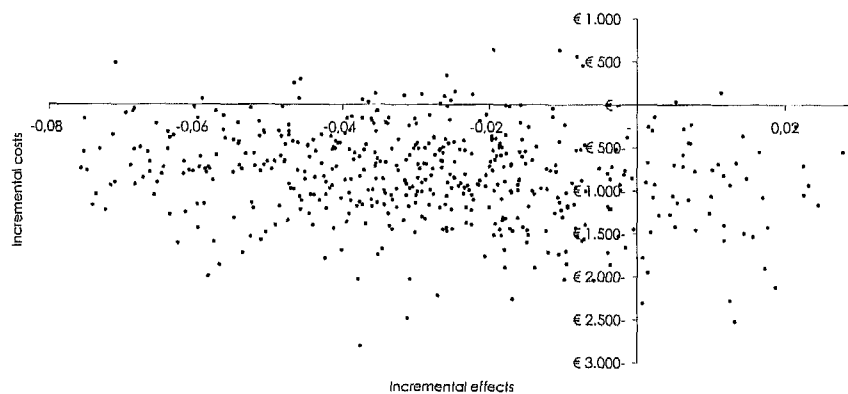
Mean difference adjustment A
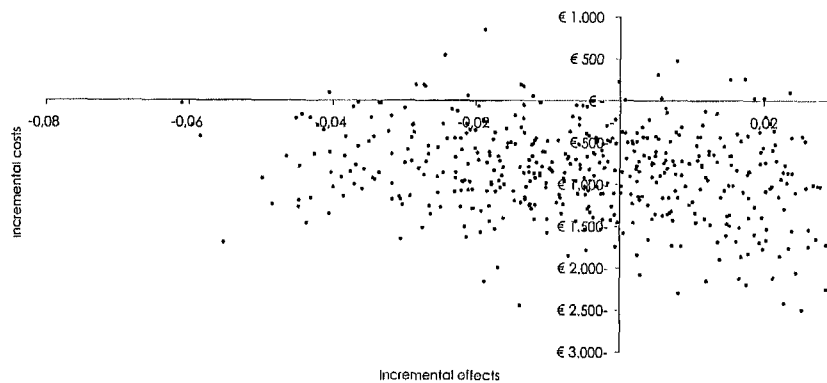
Figure 1: continued

Delta adjustment
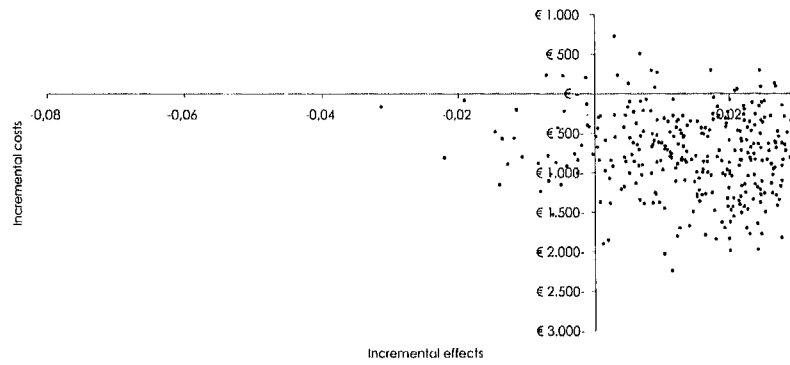
Regression-based utility adjustment
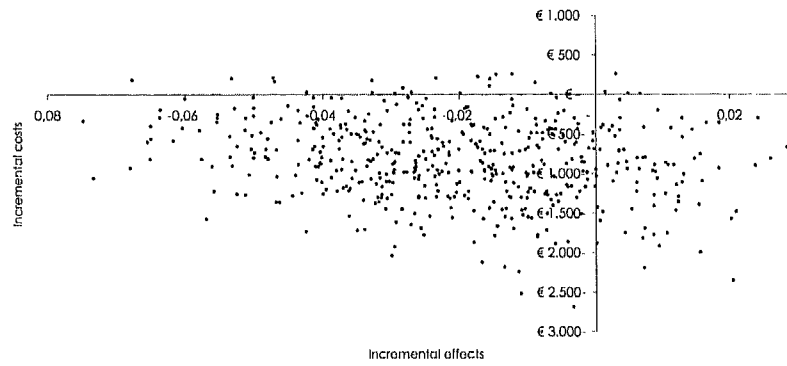
Regression-based QALY adjustment
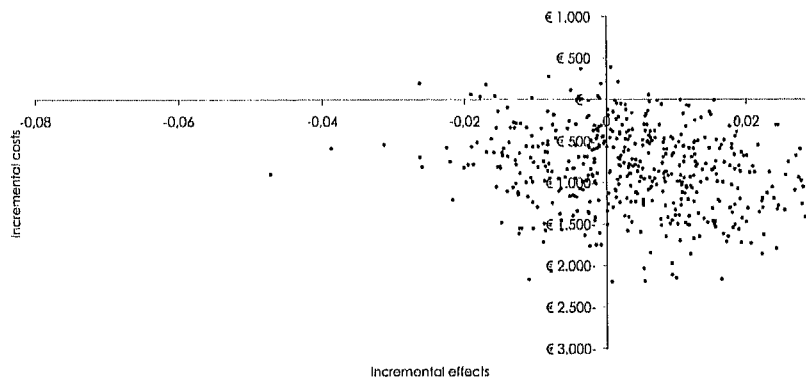
Figure 2 presents the results of the CEACs for the different applied QALY calculation methods. As expected based on the results from the CE planes, the conventional QALY method and the regression-based utility adjustment are represented by the two lowest curves. When choosing a specific ceiling ratio these two curves have a lower probability of acceptance compared to the other higher curves. The delta adjustment showed the highest curve, followed by the regression-based QALY adjustment and the curve of the mean difference adjustment A. All curves except for the highest one (delta adjustment) show a decline of the ICER with an increasing ceiling ratio, indicating that the probability of ICER acceptance is decreasing with an increase of the ICER limit. The CEA-curve of the delta adjustment is the only curve that is not influenced by increased ceiling ratios.

Figure 2: Cost-effectiveness acceptability curves for the cost-effectiveness of the experimental versus the control group with no baseline adjustment and the four adjustment methods
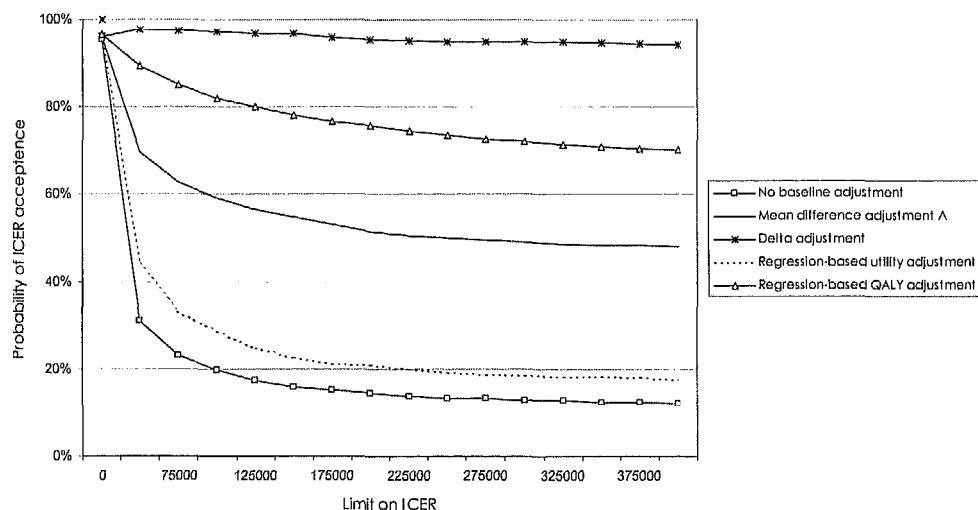
Discussion

Summary of study findings

The results of this study show that not applying a correction method for the calculation of QALYs when baseline utility imbalances between treatment groups occur, resulted in different cost-effectiveness outcomes compared to when various correction methods for the calculation of QALYs were applied. The magnitude of the impact of these correction methods on the study outcome are dependent on the type of correction method used.

Impact of baseline differences

If a randomization procedure in a randomised controlled trial is performed according to the recommendations (is patient really randomly assigned to one of the treatment arms (Altman, 2001)) there is still the possibility of imbalances of baseline characteristics between both treatment arms. It is well accepted to correct for these baseline disparities in prognostic characteristics by means of a covariate-adjusted analysis (Pocock, 2002). In trial based cost-effectiveness studies baseline utility scores can be imbalanced between treatment groups as well. The impact of these imbalances in baseline utility can be substantial as these utilities are directly incorporated in the calculation of the study outcome, the QALY. Corrections of imbalances in baseline utility in trial based cost-effectiveness studies are rarely performed (Richardson & Manca, 2004). For the interpretation of cost-effectiveness studies it is important to report any measured variation of the study outcomes, because only when data at patient level are available, bootstrap analysis can be applied and presentation of results in CEA-curves can be done.

131

Mean difference adjustment

Based on the theoretical advantages and disadvantages, the following considerations can be made for the different adjustment methods proposed. The mean difference adjustment methods (A-C) are not published in the literature. For the incremental effects it makes no difference which of the variants (A-C) are used. Apart from variation due to the bootstrap procedure, they all will result in the same CE planes and CEACs. These mean difference adjustment methods are the least preferred methods for correction as they have two important drawbacks. Firstly, they replace the individual baseline utility scores by group scores. Using this method the mean difference of utilities of the total group or a specific treatment arm at baseline is subtracted or added to the individual scores, which results in a higher percentage of patients that end up having a utility of one or higher (in our case study = 13.6%). As utilities are at a maximum 1, all values over 1 need to be adapted. The second drawback, of the mean difference adjustment method is that this method does not correct for the regression towards the mean phenomenon, because, baseline values negatively correlate with change as patients with low scores at baseline generally improve more than those with high scores (Bland & Altman, 1994).

Delta adjustment

This last disadvantage also applies to the delta adjustment method. This method is a well accepted and commonly used method (Cheng, 2000; Guthrie, 1999; Hurskainen, 2001; Lee, 2002; Patterson, 1995). Another disadvantage of the delta adjustment is that the comparison of the study findings using delta adjustment with other QALY calculations is only possible if the mean baseline utility of the total group is added. Besides these two drawbacks, the advantages of this method are twofold: it is suitable for both small (less then 50 patients) and large study populations, and if

the correlation between baseline and follow-up is high ($r > .8$) the power is still acceptable (Vickers, 2001).

Regression adjustment

Manca et al. were the first to propose a regression based technique to correct for baseline imbalances in utilities when calculating of QALYs (Manca, 2005). However, this method is not suitable for the calculation of individual QALYs and therefore ICER uncertainty cannot be explored. The two regression based correction methods we propose make individual QALY calculation possible. Compared to the previous described correction methods, they have the highest statistical power and do correct for regression towards the mean. There are also disadvantages of these two correction methods. Firstly, the assumptions for applying regression analysis need to be fulfilled. The most important assumption is that the baseline utilities need to be normally distributed. In small studies this is rarely the case. In large studies it is difficult to transform the data because of the wide range of possible health states. The second disadvantage of the regression-based utility adjustment is the possibility of over or under correction, as the 'uncorrected' baseline utility is used to calculate the adjusted QALY.

Study limitations

The generilisability of our findings is questionable. In the reported case study, there is only one follow-up moment. As a consequence, the impact of the baseline difference is larger compared to other studies where there are more follow-up measurements. Besides this, the impact of the different correction methods can vary if the differences in costs between the two treatment groups are less robust than found in the present case study. Therefore, exploring the impact of the different methods for correcting baseline utility imbalances should be repeated using the data of other trial based cost-effectiveness studies.

Recommendations

Researchers should be aware of imbalances in baseline utility in trial based cost-effectiveness studies. If an imbalance is observed, the following approach can be used:

Firstly, be sure that baseline imbalances are the result of coincidences and not due to shortcomings in the performed randomization procedure. When it is expected that there is a problem with the applied randomization procedure, propensity score matching can be used as a correction method (Joffe & Rosenbaum, 1999).

Secondly, always report baseline and follow-up utilities of both groups in mean and variance (standard deviation or standard error).

Thirdly, if baseline imbalances of utilities occur, we propose on theoretical grounds to use the regression-based QALY adjustment method in the base case analysis. Assuming that an adjustment is necessary, methods such as the mean difference and the delta adjustment methods mentioned in the present paper will not adjust for regression to the mean effects, and therefore will probably over adjust for baseline differences. They will also result in an unjustified strong influence of outliers in the baseline measurement. The regression-based QALY method takes into account the regression towards the mean phenomenon and will therefore be the most accurate adjustment method. As the QALY calculation uses the baseline value as such, solely adjusting the follow-up utility solves the problem only partly. Thus, using this regression-based QALY adjustment method will not only correct the follow-up utility values but will also prevent that the original baseline utility values are again used to calculate the QALY.

Furthermore, the impact of the other baseline adjustment methods can be explored as an additional sensitivity analysis. When applying regression based correction methods, check if the assumptions for regression analysis are fulfilled (e.g., normality of baseline utility, n>50). If baseline utilities are not normally distributed, transform data to a log normal distri-

bution. The choice to use one of the mean difference adjustment methods should be based on one of the three methods that lead to the minimum number of unrealistic high utility values. For the interpretation of the results of the mean difference adjustment method and delta adjustment method, always be aware of regression towards the mean phenomenon.

Conclusion

For cost-effectiveness studies with imbalances in the baseline utilities, only reporting study findings using the conventional QALY calculation is not sufficient. We suggest using the regression-based QALY adjustment for dealing with imbalances in baseline utility between treatment groups. Ignoring baseline utility imbalances in QALY calculations can lead to misleading interpretation of cost-effectiveness analysis results. Researchers should always be aware of this problem.

CHAPTER

Summary and general discussion

This chapter contains six major parts, which will summarize and discuss the contents of this thesis. In the first part (1) a summary will be given of the different studies performed. The barriers for implementation of the short-stay intensive care intervention will be discussed in the second part (2). In the third part (3) aspects of framing future research will be discussed. In the fourth part (4) the use of different quality of life measures for future research will be discussed. In the fifth part (5) the impact of using adjustment methods to correct imbalances in baseline utility in QALY calculations on the SSIC study findings will be discussed. In the last two parts (6 and 7) a summary of recommendations for future research will be given.

1 Summary

1.1 State of the art of fast-track treatments for low-risk CABG patients (chapter 2)

In the second chapter the results of a systematic review are described. None of the previous studies that evaluated fast-track treatments focussed on IC discharge criteria. They investigated related topics as type and dose of anesthetics (high dose: fentanyl>15μg/kg, low dose: fentanyl<15μg/kg) or extubation protocols used (the intention to extubate patients within or after 8 hours postoperative). 27 randomised controlled trials evaluating fast-track treatments were identified. The main finding of this review and meta–regression analysis is that the introduction of an early extubation protocol is an important predictor for the decrease of IC and total hospital stay in low-risk CABG patients.

The epidemiological quality of the studies was moderate. Only one study fulfilled all three methodological criteria: the use of intention-to-treat analysis for the primary study outcome IC stay, allocation concealment (is patient really randomly assigned to one of the treatment arms) and blinding of person(s) deciding on the length of IC stay.

The quality of the economic evaluations (EEs) of the 7 studies reporting on costs was assessed by means of the Consensus Health Economics Criteria list (CHEC-list). The CHEC-list scores on the 17 evaluated items of the various studies were low (on average 8 points). The drawbacks of these studies were mainly related to the choice of study design (not appropriate to the stated objective) and costing methods (irrelevant costs were identified and not appropriately valuated). The findings of our review indicate that there is a need for quality improvement of both the study design and the design of EEs.

1.2 Clinical effectiveness and cost-effectiveness of the SSIC protocol (chapter 3)

The short-stay intervention is evaluated in low-risk CABG patients by means of a randomised controlled clinical equivalence trial and an EE performed alongside it. We defined low-risk patients as patients who fulfilled the following criteria: having a stable cardiological situation (not having a cardiogenic shock, no need for inotropic therapy, no need for intraaortic balloon pump, and no ongoing or recent myocardial infarction), age < 78 years, ejection fraction of > 30% and not requiring hemodialysis or having pulmonary hypertension. Patients assigned to the SSIC group were transferred from IC to MC within 8 hours if they fulfilled the following discharge criteria: extubated for at least 30 minutes, normal breathing, arterial blood gas pO_2 > 10 Kpa and PCO_2 < 6.0 Kpa, cardiac stable, fluid balance in control ($\leq$ 100 ml chest tube drainage per hour and a diuresis > 0.5 ml/kg/hour), no signs/symptoms of neurological complications, no need for hemodynamic and supporting therapy.

The primary study endpoints were IC readmission and total hospital stay. IC readmission was 1.34% (n=4) and 2.68% (n=8) in the control group and the SSIC group respectively. The difference of 1.13% was not significantly different (P-value = 95% CI: -.9% to 2.9%). No differences were found between the two groups with respect to 'total hospital stay' and the sec-

ondary outcomes measures 'postoperative morbidity' and '30 day-mortality'. The patients in the SSIC group spent significantly fewer hours at the IC compared to control group patients. In the intervention group the mean costs per patient were €4,625 versus €5,441 in the control group. The difference was €-816, mainly due to lower costs of clinical procedures and IC stay. The generic quality of life was valuated by societal health values. The delta quality adjusted life months (Δ QALM) was significantly different between the two groups, in favour of the SSIC group. The incremental cost-effectiveness ratio and 98% of the bootstrapped ICERs were situated in the Southeast quadrant, indicating that the SSIC group is dominant over the control group. The conclusion of this study is that SSIC is a safe and cost-effective approach compared to usual care.

1.3 Health-related quality of life after fast-track treatment (chapter 4)

The objective of this prospective randomised equivalence trial was to evaluate disease specific (Multidimensional Index of Life Quality), generic (EQ-5D) and domain specific (Beck Depression Inventory, State-Trait Anxiety Inventory) health-related quality of life (QoL) until one year after fast-track treatment for low-risk coronary artery bypass patients (CABG patients). Quality of life was measured at baseline, one and twelve months after treatment. The main findings were that in a multivariate analysis at one year after surgery no statistically significant differences in QoL between the two groups were shown. In addition, after controlling for potential confounders at one year no statistical significance between both groups was found. The conclusion of this study is that QoL of SSIC patients are similar to patients receiving care as usual, and that none of the observed QoL differences between the two groups can be defined as clinically meaningful.

140

1.4 Testing validity and reliability of the EuroQoL and Multidimensional Index of Life Quality (chapter 5)

The objective of this prospective comparative study was to test the validity and reliability of the EuroQoL (EQ-5D) and the Multidimensional Index of Life Quality (MILQ, a disease specific instrument) in coronary artery bypass surgery patients.

The main findings of this comparative study were that the discriminative validity was limited due to substantial ceiling effects in all five EQ-5D dimensions shown. These effects were also observed in the MILQ dimensions financial status and healthcare professionals. The predictive validity at one month of the EQ-5D domain anxiety/depression, EQ-5D VAS, MILQ domain physical health and the MILQ-score was moderate. At 12 months after surgery the EQ-5D VAS, MILQ-score and the two physical dimensions of the MILQ had a good predictive validity. Only the EQ-5D domain usual activities and the EQ-5D VAS distinguished between patients of different New York Heart Association functional classification categories (external criterion validity). The test-retest reliability of both questionnaires can be classified as not adequate.

The conclusions of this study are that the discriminative validity of the EQ-5D domains, the external criterion and the predictive validity of both the EQ-5D and the MILQ were limited. In addition the reliability of both questionnaires was not adequate. These questionnaires are not the first choice when evaluating QoL in CABG patients.

1.5 Adjustment methods for imbalances in baseline utility between treatment groups in QALY calculations (chapter 6)

The objective of this study was to evaluate different strategies, which correct imbalances in baseline utility in quality adjusted life year (QALY) calculations and are suited for the analysis of data on a patient level.

The data of a trial based cost-effectiveness study evaluating a short-stay intensive care treatment for coronary artery bypass patients was used as

141

a case study. Five different methods were used for calculation of QALY differences: 1) Conventional QALY calculation method (no adjustment for imbalances in baseline utility); 2) Mean difference adjustment method: correction with the differences in the mean baseline utility of the treatment groups; 3) Delta adjustment: uses the change in utility over time; 4) Regression-based utility adjustment: adjusting the follow-up utility with a regression model with the follow-up utility as dependent variable, and baseline utility as independent variable; 5) Regression-based QALY adjustment: adjusting the QALY with a regression model with the QALY as dependent variable, and baseline utility as independent variable.

The results of this study show that not applying a correction method for the calculation of QALYs when imbalances in baseline utility occur, results in a different cost-effectiveness outcome compared to various simple or advanced correction methods for the calculation of QALYs having been applied. The magnitude of the impact of these correction methods on the study outcome is dependent on the type of correction method used. We recommend in case of imbalances in baseline utility to report the outcomes of the regression-based QALY adjustment in the base case analysis and to report the outcomes of the other proposed adjustment methods in additional sensitivity analyses.

2 Possibilities for implementation of the current SSIC protocol

The encouraging results of the SSIC study make it desirable to implement the SSIC protocol. We know that in the Netherlands hospitals use fast-track protocols; at the moment, however, more than three years after publication of the main study findings, we did not get any information about hospitals currently using our protocol.

One may ask if the findings of one randomised controlled trial and an EE performed alongside it are sufficient evidence to policy makers to actually decide to use the SSIC protocol in their hospitals? This study demonstrates the second-best level of evidence (level 1B): an individual ran-

domised controlled trial with narrow confidence intervals. In other words an randomised controlled trial which excludes clinical harm from the new treatment (Straus, 2005). The best level of evidence is not available as this was the first study that evaluated this type of fast-track treatment (SSIC-protocol) and therefore no systematic review with homogeneity of randomised controlled trials could be performed (Level 1A).

Reliability and transferability of the study findings will be discussed below as they are also important to decision makers in making decisions on whether or not to implement a new treatment strategy (Drummond & Pang, 2001; Mason & Mason, 2006). Decision makers need to be convinced of the reliability of the study findings, in other words; does the study give an accurate estimate of the study outcomes, which are free of biases? This may apply to the main effect measure of the EE, the QALY. Several surveys indicate that policymakers are less comfortable with the methods of QALY calculations. They find the concepts behind the measurements difficult to understand, they have concerns about the reliability of the estimation methods themselves, and finally, they have a general concern about the aggregation of healthcare benefits in a single index (Barbieri, 2005). The transferability of a study means, to which extent study findings of a specific study are directly applicable to the decision maker's own setting. This is not always the case with cost-effectiveness findings, as they can vary by settings, by levels of clinical experience, by disease severity, by co-morbidity, by resource use, by prices and by utilities (Manca, 2005; Mason & Mason, 2006; Sculpher, 2006). This may be a reason for decision makers not to implement the SSIC protocol. However, not implementing the protocol for transferability reasons was minimized by thoroughly and accurately describing the study design, conduct, and analysis in the published articles. By doing so, policy makers were able to distillate which part of the findings do or do not apply to their own specific healthcare setting.

143

An important step in the implementation process is to determine various potential barriers that impede implementation (Grol & Wensing, 2001). In the following sections the potential barriers of implementation of the SSIC protocol related to the system (infrastructure and budgets (2.1.1-2.1.2)) and the social environment (healthcare professionals and patients) will be discussed (2.1.3-2.1.4).

2.1.1 System: Infrastructure of the IC and MC

Cardio Thoracic (CT) ICs in the Netherlands can have different structures: being a separated CT IC, the CT unit is situated on the general IC, and a combined CT-cardiology IC. An MC care facility is not always available and therefore patients sometimes need to be directly transferred to the nursing ward. If centers do not have a separated CT IC nor a MC directly attached to the IC, this may pose a barrier to the implementation of the SSIC protocol, and, subsequently, to patient logistics.

2.1.2 System: Budgets

From a health economic point of view the available resources as a result of the reduced IC hours and clinical procedures can be used for the treatment of more complex patients. Another possibility of a more efficient use of IC and MC capacity can be that more surgeries can be performed and fewer patients are being refused for IC admission. However, although decision makers will understand these abovementioned arguments their main concern is the financial budget constraints. Currently the reimbursement per treated patient is not related to the total stay at the various departments, but to all aspects of a specific clinical pathway (DBC; Dutch for Diagnosis treatment Combination). Besides, another important aspect for not implementing the SSIC protocol will be the costs of the actual implementation of the intervention itself. These are for instance the costs of rebuilding the IC and/or MC departments and the costs of education and organizing meetings for healthcare professionals.

144

2.1.3 Social environment: healthcare professionals

Due to the fact that it is not possible to predict exactly at what time of the day which and how many SSIC patients will be ready to be transferred to medium care, the planning of personnel can be a barrier to healthcare professionals to implement the protocol. In our study 55% of the SSIC patients fulfilled the discharge criteria and were transferred from IC to MC within 8 hours. As the decision to transfer SSIC patients can be made not only at 8 hours but at any time a patient has a stable clinical situation (for instance at 6 or 12 hrs after surgery), these percentages will probably be higher in daily practice. Further research is needed to confirm these expectations.

Another barrier related to healthcare professionals can be the type of profession of the general manager of the IC department. He or she can decide whether or not to implement the SSIC protocol. For example, the development of the SSIC protocol by a cardiac surgeon might affect the acceptability for implementation by other professions, like intensivists or anesthesiologists who can also be the general managers of the IC.

Other potential barriers might be the attitude and resistance to change felt by IC and MC nursing staff. During the Maastricht trial e.g. nurses were very sceptical about the benefits of treating patients according to the SSIC protocol, as they feared that their workload would increase (for example, patients are not constantly asleep any more during their stay at the IC department).

2.1.4 Social environment: patients

The following data indicate that the motivation for implementing a SSIC protocol specially developed for low-risk CABG patients may have decreased in recent years due to changes in the CABG-patient population. From 2004 to 2006 a decline of the number of hospitalizations due to CABG and an increase of the number of hospital admissions for percutaneous coronary interventions (like stent placement) was observed in

145

the Netherlands (Prismant, 2008). A comparable trend was also observed in other western countries (Nallamothu, 2007). These results, together with ageing, resulted in a greater proportion of higher risk cardiac patients who were selected for elective surgery (Cheng & Barash, 2006; Flynn, 2004; Goss, 2006).

3 Framing a future study

As discussed, implementation of the SSIC protocol is difficult for various reasons. Therefore it is important that future studies in this field must diminish the implementation barriers by taking them into account as much as possible in the design phase of the study. The development of a new EE in this field will be discussed by means of the concept of framing. Framing can be defined as a series of decisions that collectively define and describe the EE study to be undertaken (Gold MR, 1996). The main aspects of framing are; defining the study objective and audience (3.1), the type of analysis (3.2), the perspective of the analysis (3.3), the intervention and the comparison treatment and the target population (3.4), and finally the boundaries and the time horizon (3.5).

3.1 Objective and audience

In our opinion the objective of a future study must be: performing a before-after multicenter trial with EE evaluating the entire clinical pathway of CABG patients. In the first 12 months before evaluating the clinical pathway, all relevant clinical outcomes, quality of life and cost data need to be gathered for the care as usual in various hospitals. In the second period of 12 months the same data will be collected when treating patients according to the clinical pathway.

The primary audience of this new study is the general managers of the IC units and the board of directors of the hospitals. Therefore it is important to focus the analysis on topics that are relevant for these two groups of decision makers. For instance, a future study should focus more on an

intervention that can easily be implemented in different hospital settings. Besides the costs and consequences that are part of a conventional EE, other aspects that are important to decision making should be assessed as well. Issues that need to be included are for example: calculations of the implementation costs, and scenario analyses on the impact of the new treatment on changes in patient flow and waiting lists.

3.2 Type of analysis

As it is expected that the clinical pathway of CABG patients is more effective compared to the usual care, the preferred type of analysis of a future study is CEA (or CUA). However, given the idea described above that other aspects are relevant to decision makers, the framework of cost-outcome analysis, being a specific method of EE, might be considered for this purpose to cover the outcomes that cannot be expressed as either costs or patient outcome.

3.3 Perspective of analysis

For a future study we would advise to choose both the hospital and societal perspective for the EE. A hospital perspective should be used, because at this level the decisions about the use of the clinical pathway of CABG patients will be made. A societal perspective is also important for the foundation of arguments for implementation. Due to the proposed changes in treatment it is expected that treatment according to the clinical pathway will have a positive effect on recovery (like reduced overall hospital stay, higher scores on physical aspects of quality of life) compared to care as usual. Patients treated according to the clinical pathway for instance will be discharged earlier from hospital after surgery. In addition, cost-savings will probably take place. The direct medical costs, like the hospital costs (costs of inpatient days and clinical procedures) will probably decrease, but some indirect medical costs will probably increase (for instance more informal care or paid help after

147

hospital discharge will be needed). The indirect non-medical costs, however, will be lower as it is expected that the intervention patients restart normal and daily life sooner (shorter absence from (volunteer) work) compared to patients who are not treated according to the new clinical pathway.

3.4 Defining the intervention, comparison treatment and target population

Details of the clinical pathway must be clearly specified in a future study. This means that not only the intervention on the different departments (IC, MC and ward) needs to be described in detail but also the care as usual. In the description of the IC protocol details of early extubation, frequencies of regular checks, temperature management, sedation and pain management are important. The difference between this future study and the performed SSIC study is that not only must it focus on the evaluation of an intervention in the first hours of intensive care treatment, but on all aspects of the pre-, peri- and postoperative care of CABG patients. An important aspect of this new study should be that patients who fulfill the IC discharge criteria are immediately transferred to a lower care facility (MC or ward). In addition not only discharge after 8 hours IC treatment must be considered as an option for transferring patients, but other time points in a range of 6 to 12 hours should also be included. These protocols must also describe details of, for instance, the preoperative screening day, temperature management during surgery, anesthetic doses, mobilization of patients by physiotherapist and nurses. As discussed in the implementation section the care as usual can vary between different hospitals and settings. It is important to compare, if possible, the new clinical pathway with the most important comparators of care, which are the most realistic policy choices.

The target population of a future study should contain both low-risk and high-risk CABG patients and the eligibility criteria need to be based on current standards of risk stratification (Michel, 2003; Nilsson, 2006).

148

3.5 Boundaries and time horizon of the economic evaluation

The scope of a future study should be the evaluation of the clinical effectiveness and cost-effectiveness of the clinical pathway for all CABG patients. Aspects that are not primarily related to the abovementioned scope, like the influences on quality of life of family members and costs made by those relatives due to illness of the patients under study, should not be incorporated in the analysis.

In a future study the follow-up must be until one year after treatment, for all relevant study outcomes: clinical outcomes, quality of life and costs. One year follow-up can be chosen as the time frame for this study as it is a well accepted follow-up for rehabilitation after CABG surgery (Badia, 2001; Borkon, 2002; Cheng & Barash, 2006).

4 Choosing quality of life measures for a future study

4.1 Generic quality of life measures

In many ways it is important to know the possible effects of a new treatment on quality of life. One of its merits is that it can help clinicians to explain to patients what they can expect from a specific treatment. Besides, if more than one therapeutic option is available, this knowledge -in addition too clinical and cost-effectiveness- contributes to making rational choices concerning the most appropriate therapy. As the measurement of QoL is subjective it is very important to use validated and reliable instruments. The EQ-5D was chosen in the SSIC study as generic measure for several reasons. Firstly, it is has good psychometric proportions (Brazier, 1993). Secondly, it is brief, simple and easy to administer (only 6 items). In addition it is suitable for EE as it can generate a single index score for health (utility) by means of population weights (Dolan, 1997). Besides using it for the EE (chapter 3) we wanted to use the same instrument for describing patients' quality of life in the five different dimensions, so we could use this in the QoL-study (chapter 4).

149

Based on our findings of the validity and reliability study (chapter 5) it may be better not to choose the EQ-5D but another generic instrument for the evaluation of QoL in evaluation study of the clinical pathway for CABG patients. There are several other preference-based measures that can be considered. The other questionnaires most commonly used with the same possibilities as the EQ-5D are the Short Form-36 (SF-36) and the Health Utilities Index (HUI). The HUI contains 15 questions and classifies patients in either HUI2 or HUI3 health states. In our opinion this questionnaire contains attributes that are not relevant for the evaluation of the clinical pathway: sensation and fertility in the HUI2; vision, hearing, speech and fertility in the HUI3.

Should we then use the SF-36 (and therefore the SF-6D) for a future study instead of EQ-5D? A theoretical advantage of the SF-6D is that it contains a larger descriptive system (i.e. 18.000 unique health states can be described by SF-6D compared to only 243 by EQ-5D), therefore it potentially has greater ability to identify small changes (Bryan & Longworth, 2005). Another advantage of using the SF-6D is its better ability to detect improvements in the upper range of the utility scale in cardiovascular disorders (Moock & Kohlmann, 2008). A disadvantage of the SF-6D is that it does not describe health states at the lower end of the scale (Longworth & Bryan, 2003). Based on a head-to-head comparison of EQ-5D and SF-6D in patients with coronary heart disease Van Stel concluded there is no clear benefit of using SF-6D in clinical studies instead of EQ-5D (van Stel & Buskens, 2006). Others do not prefer one to the other (Kopec & Willison, 2003) either as both questionnaires have several disadvantages. Some authors found no sensitivity to change of SF-6D after the intervention (Longworth & Bryan, 2003; Smith, 2000; van Stel & Buskens, 2006). We were also not able to demonstrate sensitivity to change for EQ-5D (chapter 5). In our opinion based on the above-mentioned advantages and disadvantages and the results of our study (chapter 5) either SF-6D or EQ-5D

should be chosen for the measurement of generic quality of life in a future study evaluating the entire clinical pathway for CABG patients.

4.2 Disease specific quality of life measures

For the assessment of QoL of people with ischemic heart disease it is recommended to use a disease specific QoL questionnaire (Dempster & Donnelly, 2000) in addition to a generic questionnaire. In the QoL-study (chapter 4) we used the Multidimensional index of quality of life (MILQ) as disease specific instrument. There were two reasons why we chose this questionnaire in our evaluation study: a Dutch version of this questionnaire was already available, and both psychometric proportions (validity and reliability) were extensively examined by Avis et al. (Avis, 1996). The conclusion of these authors was that this instrument was valid and reliable, contrary to our study findings (chapter 5).

At the moment there are various other options for the evaluation of disease specific QoL of the clinical pathway of CABG patients. The two most commonly used are the Seattle Angina Questionnaire (SAQ) (Dougherty, 1998; Spertus, 1995) and the MacNew Heart Disease Questionnaire (MacNew) (Lim, 1993). The SAQ consists of 19 items, which are grouped in five separate domains: physical limitation, angina stability, angina frequency, treatment satisfaction and disease perception. The MacNew consists of 27 items, which fall into the following three domains: physical limitations, emotional function and social function. All domains of the MacNew and SAQ are responsive and their test-retest reliability can also be classified as high (Dougherty, 1998; Hofer, 2003; Hofer, 2004; Spertus, 1994; Spertus, 1995). Furthermore, it was found that none of the SAQ domains and most of the MacNew domains are able to discriminate between patients in Canadian Cardiovascular Society functional classes I-IV (Hofer, 2003). Dougherty et al., however, found that all the SAQ domains except for treatment satisfaction were able to discriminate between Canadian Cardiovascular Society functional classes I-III. Based on previous

151

findings we can conclude that the psychometric proportions of both questionnaires can be classified as good. However, none of the SAQ and only a few of the studies evaluating the reliability and validity of the MacNew are performed in the Netherlands. In addition to this the study population consisted mostly of patients with a stable angina and not of a population of CABG patients. Therefore in our opinion further research is needed to either prefer the SAQ to the MacNew for the measurement of disease specific QoL in a future study evaluating the clinical pathway for CABG patients.

5 Effects of using correction methods for imbalances in baseline utility in QALY-calculations on SSIC study findings

In our EE (chapter 2) we corrected for imbalances in baseline utility by means of the delta method. The conclusion of the EE based on using this correction method was that the SSIC was a cost-effective treatment. Although this correction method is generally accepted (Cheng, 2000; Guthrie, 1999; Hurskainen, 2001; Lee, 2002; Patterson, 1995), it has an important drawback: it does not correct for regression towards the mean phenomenon. In an additional study (chapter 6) we investigated the impact of different correction methods and concluded that the regression-based QALY adjustment was the preferred correction method. If we calculate the ICERs using the regression-based QALY adjustment, this method showed more uncertainty about the study findings of the EE compared to calculation of the ICERs using the delta adjustment method. When using this correction method after bootstrapping, 98% of the ICERs were situated in the Southeast Quadrant compared to 63% using regression-based QALY adjustment. Can we therefore still say that the SSIC is a cost-effective treatment? Yes, because the point estimates indicate dominance of the SSIC treatment; there is, however, more uncertainty about this, as 33% of the ICERs lay in the Southwest quadrant, indicating

that the SSIC intervention is cost saving but not more effective compared to care as usual.

6 Summary of recommendations

Based on the evidence of this thesis we would recommend:

1. a before-after multicenter trial that evaluates the entire clinical pathway instead of a randomised controlled trial in a single center with the focus on an intervention that only evaluates the first eight hours of intensive care treatment as study design;
2. not only a CEA as analysis design, but also outcomes that are more relevant to the primary audience, like costs of implementation;
3. not only a hospital, but also a societal perspective;
4. to include both low and high-risk CABG patients instead of low-risk on a long-term follow-up (of one year) for all study outcomes instead of the one month follow-up for clinical outcomes and costs used in the SSIC study;
5. the SF-6D or EQ-5D for the measurement of generic QoL;
6. the SAQ or MacNew instead of the MILQ for the measurement of disease specific QoL;
7. the regression-based QALY adjustment method instead of the delta adjustment method, if baseline imbalances in utility occur.

In conclusion, all the above listed recommendations are important for future research of studies evaluating new treatments in CABG patients, the last recommendation, however, applies to cost-effectiveness analyses in general, too.

7 Practical recommendations for future research

In this last part some practical recommendations for future research will be given based on the experience of performing the SSIC study.

153

7.1 Information to healthcare professionals

- Organize a kick off meeting at the start of the study for all health-care professionals.

- Keep the study protocol, summary of treatment protocol, and telephone list available and up-to-date at a central place.

- Inform all healthcare professionals once a month about study progress, and immediately about protocol changes. Do not forget to inform new employees.

7.2 Treatment evaluation

- Perform multidisciplinary meetings once a week to evaluate the treatment of control and intervention patients. Record all main findings and issues.

- Give feedback on treatment issues and applications of case record forms to healthcare professionals as soon as possible.

7.3 Informed consent procedure and patient guidance

- Inform patients on the day of admission in the morning hours in an individual approach.

- Ask them to fill in questionnaires and let them write down questions concerning the study. Collect the questionnaires in the afternoon, and check missing items therefore.

- Visit patients frequently during hospital stay. Visit patients also at the day of discharge and inform them about their follow-up (e.g. questionnaires and hospital visits).

CHAPTER

8

Samenvatting

In dit hoofdstuk worden de belangrijkste bevindingen van dit proefschrift samengevat. In het eerste gedeelte van dit hoofdstuk wordt een samenvatting van de verschillende studies gegeven (1). De barrières voor implementatie van de short-stay intensive-care interventie worden besproken in het tweede gedeelte (2). In de laatste twee gedeeltes (3 en 4) wordt een samenvatting van de aanbevelingen voor verder onderzoek gegeven.

1 Samenvatting van de studies in het proefschrift

1.1 Huidige kennis met betrekking tot fast-track behandeling voor laagrisico CABG-patiënten (hoofdstuk 2)

In het tweede hoofdstuk van dit proefschrift worden de resultaten van een systematische literatuurstudie beschreven. 27 gerandomiseerde, gecontroleerde studies die fast-track behandelingen evalueerden werden gevonden in de MEDLINE en in de Cochrane databestanden. Omdat de verschillende studies heterogeen waren wat betreft de onderzochte interventies is een metaregressie analyse uitgevoerd. Hierbij zijn de studies ingedeeld op grond van belangrijke kenmerken van fast-track behandelingen: type en dosis anestheticum, extubatietijd en temperatuurmanagement. De belangrijkste bevinding van deze metaregressie analyse is dat de introductie van een protocol dat gericht is op snelle extubatie een belangrijke voorspeller is voor de afname van het totaal aantal uren intensive-care verblijf en ziekenhuisverblijf bij laagrisico CABG-patiënten.

In dit hoofdstuk zijn ook de verschillende studies onderzocht op hun methodologische kwaliteit. De epidemiologische kwaliteit van de 27 studies was redelijk. Eén studie voldeed aan alle drie methodologische criteria: het gebruik van intention-to-treat analyse voor de primaire studie uitkomst (intensive-care verblijf), allocation concealment (het geheim houden of blinderen van de toewijzing van patiënten aan de verschillende onderzoeksgroepen) en blindering van de personen die de lengte van

158

de intensive-care duur bepalen. De kwaliteit van de economische evaluaties van de studies is onderzocht aan de hand van de Consensus Health Economics Criteria list (CHEC-list). The CHEC-list scores van de 17 gescoorde items van de studies was laag (gemiddeld 8 punten). Dit was met name te wijten aan een verkeerde keuze van het onderzoeksdesign en verkeerd gebruik van methoden om de kosten te berekenen. De bevindingen van deze literatuurstudie laten zien dat er een noodzaak is om zowel de kwaliteit van de initiële studies alsmede die van de economische evaluaties te verbeteren.

1.2 Klinische effectiviteit en kosteneffectiviteit van het short-stay intensive care protocol (hoofdstuk 3)

Het effect van een short-stay intensive-care verblijf is onderzocht bij zeshonderd laagrisico CABG-patiënten door middel van een gerandomiseerde, gecontroleerde klinische equivalentie trial en een hieraan gekoppelde economische evaluatie. Patiënten die een laagrisico hadden op complicaties mochten meedoen in deze studie. Inclusie criteria waren bijvoorbeeld: jonger dan 78 jaar, een ejectiefractie van meer dan 30%, geen dialyse nodig hebben, geen pulmonaire hypertensie. De cardiologische toestand van de patiënten moest eveneens stabiel zijn. Patienten ingedeeld in de SSIC-groep werden binnen 8 uur overgeplaatst van de intensive care naar de medium care als ze voldeden aan de volgende ontslagcriteria: extubatie gedurende minstens 30 minuten, normale ademhaling, arteriële bloed gassen $Pao_2 > 10$ kPa and $Paco_2 < 6.0$ kPa, cardiologisch stabiel, vochtbalans onder controle (≤ 100 ml thorax drain per uur en een diurese > 0.5 ml/kg/uur), geen signalen en symptomen van neurologische complicaties en geen noodzaak voor hemodynamische en ondersteunende therapie.

De primaire studie uitkomsten waren het aantal intensive-care heropnames en de totale duur van de ziekenhuisopname. Vier controlegroep patiënten (1.34%) en acht (2.68%) SSIC-patiënten werden heropgeno-

159

men op de intensive-care afdeling. Het verschil van 1.13% tussen beide groepen was niet significant verschillend (P-waarde = .241; 95% betrouwbaarheidsinterval: -.9% tot 2.9%). Er werd ook geen verschil gevonden in de twee groepen met betrekking tot de totale duur van de ziekenhuisopname en de andere secundaire uitkomsten postoperatieve morbiditeit en 30 dagen mortaliteit. De patiënten in de SSIC-groep verbleven significant minder uren op de intensive care vergeleken met de controlegroep patiënten. In de interventiegroep waren de totale kosten gemiddeld €816,- lager vergeleken met de gemiddelde kosten van de patiënten in de controlegroep. De kostenbesparing in de SSIC-groep werd voornamelijk veroorzaakt door lagere kosten van klinische verrichtingen en intensive- care verblijf. De kwaliteit van leven uitgedrukt in - delta van voor kwaliteit van leven gecorrigeerde levensmaanden (Δ QALM)- was significant verschillend tussen beide groepen, en hoger voor de SSIC-patiënten. De incrementele kosteneffectiviteitsratio en de 98% van de gebootstrapte ICERs lagen in het zuidwest-kwadrant. Dit geeft aan dat de SSIC-interventie dominant is ten opzichte van de standaard zorg. De conclusie van deze studie is dat de SSIC een veilige en kosteneffectieve behandeling is vergeleken met de huidige standaard behandeling (24-uur intensive-care zorg).

1.3 Kwaliteit van leven na een fast-track behandeling (hoofdstuk 4)

Het doel van deze gerandomiseerde, gecontroleerde equivalentie studie was het evalueren van ziekte specifieke, algemene en domein specifieke kwaliteit van leven tot een jaar na de short-stay intensive-care behandeling. Ziekte specifieke kwaliteit van leven werd gemeten met behulp van de Multidimensional Index of Life Quality, de algemene kwaliteit van leven door middel van de EQ-5D en de domein specifieke door middel van de Beck Depression Inventory en State-Trait Anxiety Inventory. De metingen werden verricht op de dag vóór de operatie en op één en twaalf maanden erna. De belangrijkste bevindingen van deze

160

studie zijn dat na een multivariate analyse op één jaar na de operatie geen significant verschil in kwaliteit van leven tussen SSIC-patiënten en controlegroep patiënten waar te nemen is. Ook na de correctie van mogelijke verstorende factoren (zoals bijvoorbeeld co-morbiditeiten en leeftijd) was er op één jaar na de operatie geen statistisch significant verschil tussen beide groepen waargenomen. De conclusie van deze studie is dat op het gebied van de kwaliteit van leven geen van de waargenomen verschillen tussen beide behandelingsgroepen klinisch relevant verschillend zijn.

1.4 Het testen van de validiteit en de betrouwbaarheid van de EuroQoL en Multidimensional Index of Life Quality (hoofdstuk 5)

Het doel van deze prospectieve studie was het testen van de validiteit en de betrouwbaarheid van de EuroQoL (EQ-5D) en de Multidimensional Index of Life Quality (MILQ) in CABG-patiënten.

De belangrijkste bevindingen van deze studie waren dat de discriminatieve validiteit in de vijf EQ-5D dimensies werd beperkt door plafondeffecten (het percentage patiënten met een maximale score). Deze effecten traden ook op bij twee van de negen MILQ-domeinen. De predictieve validiteit op één maand na de operatie van het EQ-5D domein angst/depressie, EQ-5D VAS, MILQ-domein fysieke gezondheid en de MILQ-score waren redelijk. Op 12 maanden na de operatie hadden de EQ-5D VAS, de MILQ-score en de twee fysieke dimensies van de MILQ een goede predictieve validiteit. Alleen het EQ-5D domein dagelijkse activiteiten en de EQ-5D VAS kunnen onderscheid maken tussen patiënten ingedeeld in de verschillende New York Heart Association functional classification categorieën (externe criterium validiteit). De test-hertest betrouwbaarheid van beide vragenlijsten kan worden geclassificeerd als onvoldoende.

De conclusies van deze studie zijn dat de discriminatieve validiteit van de EQ-5D domeinen, de externe criterium validiteit en de predictieve

161

validiteit van zowel de EQ-5D en de MILQ beperkt zijn. Daarnaast is de betrouwbaarheid van beide vragenlijsten niet toereikend. Op basis van de bevindingen van onze studie kunnen we concluderen dat beide vragenlijsten niet de voorkeur hebben bij het evalueren van de kwaliteit van leven van CABG-patiënten.

1.5 Correctiemethoden voor baseline utiliteitsverschillen tussen twee behandelingsstrategieën in voor kwaliteit van leven gecorrigeerde levensjaar berekeningen (hoofdstuk 6)

Het doel van deze laatste studie was het evalueren van verschillende strategieën die kunnen corrigeren voor baseline utiliteit verschillen in voor kwaliteit van leven gecorrigeerde levensjaar (QALY) berekeningen zodat er analyses kunnen worden gedaan op patiëntniveau.

De data van de trial die de short-stay intensive-care behandeling voor CABG-patiënten onderzocht, werd gebruikt als case studie. Vijf verschillende methoden werden toegepast: 1) de conventionele QALY berekening, waarbij geen baseline correctie plaatsvindt; 2) Correctie met de verschillen tussen de gemiddelde baseline utiliteit van beide behandelingsgroepen; 3) De Delta correctiemethode, waarbij alleen de veranderingen in de utiliteit tussen beide metingen wordt gebruikt; 4) een op regressie gebaseerde utiliteitscorrectie, waarbij een regressie-model wordt gebruikt met daarin de follow-up utiliteit als de afhankelijke variabele en de baseline utiliteit als de onafhankelijke variabele; 5) een op regressie gebaseerde QALY correctie, waarbij de QALY de afhankelijke en de baseline utiliteit de onafhankelijke variabele is in het regressie-model.

De resultaten van deze studie laten zien dat het niet-gebruiken van een correctiemethode wanneer er wel baseline utiliteitsverschillen optreden, resulteert in verschillende kosteneffectiviteit uitkomsten. De mate van de invloed van de correctiemethoden is afhankelijk van welk type correctiemethode er wordt gebruikt. Op basis van de resultaten van deze studie wordt aanbevolen dat wanneer er een baseline verschil in de utiliteit

wordt waargenomen deze het beste kan worden gecorrigeerd met behulp van de regressie-gebaseerde QALY correctie. De uitkomsten van de andere correctiemethoden kunnen worden gerapporteerd in aanvullende sensitiviteitsanalyses.

2 Mogelijkheden voor implementatie van het huidige short-stay intensive care protocol

De bevindingen van de SSIC-studie laten zien dat de resultaten aanleiding zouden kunnen zijn om het SSIC-protocol te implementeren in de Nederlandse ziekenhuizen. Echter meer dan drie jaar na de publicatie van de belangrijkste bevindingen van de SSIC-studie hebben we echter geen informatie dat ziekenhuizen op dit moment het SSIC- protocol gebruiken.

De redenen waarom dit tot op heden nog niet gebeurd is, worden uitgebreid besproken in hoofdstuk 7, paragraaf 2. Een van de redenen die kan worden geopperd, is dat de implementatie op basis van één onderzoek niet voldoende reden is voor beleidsmakers om volgens het SSIC-protocol te gaan werken. Daarnaast zijn er een heel aantal potentiële barrières te bedenken die implementatie van het SSIC-protocol in de huidige vorm in de weg kunnen staan. Zoals de infrastructuur van de verschillende intensive en medium care afdelingen in de diverse cardio-thorale centra in Nederland. Als bijvoorbeeld CABG-patiënten op gemengde IC's liggen, heeft dit direct gevolgen voor de patiëntenlogistiek. Een ander voorbeeld van een barrière kan zijn dat er hoge kosten moeten worden gemaakt voor het verbouwen van de IC en MC afdelingen als er geen aparte MC aanwezig is. Een belangrijke andere barrière is dat het met het huidige protocol niet mogelijk is om exact te bepalen hoeveel patiënten wanneer op welke afdeling liggen, waardoor het moeilijk is om een personeelsplanning te maken. Daarnaast is het zo dat door recente veranderingen in de patiëntenpopulatie er minder laagri-

163

sico CABG-patiënten zijn, waardoor het implementeren van een protocol specifiek voor deze groep geen hoge prioriteit heeft.

3 Samenvatting van de aanbevelingen

Gebaseerd op de bevindingen van dit proefschrift kunnen de volgende aanbevelingen worden gedaan:

1. gebruik een vóór-na trial die de totale behandeling van CABG-patiënten evalueert in verschillende ziekenhuizen in plaats van een gerandomiseerde, gecontroleerde trial die alleen de eerste 8 uur intensive-care behandeling in één ziekenhuis evalueert;

2. neem naast de klinische en kosteneffectiviteit ook uitkomsten mee die relevant zijn voor het primaire publiek van de studie (managers van de IC's en directie van ziekenhuizen), zoals de kosten van implementatie en scenario analyses die de impact van een nieuwe behandeling op veranderingen van patiëntstromen en wachtlijsten inzichtelijk maken;

3. neem een maatschappelijk en ziekenhuis perspectief als uitgangspunt voor de economische evaluatie in plaats van alleen een ziekenhuis perspectief;

4. meet alle studieuitkomsten bij zowel hoog- als laagrisico CABG-patiënten met een lange termijn follow-up (1 jaar) in plaats van alleen laagrisico en een korte termijn follow-up voor klinische effecten en kosten zoals in de SSIC-studie;

5. gebruik de SF-6D of de EQ-5D voor het meten van algemene kwaliteit van leven;

6. gebruik de Seatlle Angina of MacNew vragenlijst in plaats van de MILQ voor de meting van ziekte specifieke kwaliteit van leven;

164

7. gebruik, als er baseline verschillen in utilitieitsmetingen op-
treden, de op QALY gebaseerde regressie correctieme-
thode voor de berekening van voor kwaliteit van leven
gecorrigeerde levensjaren in plaats van de delta correc-
tiemethode.

Samenvattend: alle aanbevelingen die hierboven zijn opgenoemd
zijn belangrijk voor toekomstig onderzoek die nieuwe behandelingen
voor CABG-patiënten evalueren. De laatste aanbeveling is echter
ook van belang voor kosteneffectiviteitsanalyses in het algemeen.

4 Praktische aanbevelingen voor toekomstig onderzoek

In dit laatste gedeelte wordt een aantal praktische aanbevelingen ge-
daan voor toekomstig onderzoek.

4.1 Informatie voor personeel in de gezondheidszorg

- Organiseer een kick-off meeting aan het begin van de studie
  voor al het betrokken personeel.
- Houd het onderzoeksprotocol, een samenvatting van het be-
  handelingsprotocol en een up-to-date telefoonlijst beschikbaar
  op een centrale plaats.
- Informeer al het personeel eens per maand over de studie voort-
  gang, en onmiddellijk over wijzigingen in het protocol. Vergeet
  niet nieuwe medewerkers te informeren.

4.2 Evaluatie van de behandeling

- Houd wekelijks multidisciplinaire meetings en evalueer de be-
  handeling van zowel interventie- alsook controlegroep patiënten.
  Notuleer alle belangrijke bevindingen en onderwerpen.

165

- Geef zo snel als mogelijk feedback aan het personeel over behandelingsgerelateerde onderwerpen en het invullen van Case Record Forms.

4.3 Informed consent procedure en patiëntenbegeleiding

- Informeer patiënten persoonlijk op de dag van de opname, in de ochtend.

- Vraag of ze de vragenlijsten willen invullen en vragen te noteren. Verzamel de vragenlijsten in de namiddag en controleer op eventueel onbrekende items.

- Bezoek de patiënten regelmatig gedurende hun verblijf in het ziekenhuis. Bezoek hen ook op de dag dat ze naar huis gaan en informeer ze dan over de follow-up (bijvoorbeeld vragenlijsten en poliklinische bezoeken).

References

References

AHA (2008). American Heart Association, Heart Disease and Stroke Statistics-Update 2008. Available at: http://www.americanheart.org.

Ahonen J, Olkkola KT, Hynynen M, Seppala T, Ikavalko H, Remmerie B, and Salmenpera M (2000). Comparison of alfentanil, fentanyl and sufentanil for total intravenous anaesthesia with propofol in patients undergoing coronary artery bypass surgery. Br J Anaesth 85(4):533-40.

Al-Ruzzeh S, Athanasiou T, Mangoush O, Wray J, Modine T, George S, and Amrani M (2005). Predictors of poor mid-term health related quality of life after primary isolated coronary artery bypass grafting surgery. Heart 91(12):1557-62.

Alpert JS, Thygesen K, Antman E, and Bassand JP (2000). Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. J Am Coll Cardiol 36(3):959-69.

Altman DG (1999). Practical statistics for medical research. New York: Chapman and Hall.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, and Lang T (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 134(8):663-94.

Arom KV, Emery RW, Petersen RJ, and Schwartz M (1995). Cost-effectiveness and predictors of early extubation. Ann Thorac Surg 60(1):127-32.

Avis NE, Smith KW, Hambleton RK, Feldman HA, Selwyn A, and Jacobs A (1996). Development of the multidimensional index of life quality. A quality of life measure for cardiovascular disease. Med Care 34(11):1102-20.

Badia X, Diaz-Prieto A, Gorriz MT, Herdman M, Torrado H, Farrero E, and Cavanilles JM (2001). Using the EuroQol-5D to measure changes in quality of life 12 months after discharge from an intensive care unit. Intensive Care Med 27(12):1901-7.

Bando K, Sun K, Binford RS, and Sharp TG (1997). Determinants of longer duration of endotracheal intubation after adult cardiac operations. Ann Thorac Surg 63(4):1026-33.

Barbieri M, Drummond M, Willke R, Chancellor J, Jolain B, and Towse A (2005). Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: lessons for inferring generalizability. Value Health 8(1):10-23.

Beck AT and Steer RA (1987). Beck Depression Inventory Manual. Toronto: Hatcourt, Brace, Jovanovich.

Berry PD, Thomas SD, Mahon SP, Jackson M, Fox MA, Fabri B, Weir WI, and Russell GN (1998). Myocardial ischaemia after coronary artery bypass grafting: early vs late extubation. Br J Anaesth 80(1):20-5.

BHF (2008). British Heart Foundation, British Heart Foundation's Statistics Available at: http://www.heartstats.org.

Bland JM and Altman DG (1994). Regression towards the mean. Bmj 308(6942):1499.

Books RR and de Charro F (2003). The measurement and valuation of Health status using EQ-5D: a european perspective. Evidence from the EuroQoL BIOMED Research Programme. Dordrecht: Kluwer academic publisher.

Borkon AM, Muehlebach GF, House J, Marso SP, and Spertus JA (2002). A comparison of the recovery of health status after percutaneous coronary intervention and coronary artery bypass. Ann Thorac Surg 74(5):1526-30; discussion 30.

Boudrez H and De Backer G (2001). Psychological status and the role of coping style after coronary artery bypass graft surgery. Results of a prospective study. Qual Life Res 10(1):37-47.

Bowler I, Djaiani G, Abel R, Pugh S, Dunne J, and Hall J (2002). A combination of intrathecal morphine and remifentanil anesthesia for fast-track cardiac anesthesia and surgery. J Cardiothorac Vasc Anesth 16(6):709-14.

Brazier J, Jones N and Kind P (1993). Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 2(3):169-80.

Briggs AH and O'Brien BJ (2001). The death of cost-minimization analysis? Health Econ 10(2):179-84.

Briggs AH, Wonderling DE and Mooney CZ (1997). Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ 6(4):327-40.

Brooks R (1996). EuroQol: the current state of play. Health Policy 37(1):53-72.

Bryan S and Longworth L (2005). Measuring health-related utility: why the disparity between EQ-5D and SF-6D? Eur J Health Econ 6(3):253-60.

Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, and Vray M (1997). Modelling in economic evaluation: an unavoidable fact of life. Health Econ 6(3):217-27.

Caine N, Sharples LD and Wallwork J (1999). Prospective study of health related quality of life before and after coronary artery bypass grafting: outcome at five years. Heart 81(4):347-51.

CBS (2003). Centraal Bureau voor de statistiek, prijsindexcijfers gezondheidszorg. Available at: www.cbs.nl.

Cheng AK, Rubin HR, Powe NR, Mellon NK, Francis HW, and Niparko JK (2000). Cost-utility analysis of the cochlear implant in children. Jama 284(7):850-6.

Cheng DC and Barash PG (2006). Is fast-track intensive care unit management still on the express track? Crit Care Med 34(6):1826-8.

Cheng DC, Karski J, Peniston C, Asokumar B, Raveendran G, Carroll J, Nierenberg H, Roger S, Mickle D, Tong J, Zelovitsky J, David T, and Sandler A (1996a). Morbidity outcome in early versus conventional tracheal extubation after coronary artery bypass grafting: a prospective randomized controlled trial. J Thorac Cardiovasc Surg 112(3):755-64.

Cheng DC, Karski J, Peniston C, Raveendran G, Asokumar B, Carroll J, David T, and Sandler A (1996b). Early tracheal extubation after coronary artery bypass graft surgery reduces costs and improves resource use. A prospective, randomized, controlled trial. Anesthesiology 85(6):1300-10.

Cheng DC, Newman MF, Duke P, Wong DT, Finegan B, Howie M, Fitch J, Bowdle TA, Hogue C, Hillel Z, Pierce E, and Bukenya D (2001). The efficacy and resource utilization of remifentanil and fentanyl in fast-track coronary artery bypass graft surgery: a prospective randomized, double-blinded controlled, multi-center trial. Anesth Analg 92(5):1094-102.

Cheng DC, Wall C, Djaiani G, Peragallo RA, Carroll J, Li C, and Naylor D (2003). Randomized assessment of resource use in fast-track cardiac surgery 1-year after hospital discharge. Anesthesiology 98(3):651-7.

Chocron S, Etievent JP, Viel JF, Dussaucy A, Clement F, Alwan K, Neidhardt M, and Schipman N (1996). Prospective study of quality of life before and after open heart operations. Ann Thorac Surg 61(1):153-7.

Chong JL, Grebenik C, Sinclair M, Fisher A, Pillai R, and Westaby S (1993). The effect of a cardiac surgical recovery area on the timing of extubation. J Cardiothorac Vasc Anesth 7(2):137-41.

Chong JL, Pillai R, Fisher A, Grebenik C, Sinclair M, and Westaby S (1992). Cardiac surgery: moving away from intensive care. Br Heart J 68(4):430-3.

Cohen J (1988). Statistical power analysis for behavioral sciences. New York: Lawrence Erlbaum and Associates.

Coll AM, Ameen JR and Mead D (2004). Postoperative pain assessment tools in day surgery: literature review. J Adv Nurs 46(2):124-33.

Cuthbertson BH, Scott J, Strachan M, Kilonzo M, and Vale L (2005). Quality of life before and after intensive care. Anaesthesia 60(4):332-9.

Dempster M and Donnelly M (2000). Measuring the health related quality of life of people with ischaemic heart disease. Heart 83(6):641-4.

Doering LV, Imperial-Perez F, Monsein S, and Esmailian F (1998). Preoperative and postoperative predictors of early and delayed extubation after coronary artery bypass surgery. Am J Crit Care 7(1):37-44.

Dolan P (1997). Modeling valuations for EuroQol health states. Med Care 35(11):1095-108.

References

Donaldson C MM, Vale L (2002). Criteria for conducting systematic reviews based on economic evaluation studies (chapter 8). In Evidence-based Health economics: BMJ-Books.

Dougherty CM, Dewhurst T, Nichol WP, and Spertus J (1998). Comparison of three quality of life instruments in stable angina pectoris: Seattle Angina Questionnaire, Short Form Health Survey (SF-36), and Quality of Life Index-Cardiac Version III. J Clin Epidemiol 51(7):569-75.

Dowie J (2002). Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Econ 11(1):1-8.

Drummond MF, O'Brien B, Stoddart GL, and Torrance GW (1996). Methods for the economic evaluation of health care programmes. Oxford: Oxford medical publications.

Drummond MF and Pang E (2001). Transferability of economic evaluation results. In Economic evaluation in health care; merging theory with practice, edited by MF Drummond and A McGuire, pp. 256-76. Oxford: Oxford University Press.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, and Stoddart GL (2005). Methods for the economic evaluation of health care programmes Oxford: Oxford medical publications.

Duits AA, Boeke S, Taams MA, Passchier J, and Erdman RA (1997). Prediction of quality of life after coronary artery bypass graft surgery: a review and evaluation of multiple, recent studies. Psychosom Med 59(3):257-68.

Dumas A, Dupuis GH, Searle N, and Cartier R (1999). Early versus late extubation after coronary artery bypass grafting: effects on cognitive function. J Cardiothorac Vasc Anesth 13(2):130-5.

Eisenberg E, Pultorak Y, Pud D, and Bar-El Y (2001). Prevalence and characteristics of post coronary artery bypass graft surgery pain (PCP). Pain 92(1-2):11-7.

Engelman RM (1996). Mechanisms to reduce hospital stays. Ann Thorac Surg 61(2 Suppl):S26-9; discussion S33-4.

Engelman RM, Rousou JA, Flack JE, 3rd, Deaton DW, Humphrey CB, Ellison LH, Allmendinger PD, Owen SG, and Pekow PS (1994). Fast-track recovery of the coronary bypass patient. Ann Thorac Surg 58(6):1742-6.

Engoren M, Luther G and Fenn-Buderer N (2001). A comparison of fentanyl, sufentanil, and remifentanil for fast-track cardiac anesthesia. Anesth Analg 93(4):859-64.

Engoren MC, Kraras C and Garzia F (1998). Propofol-based versus fentanyl-isoflurane-based anesthesia for cardiac surgery. J Cardiothorac Vasc Anesth 12(2):177-81.

EuroQoL-Group (1990). EuroQoL-a new facility for the measurement of health-related quality of life. Health Policy 16(3):199-208.

Falger PS, EHWJ van Leuteren, KSJ (2000). Wederzijdse beinvloeding van de kwaliteit van leven van patienten en significante anderen na een hartinfarct of een 'coronary artery bypass-graft'-operatie. Gedrag en Gezondheid 28(5):274-87.

Fenwick E, O'Brien BJ and Briggs A (2004). Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. Health Econ 13(5):405-15.

Fillinger MP, Yeager MP, Dodds TM, Fillinger MF, Whalen PK, and Glass DD (2002). Epidural anesthesia and analgesia: effects on recovery from cardiac surgery. J Cardiothorac Vasc Anesth 16(1):15-20.

Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, and Mowat A (1993). A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. J Rheumatol 20(3):429-36.

Flynn M, Reddy S, Shepherd W, Holmes C, Armstrong D, Lunn C, Khan K, and Kendall S (2004). Fast-tracking revisited: routine cardiac surgical patients need minimal intensive care. Eur J Cardiothorac Surg 25(1):116-22.

Foster GH, Conway WA, Pamulkov N, Lester JL, and Magilligan DJ, Jr. (1984). Early extubation after coronary artery bypass: brief report. Crit Care Med 12(11):994-6.

Gelling L (1999). Causes of ICU psychosis: the environmental factors. Nurs Crit Care 4(1):22-6.

Gold MR, Siegel JE, Russell LB, and Weinstein MC (1996). Cost-effectiveness in health and medicine. New York: Oxford University Press.

Gold MR SJ, Russell LB, Weinstein MC (1996). Cost-effectiveness in health and medicine. New York: Oxford University Press.

Goss JR, Maynard C, Aldea GS, Marcus-Smith M, Whitten RW, Johnston G, Phillips RC, Reisman M, Kelley A, and Anderson RP (2006). Effects of a statewide physician-led quality-improvement program on the quality of cardiac care. Am Heart J 151(5):1033-42.

Grol R and Wensing M (2001). Implementatie, effectieve verandering in de patientenzorg [In Dutch]. Maarssen: Elsevier gezondheidszorg.

Guthrie E, Moorey J, Margison F, Barker H, Palmer S, McGrath G, Tomenson B, and Creed F (1999). Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. Arch Gen Psychiatry 56(6):519-26.

Guyatt GH (2002). Commentary on Jack Dowie, "Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions". Health Econ 11(1):9-12; discussion 21-2.

Guyatt GH, Veldhuyzen Van Zanten SJ, Feeny DH, and Patrick DL (1989). Measuring quality of life in clinical trials: a taxonomy and review. Canadian medical Association Journal 140(12):1441-8.

Habib RH, Zacharias A and Engoren M (1996). Determinants of prolonged mechanical ventilation after coronary artery bypass grafting. Ann Thorac Surg 62(4):1164-71.

Hall RI, Murphy JT, Moffitt EA, Landymore R, Pollak PT, and Poole L (1991). A comparison of the myocardial metabolic and haemodynamic changes produced by propofol-sufentanil and enflurane-sufentanil anaesthesia for patients having coronary artery bypass graft surgery. Can J Anaesth 38(8):996-1004.

Hawkes C, Dhileepan S and Foxcroft D (2003). Early extubation for adult cardiac surgical patients. Cochrane Database Syst Rev 4:CD003587.

Herlitz J, Brandrup-Wognsen G, Caidahl K, Hartford M, Haglid M, Karlson BW, Karlsson T, and Sjoland H (2005). Determinants for an impaired quality of life 10 years after coronary artery bypass surgery. Int J Cardiol 98(3):447-52.

Herlitz J, Wiklund I, Caidahl K, Karlson BW, Sjoland H, Hartford M, Haglid M, and Karlsson T (1999). Determinants of an impaired quality of life five years after coronary artery bypass surgery. Heart 81(4):342-6.

Heyland DK, Guyatt G, Cook DJ, Meade M, Juniper E, Cronin L, and Gafni A (1998). Frequency and methodologic rigor of quality-of-life assessments in the critical care literature. Crit Care Med 26(3):591-8.

Higgins TL (1992). Pro: early endotracheal extubation is preferable to late extubation in patients following coronary artery surgery. J Cardiothorac Vasc Anesth 6(4):488-93.

Hofer S, Benzer W, Schussler G, von Steinbuchel N, and Oldridge NB (2003). Health-related quality of life in patients with coronary artery disease treated for angina: validity and reliability of German translations of two specific questionnaires. Qual Life Res 12(2):199-212.

Hofer S, Lim L, Guyatt G, and Oldridge N (2004). The MacNew Heart Disease health-related quality of life instrument: a summary. Health Qual Life Outcomes 2:3.

Holland R, Smith RD, Harvey I, Swift L, and Lenaghan E (2004). Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. Health Econ 13(8):793-805.

Howie MB, Cheng D, Newman MF, Pierce ET, Hogue C, Hillel Z, Bowdle TA, and Bukenya D (2001). A randomized double-blinded multicenter comparison of remifentanil versus fentanyl when combined with isoflurane/propofol for early extubation in coronary artery bypass graft surgery. Anesth Analg 92(5):1084-93.

Hurel D, Loirat P, Saulnier F, Nicolas F, and Brivet F (1997). Quality of life 6 months after intensive care: results of a prospective multicenter study using a generic health status scale and a satisfaction scale. Intensive Care Med 23(3):331-7.

Hurskainen R, Teperi J, Rissanen P, Aalto AM, Grenman S, Kivela A, Kujansuu E, Vuorma S, Yliskoski M, and Paavonen J (2001). Quality of life and cost-effectiveness of levonorgestrel-releasing intrauterine system versus hysterectomy for treatment of menorrhagia: a randomised trial. Lancet 357(9252):273-7.

171

Hwang NC, Shankar S, Ong BC, Leong CK, Chin E, Chua YL, Tan YS, and Ng J (1999). Changing the institutional practice of prolonged mechanical ventilation after coronary artery bypass graft surgery to early extubation. Ann Acad Med Singapore 28(4):534-41.

ISPOR (2008). International Society for Pharmacoeconomics and Outcomes Research, Pharmacoeconomic Guidelines Around the World. Available at: http://www.ispor.org/peguidelines/index.asp.

Jenkinson C and McGee H (1998). Health status measurement: a brief but critical introduction. Oxford: Radcliffe Medical Press.

Joffe MM and Rosenbaum PR (1999). Invited commentary: propensity scores. Am J Epidemiol 150(4):327-33.

Johnson RG (2006). Grading the evidence for shortened intensive care unit stays after cardiac operations. Crit Care Med 34(1):242-3.

Kadoi Y, Saito S, Kunimoto F, Goto F, and Fujita N (2003). Comparative effects of propofol versus fentanyl on cerebral oxygenation state during normothermic cardiopulmonary bypass and postoperative cognitive dysfunction. Ann Thorac Surg 75(3):840-6.

Karski JM (1995). Practical aspects of early extubation in cardiac surgery. J Cardiothorac Vasc Anesth 9(5 Suppl 1):30-3.

Karski JM, Dowd NP, Joiner R, Carroll J, Peniston C, Bailey K, Glynn MF, Teasdale SJ, and Cheng DC (1998). The effect of three different doses of tranexamic acid on blood loss after cardiac surgery with mild systemic hypothermia (32 degrees C). J Cardiothorac Vasc Anesth 12(6):642-6.

Klineberg PL, Geer RT, Hirsh RA, and Aukburg SJ (1977). Early extubation after coronary artery bypass graft surgery. Crit Care Med 5(6):272-4.

Konopad E, Noseworthy TW, Johnston R, Shustack A, and Grace M (1995). Quality of life measures before and one year after admission to an intensive care unit. Crit Care Med 23(10):1653-9.

Konstantakos AK and Lee JH (2000). Optimizing timing of early extubation in coronary artery bypass surgery patients. Ann Thorac Surg 69(6):1842-5.

Kopec JA and Willison KD (2003). A comparative review of four preference-weighted measures of health-related quality of life. J Clin Epidemiol 56(4):317-25.

Langeluddecke P, Fulcher G, Baird D, Hughes C, and Tennant C (1989). A prospective evaluation of the psychosocial effects of coronary artery bypass surgery. J Psychosom Res 33(1):37-45.

Latham P, Zarate E, White PF, Bossard R, Shi C, Morse LS, Douning LK, and Chi L (2000). Fast-track cardiac anesthesia: a comparison of remifentanil plus intrathecal morphine with sufentanil in a desflurane-based anesthetic. J Cardiothorac Vasc Anesth 14(6):645-51.

Lazar HL, Fitzgerald CA, Ahmad T, Bao Y, Colton T, Shapira OM, and Shemin RJ (2001). Early discharge after coronary artery bypass graft surgery: are patients really going home earlier? J Thorac Cardiovasc Surg 121(5):943-50.

Lee DS, Jang MJ, Cheon GJ, Chung JK, and Lee MC (2002). Comparison of the cost-effectiveness of stress myocardial SPECT and stress echocardiography in suspected coronary artery disease considering the prognostic value of false-negative results. J Nucl Cardiol 9(5):515-22.

Lee JH, Kim KH, vanHeeckeren DW, Murrell HK, Cmolik BL, Graber R, Effron B, and Geha AS (1996). Cost analysis of early extubation after coronary bypass surgery. Surgery 120(4):611-7; discussion 17-9.

Lim LL, Valenti LA, Knapp JC, Dobson AJ, Plotnikoff R, Higginbotham N, and Heller RF (1993). A self-administered quality-of-life questionnaire after acute myocardial infarction. J Clin Epidemiol 46(11):1249-56.

Longworth L and Bryan S (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health Econ 12(12):1061-7.

Macran S, Weatherly H and Kind P (2003). Measuring population health: a comparison of three generic health status measures. Med Care 41(2):218-31.

Manca A, Hawkins N and Sculpher MJ (2005). Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 14(5):487-96.

Mangano DT, Siliciano D, Hollenberg M, Leung JM, Browner WS, Goehner P, Merrick S, and Verrier E (1992). Postoperative myocardial ischemia. Therapeutic trials using intensive analgesia following surgery. The Study of Perioperative Ischemia (SPI) Research Group. Anesthesiology 76(3):342-53.

Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, Esdaile JM, and Anis AH (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med 60(7):1571-82.

Mason JM and Mason AR (2006). The generalisability of pharmacoeconomic studies: issues and challenges ahead. Pharmacoeconomics 24(10):937-45.

Mayou R and Bryant B (1993). Quality of life in cardiovascular disease. Br Heart J 69(5):460-6.

McDowell I and Newell C (1996). Measuring health: a guide to rating scales and questionnaires. New York: Oxford University Press.

McHorney CA and Tarlov AR (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 4(4):293-307.

Meade MO, Guyatt G, Butler R, Elms B, Hand L, Ingram A, and Griffith L (2001). Trials comparing early vs late extubation following cardiovascular surgery. Chest 120(6 Suppl):445S-53S.

Michalopoulos A, Nikolaides A, Antzaka C, Deliyanni M, Smirli A, Geroulanos S, and Papadimitriou L (1998). Change in anaesthesia practice and postoperative sedation shortens ICU and hospital length of stay following coronary artery bypass surgery. Respir Med 92(8):1066-70.

Michel P, Roques F and Nashef SA (2003). Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg 23(5):684-7; discussion 87.

Mollhoff T, Herregods L, Moerman A, Blake D, MacAdams C, Demeyere R, Kirno K, Dybvik T, and Shaikh S (2001). Comparative efficacy and safety of remifentanil and fentanyl in 'fast track' coronary artery bypass graft surgery: a randomized, double-blind study. Br J Anaesth 87(5):718-26.

Moock J and Kohlmann T (2008). Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 17(3):485-95.

Mora CT, Dudek C, Torjman MC, and White PF (1995). The effects of anesthetic technique on the hemodynamic response and recovery profile in coronary revascularization patients. Anesth Analg 81(5):900-10.

Myles PS, Buckland MR, Weeks AM, Bujor MA, McRae R, Langley M, Moloney JT, Hunt JO, and Davis BB (1997). Hemodynamic effects, myocardial ischemia, and timing of tracheal extubation with propofol-based anesthesia for cardiac surgery. Anesth Analg 84(1):12-9.

Myles PS, Daly DJ, Djaiani G, Lee A, and Cheng DC (2003). A systematic review of the safety and effectiveness of fast-track cardiac anesthesia. Anesthesiology 99(4):982-7.

Myles PS, Hunt JO, Fletcher H, Watts J, Bain D, Silvers A, and Buckland MR (2002). Remifentanil, fentanyl, and cardiac surgery: a double-blinded, randomized, controlled trial of costs and outcomes. Anesth Analg 95(4):805-12, table of contents.

Nallamothu BK, Young J, Gurm HS, Pickens G, and Safavi K (2007). Recent trends in hospital utilization for acute myocardial infarction and coronary revascularization in the United States. Am J Cardiol 99(6):749-53.

NHS (2006). Nederlandse Hartstichting, hart- en vaatziekten in nederland. Available at: www.hartstichting.nl.

Nilsson J, Algotsson L, Hoglund P, Luhrs C, and Brandt J (2006). Comparison of 19 preoperative risk stratification models in open-heart surgery. Eur Heart J 27(7):867-74.

Nowels D, McGloin J, Westfall JM, and Holcomb S (2005). Validation of the EQ-5D quality of life instrument in patients after myocardial infarction. Qual Life Res 14(1):95-105.

References

NYHA (1994). New York Heart Association Classification, Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. pp. 253-56. Boston: Little, Brown & Co.

O'Rourke (2002). Mixed means and medians: U unified approach to deal with disparate outcome summaries, presented at Symposium on systematic reviews: publishing the boundaries. Oxford.

Oostenbrink JB, Buijs-Van der Woude T, van Agthoven M, Koopmanschap MA, and Rutten FF (2003). Unit costs of inpatient hospital days. Pharmacoeconomics 21(4):263-71.

Oostenbrink JB, Koopmanschap MA and Rutten FFH (2000). Handleiding voor kostenonderzoek, methoden en richtlijnen voor economische evaluatie in de gezondheidszorg. Amstelveen: College van Zorgverzekeringen.

Parsonnet V, Dean D and Bernstein AD (1989). A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. Circulation 79(6 Pt 2):I3-12.

Patterson RE, Eisner RL and Horowitz SF (1995). Comparison of cost-effectiveness and utility of exercise ECG, single photon emission computed tomography, positron emission tomography, and coronary angiography for diagnosis of coronary artery disease. Circulation 91(1):54-65.

Petter AH, Chiolero RL, Cassina T, Chassot PG, Muller XM, and Revelly JP (2003). Automatic "respirator/weaning" with adaptive support ventilation: the effect on duration of endotracheal intubation and patient management. Anesth Analg 97(6):1743-50.

Phillips Bute B, Mathew J, Blumenthal JA, Welsh-Bohmer K, White WD, Mark D, Landolfo K, and Newman MF (2003). Female gender is associated with impaired quality of life 1 year after coronary artery bypass surgery. Psychosom Med 65(6):944-51.

Pirraglia PA, Peterson JC, Williams-Russo P, Gorkin L, and Charlson ME (1999). Depressive symptomatology in coronary artery bypass graft surgery patients. Int J Geriatr Psychiatry 14(8):668-80.

Pocock SJ, Assmann SE, Enos LE, and Kasten LE (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 21(19):2917-30.

Prakash O, Jonson B, Meij S, Bos E, Hugenholtz PG, Nauta J, and Hekman W (1977). Criteria for early extubation after intracardiac surgery in adults. Anesth Analg 56(5):703-8.

Priestley MC, Cope L, Halliwell R, Gibson P, Chard RB, Skinner M, and Klineberg PL (2002). Thoracic epidural anesthesia for cardiac surgery: the effects on tracheal intubation time and length of hospital stay. Anesth Analg 94(2):275-82, table of contents.

Prismant (2008). Ziekenhuisstatistieken. Available at: http://www.prismant.nl.

Quasha AL, Loeber N, Feeley TW, Ullyot DJ, and Roizen MF (1980). Postoperative respiratory care: a controlled trial of early and late extubation following coronary-artery bypass grafting. Anesthesiology 52(2):135-41.

Reyes A, Vega G, Blancas R, Morato B, Moreno JL, Torrecilla C, and Cereijo E (1997). Early vs conventional extubation after cardiac surgery with cardiopulmonary bypass. Chest 112(1):193-201.

Richardson G and Manca A (2004). Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Econ 13(12):1203-10.

RIVM (2008a). Rijksinstituut voor Volksgezondheid en Milieu, Cost of illness in the Netherlands 2005. Available at: www.costofillness.nl.

RIVM (2008b). Rijksinstituut voor Volksgezondheid en Milieu, Nationaal Kompas Volksgezondheid: Gezondheid en ziekte. Available at: http://www.nationaalkompas.nl.

Rosow C (1993). Remifentanil: a unique opioid analgesic. Anesthesiology 79(5):875-6.

Royse C, Royse A, Soeding P, Blake D, and Pang J (2003). Prospective randomized trial of high thoracic epidural analgesia for coronary artery bypass surgery. Ann Thorac Surg 75(1):93-100.

174

Rumsfeld JS, Magid DJ, O'Brien M, McCarthy M, Jr., MaWhinney S, Scd, Shroyer AL, Moritz TE, Henderson WG, Sethi GK, Grover FL, and Hammermeister KE (2001). Changes in health-related quality of life following coronary artery bypass graft surgery. Ann Thorac Surg 72(6):2026-32.

Samuelson PN, Reves JG, Kirklin JK, Bradley E, Jr., Wilson KD, and Adams M (1986). Comparison of sufentanil and enflurane-nitrous oxide anesthesia for myocardial revascularization. Anesth Analg 65(3):217-26.

Scholte op Reimer WJM, Gitt AK, Boersma E, and Simoons ML (2006). Cardiovascular Diseases in Europe. Euro Heart Survey – 2006. Sophia Antipolis: European Society of Cardiology.

Schrader G, Cheok F, Hordacre AL, and Guiver N (2004). Predictors of depression three months after cardiac hospitalization. Psychosom Med 66(4):514-20.

Schweikert B, Hahmann H and Leidl R (2006). Validation of the EuroQol questionnaire in cardiac rehabilitation. Heart 92(1):62-7.

Sculpher MJ, Claxton K, Drummond M, and McCabe C (2006). Whither trial-based economic evaluation for health care decision making? Health Econ 15(7):677-87.

Sherry KM, McNamara J, Brown JS, and Drummond M (1996). An economic evaluation of propofol/fentanyl compared with midazolam/fentanyl on recovery in the ICU following cardiac surgery. Anaesthesia 51(4):312-7.

Shroff A, Rooke GA and Bishop MJ (1997). Effects of intrathecal opioid on extubation time, analgesia, and intensive care unit stay following coronary artery bypass grafting. J Clin Anesth 9(5):415-9.

Silbert BS, Santamaria JD, O'Brien JL, Blyth CM, Kelly WJ, and Molnar RR (1998). Early extubation following coronary artery bypass surgery: a prospective randomized controlled trial. The Fast Track Cardiac Care Team. Chest 113(6):1481-8.

Simchen E, Galai N, Braun D, Zitser-Gurevich Y, Shabtai E, and Naveh I (2001). Sociodemographic and clinical factors associated with low quality of life one year after coronary bypass operations: the Israeli coronary artery bypass study (ISCAB). J Thorac Cardiovasc Surg 121(5):909-19.

Simeone F, Biagioli B, Scolletta S, Marullo AC, Marchet-Ti L, Caciorgna M, and Giomarelli P (2002). Optimization of mechanical ventilation support following cardiac surgery. J Cardiovasc Surg (Torino) 43(5):633-41.

Sirio CA and Martich GD (1999). Who goes to the ICU postoperatively? Chest 115(5 Suppl):125S-29S.

Sleeper LA, Ramanathan K, Picard MH, Lejemtel TH, White HD, Dzavik V, Tormey D, Avis NE, and Hochman JS (2005). Functional status and quality of life after emergency revascularization for cardiogenic shock complicating acute myocardial infarction. J Am Coll Cardiol 46(2):266-73.

Smith HJ, Taylor R and Mitchell A (2000). A comparison of four quality of life instruments in cardiac patients: SF-36, QLI, QLMI, and SEIQoL. Heart 84(4):390-4.

Smith KW, Avis NE and Assmann SF (1999). Distinguishing between quality of life and health status in quality of life research: a meta-analysis. Qual Life Res 8(5):447-59.

Smith KW and Larson MJ (2003). Quality of life assessments by adult substance abusers receiving publicly funded treatment in Massachusetts. Am J Drug Alcohol Abuse 29(2):323-35.

Spertus JA, Winder JA, Dewhurst TA, Deyo RA, and Fihn SD (1994). Monitoring the quality of life in patients with coronary artery disease. Am J Cardiol 74(12):1240-4.

Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell M, and Fihn SD (1995). Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. J Am Coll Cardiol 25(2):333-41.

Stanley TH and Webster LR (1978). Anesthetic requirements and cardiovascular effects of fentanyl-oxygen and fentanyl-diazepam-oxygen anesthesia in man. Anesth Analg 57(4):411-6.

Staquet M, Berzon R, Osoba D, and Machin D (1996). Guidelines for reporting results of quality of life assessments in clinical trials. Qual Life Res 5(5):496-502.

References

Straus SE, Richardson WS, Glaszou P, and Haynes RB (2005). Evidence based medicine, how to practice and teach EBM. Philadelphia: Elsevier.

STS (2003). Society of Thoracic surgeons, Definitions of terms of the society of Thoracic surgeons national cardiac surgery database. Available at: http://www.sts.org.

Sulzer CF, Chiolero R, Chassot PG, Mueller XM, and Revelly JP (2001). Adaptive support ventilation for fast tracheal extubation after cardiac surgery: a randomized controlled study. Anesthesiology 95(6):1339-45.

Suter P, Arma-Ganadis A and Beaufils F (1994). Predicting outcome in ICU patients. 2nd European Consensus Conference in Intensive Care Medicine. Intensive Care Med 20(5):390-7.

Thompson DR, Meadows KA and Lewin RJ (1998). Measuring quality of life in patients with coronary heart disease. Eur Heart J 19(5):693-5.

Thompson DR and Yu CM (2003). Quality of life in patients with coronary heart disease-I: Assessment tools. Health Qual Life Outcomes 1(1):42.

Thompson SG and Higgins JP (2002). How should meta-regression analyses be undertaken and interpreted? Stat Med 21(11):1559-73.

Tritapepe L, Voci P, Di Giovanni C, Pizzuto F, Cuscianna E, Caretta Q, and Pietropaoli P (2002). Alfentanil and sufentanil in fast-track anesthesia for coronary artery bypass graft surgery. J Cardiothorac Vasc Anesth 16(2):157-62.

van der Ploeg HD, PB, Spielberger, C. (1986). Handleiding bij de Zelf-Beoordelings vragenlijst [Manual for the Staite-Trait Anxiety Inventory (STAI)]. Lisse: Swets and Zeitlinger BV.

van Hout BA, Al MJ, Gordon GS, and Rutten FF (1994). Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 3(5):309-19.

van Mastrigt GA, Heijmans J, Severens JL, Fransen EJ, Roekaerts P, Voss G, and Maessen JG (2006). Short-stay intensive care after coronary artery bypass surgery: randomized clinical trial on safety and cost-effectiveness. Crit Care Med 34(1):65-75.

van Stel HF and Buskens E (2006). Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. Health Qual Life Outcomes 4:20.

vander Salm TJ and Blair SA (1984). Effect of reduction of postoperative days in the intensive care unit after coronary artery bypass. J Thorac Cardiovasc Surg 88(4):558-61.

Velasco FT, Tarlow LS and Thomas SJ (1995). Economic rationale for early extubation. J Cardiothorac Vasc Anesth 9(5 Suppl 1):2-9.

Vickers AJ (2001). The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. BMC Med Res Methodol 1(1):6.

Welke KF, Stevens JP, Schults WC, Nelson EC, Beggs VL, and Nugent WC (2003). Patient characteristics can predict improvement in functional health after elective coronary artery bypass grafting. Ann Thorac Surg 75(6):1849-55; discussion 55.

Westaby S, Pillai R, Parry A, O'Regan D, Giannopoulos N, Grebenik K, Sinclair M, and Fisher A (1993). Does modern cardiac surgery require conventional intensive care? Eur J Cardiothorac Surg 7(6):313-8; discussion 18.

Wiebe S, Guyatt G, Weaver B, Matijevic S, and Sidwell C (2003). Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 56(1):52-60.

Yun KL, Sintek CF, Fletcher AD, Pfeffer TA, Kochamba GS, Mahrer PR, and Khonsari S (1999). Time related quality of life after elective cardiac operation. Ann Thorac Surg 68(4):1314-20.

Zarate E, Latham P, White PF, Bossard R, Morse L, Douning LK, Shi C, and Chi L (2000). Fast-track cardiac anesthesia: use of remifentanil combined with intrathecal morphine as an alternative to sufentanil during desflurane anesthesia. Anesth Analg 91(2):283-7.

176

Dankwoord

Allereerst wil ik alle patiënten en hun familie bedanken die hebben meegewerkt aan deze short-stay studie. Zonder hen was deze studie nooit tot een goed einde gekomen. Natuurlijk wil ik ook graag alle personeelsleden van de verschillende cardio-chirurgische afdelingen (metname de cardio-chirurgische intensive care (F3) en de verpleegafdeling (D4)) en operatiekamers bedanken die onze 600 patiënten hebben geopereerd en behandeld. Bedankt voor jullie inzet en vertrouwen in onze studie.

In het bijzonder en met naam wil ik de volgende personen bedanken:
mijn promotores, professor dr. Jos Maessen, professor dr. Hans Severens en professor dr. Martin Prins. Beste Jos, bedankt dat je mij de mogelijkheid hebt geboden om te kunnen promoveren. Met vereende krachten is deze klus na 8 jaar dan toch geklaard. Wie had dat gedacht? Ik heb veel waardering voor je gedrevenheid: een afdeling runnen, de hele dag opereren en ook nog een hele club onderzoekers begeleiden. Hopelijk blijft er voor jou genoeg tijd over om samen met je vrouw mooie reizen en wandelingen te maken op zoek naar beren in Kamtsjatka.
Beste Hans, ook jou wil ik graag bedanken voor je inbreng in mijn proefschrift. Je zult ook wel blij zijn dat het niet alleen is gebleven bij de 40 versies van de tijdsplanningen. Ik kon altijd op korte termijn bij je terecht om inhoudelijke zaken te bespreken. Zeker zo belangrijk voor mij was de motivatie die je mij gaf om vooral door te gaan en niet voortijdig te stoppen.
Beste Martin, bedankt voor wat je allemaal hebt geregeld om mijn aanstellingen te verlengen zodat ik bijna 7,5 jaar bij de KEMTA kon blijven werken. Het was ook fijn om met jou een artikel te schrijven: van jouw snelle en gestructureerde manier van werken heb ik veel geleerd.

178

Ik wil de leden van de beoordelingscommissie, prof. dr. J. H. Zwaveling, prof. dr. E. Buskens, prof. dr. J. W. M. Greve en dr. J. P. A. M. Schönberger bedanken voor hun beoordeling van dit proefschrift.

Monique Miesen en Monique Dols: ten tijde van de uitvoering van de trial waren we een hecht team. Dankzij jullie inzet liep de SSIC-studie gesmeerd en werden de patiënten en verpleegkundigen gemotiveerd om alles in te vullen en bij te houden. Bedankt hiervoor.

Lysette Liegeois: tijdens de studie was jouw inbreng aanzienlijk met het notuleren van alle vergaderingen alsmede door het invoeren van veel vragenlijsten. Maar ook andere zaken kon ik met een gerust hart aan jou overlaten, want een ding was/is duidelijk: als je Lysette iets vraagt, is het vandaag nog geregeld! Ook de andere dames van het secretariaat van de cardio-thoracale chirurgie, bedankt voor jullie ondersteuning in het short-stay project.

Beste John Heijmans, het was leuk en leerzaam om met jou samen te werken. Bedankt dat je jouw medische kennis met mij wilde delen.

Ook de andere leden van de SSIC-groep -Paul Roekaerts, Jos Swennen, Chris Bouts en Erik Franssen- bedankt voor jullie inbreng in de short-stay studie. Vele uren vergaderen over de short-stay patiënten heeft tot vele nieuwe inzichten geleid die hopelijk ook nog nu bruikbaar zijn.

Aangezien het al een hele tijd geleden is dat dit short-stay project heeft plaats gevonden, wil ik mij de dezen excuseren als ik iemand heb vergeten te noemen in dit dankwoord.

Beste Fred Nieman en Fons Kessels, bedankt voor het delen van jullie statistische kennis zodat ik deze kon gebruiken in mijn proefschrift.

Beste Manuela Joore, heel erg bedankt voor je hulp bij het schrijven van de twee kwaliteit van leven artikelen. Dankzij jou werden ze een stuk beter!

179

Beste Gemma Voss, jij was betrokken bij de begeleiding van mijn eerste projecten bij de KEMTA en de opzet van het SSIC project. Ik heb veel aan je gehad, bedankt hiervoor.

Irene van Hoof, bedankt voor de ondersteuning die jij bood bij het plannen van de vele meetings voor het bij elkaar krijgen van mijn promotie-team.

Thea, Brigitte en Jody. Ondertussen alle drie gepromoveerd, ja eentje moet de laatste zijn. Bedankt voor de gezellige etentjes en de motiverende gesprekken over promoties, werk en andere 'echt' belangrijke zaken in het leven.

Daniëlle Brunenberg, lief en leed hebben we gedeeld gedurende de 6 jaar dat we bij elkaar op de kamer zaten. Je was een fijne kamergenoot. Proficiat met de geboorte van je dochter Iris.

Audry, succes met de laatste loodjes van jou proefschrift.

Mijn eerder genoemde en andere ex-KEMTA-collega's, Bert, Caro, Carmen, Claire, Debbie, Daniëlle Willems, Isabelle, Janneke, Linda en Margriet, bedankt voor de inhoudelijke discussies en de onvergetelijke Café Zuid borrels.

Ook wil ik de leden van de MTA-werkgroep in Maastricht bedanken voor hun leerzame en boeiende discussies.

Lieve familie, schoonfamilie en vrienden, jullie wil ik bedanken dat jullie er altijd voor mij waren. Ik hoop dat er nu weer wat meer tijd komt om elkaar regelmatiger te zien.

Lieve Elfi en Mira, vriendinnen zoals jullie zou ik iedereen willen toewensen. Ik ben dan ook heel blij dat jullie mijn paranimfen zijn. De laatste loodjes waren zwaar, maar gelukkig kan ik altijd op jullie steun rekenen.

Lieve Stephanie, al is de afstand Maastricht-Amsterdam soms niet handig, echte vriendschappen blijven toch overeind. Ik ben blij dat jij deze mooie kaft van dit proefschrift hebt willen ontwerpen.

180

Lieve Aimée, een zus zoals jij is er maar één. Eindelijk hebben we weer meer tijd om leuke dingen te ondernemen. Bedankt voor al je steun!

Lieve papa en mama, jullie hebben mij nooit gepusht om iets te doen, maar wel mij het vertrouwen gegeven dat ik iets kon als ik mijn hart maar volgde. Wie had ooit gedacht dat deze promotie toch nog eens af zou komen? Bedankt voor jullie onvoorwaardelijke steun! En ook, vooral mama, voor alle ontelbare oppasdagen. De kinderen kunnen zich geen betere en lievere ama wensen.

Leeve Stephan, het waor neet altied mekelijk om oan mien promotie te werken, dien begrip en geduld met mich woord er deks door op de preuf gesteld. Zonder diene onveurwaordelijke steun waor dit beukste er noets gewees. Wie goan veer dit vieren? Eindelijk, eindelijk is het aof!!

Leeve Jip en Julie, geer maak alle aandere dinger in het leeve onbelangriek. Wat een vreug um uuch in mien leeve te höbbe. Ich bin gruuts op uch!

Jip, Julie en Stephan, leeve schatte, mien poemmelkes, ich hou hiel hiel veul vaan uuch!

181

Curriculum Vitae

Ghislaine van Mastrigt werd geboren op 17 augustus 1973 te Heerlen. In 1992 heeft zij na het behalen van het Mavo-diploma op het Trichtercollege te Maastricht ook de Havo afgerond. Vervolgens is zij een jaar in Frankrijk Franse Taal- en Letterkunde gaan studeren. In 1993 heeft ze de propedeuse van de studie HBO-Verpleegkunde behaald en is vervolgens Gezondheidswetenschappen gaan studeren aan de Universiteit van Maastricht. Zij studeerde af in 1999 in de studierichting Biologische Gezondheidskunde. Na het behalen van haar diploma is ze gaan werken bij de afdeling Klinische Epidemiologie en Medical Technology Assessment (KEMTA) van het academisch ziekenhuis Maastricht. Hier heeft ze 6.5 jaar gewerkt aan diverse onderzoeksprojecten. In de periode van 2002 tot 2004 heeft ze aan de Universiteit van Maastricht het epidemiologie-tracé gevolgd. Van januari 2007 tot en met juni 2008 heeft ze als onderzoeker bij de Maastro Clinic in Maastricht gewerkt. Van januari tot augustus 2009 is ze als projectleider bij de Taakgroep Statistiek Ontwikkeling van het CBS in Heerlen werkzaam geweest.

Samen met haar partner Stephan Mullers heeft zij twee dochtertjes: Jip (2006) en Julie (2008).