The DOORS-study of on-pump versus off-pump coronary artery bypass grafting: A post hoc analysis of methods for multiple imputation of missing data in economic evaluation.

Authors:

1Lars Ehlers, M.Sc., Ph.D., professor of health economics
2Morten Fenger-Grøn, M.Sc., statistician
3Søren Sønderup Beck, M.Sc., HTA Consultant
4Kim Houlin M.D., Ph.d. Assoc. Professor
5Jørgen Lauridsen, M.Sc., Ph.D., professor of health economics

Affiliations:

1Department of Business & Management, Aalborg University, Denmark
2Department of Clinical Epidemiology, Aarhus University Hospital, Denmark & Research Unit for General Practice, Aarhus University, Denmark
3HTA & Health Services Research, Centre for Public Health, Central Denmark Region & Department of HTA, School of Public Health, Aarhus University
4Dept. of Cardiothoracic and Vascular Surgery, Aarhus University Hospital, Denmark
5The Research Unit of Health Economics, University of Southern Denmark, Denmark
Abstract

Background: Former studies of the cost-effectiveness of conventional coronary artery bypass grafting (CCAGB) vs. off-pump coronary artery bypass (OPCAB) favors OPCAB, however CCABG remains the most commonly used surgical technique. Explanations for this includes uncertainty of differences in risk of side effects as well as former trials’ focusing on low risk patients and high risk patients’ liability to aberration from intended treatment and loss to follow-up.

Methods: A cost-utility analysis was conducted alongside the Danish On-pump Off-pump Randomization Study (DOORS) based on the intention to treat principle. A post hoc analysis of the problem of missing data was addressed by multiple imputation using the conditional Gaussian as well as the chained equation approach. Both methods where applied using two different models (representing a data-driven respectively a clinical reasoning selection strategy).

Results: The cost-effectiveness acceptability curve for the complete case analysis (n=779) showed 88 % probability of OPCAB being cost-effective at a threshold value of £30,000/QALY. The four analyses based on multiple imputation showed agreeing results of 73-75 %, though estimated ICER proved susceptible to choice of imputation model.

Conclusion: The result of the previously published complete-case analysis of the cost-effectiveness of OPCAB versus CCABG was reinforced by this post hoc analysis of the uncertainty due to missing data. The analysis showed similarity of results produced by the conditional Gaussian and the chained equations approach and importance of care in the choice of model. Evidence about the long term cost-effectiveness of OPCAB versus CCABG is warranted.
Introduction

The DOORS study

For centuries coronary bypass artery grafting (CCABG) has been considered one of the most cost-effective interventions in the world performed in hundred of thousands each year worldwide [1,2]. Total cost in the US is approximately $ 8 billion [2]. Conventional on-pump surgery with cardiopulmonary bypass is the most widely used technique for myocardial revascularization, although advances in surgical techniques, the use of intracoronary shunts, and improvements in epicardial stabilization devices have allowed surgeons to routinely perform multivessel off-pump coronary artery bypass (OPCAB) [3]. In US it is estimated that around 80 % is conducted as CCABG and 20 % OPCAB [4].

Six cost-effectiveness analyses of CCABG versus OPCAB conducted alongside clinical trials have been published [4-9]. All of these studies conclude that OPCAB is significantly cheaper and that there is no significant difference in cardiac outcomes and health related quality of life between the two alternatives. Among the possible explanations of why CCABG stays the preferred surgical technique is that in the randomised trials, high-risk groups, e.g. elderly patients and patients with multiple vessel disease, are under-represented, and that there may be unknown differences in adverse events. Patients operated using OPCAB may have higher risk of incomplete revascularization and graft failure than patients operated with CCABG while OPCAB may eliminate complications related to the heart- lung machine [3,10,11].

The DOORS- Study (Danish On-pump vs. Off-pump Randomization Study) was the largest trial ever of CCABG versus OPCAB powered to detect a significant difference in side-effects with a statistical power of more than 80 %. It was a pragmatic trial of patients aged 70 years and above with surgery performed at the four Danish heart
surgery centers. From January 1st 2005 to November 16th 2008, a total of 1708 patients were screened, of which 900 patients were randomly assigned to either CCABG or OPCAB.

A complete case analysis of cost-effectiveness showed a probability of OPCAB being cost-effective of approximately 88 % thus indicating the possibilities of huge savings through increased usage of OPCAB [12].

The purpose of this study is to address the problem of missing data in the DOORS-study with a multiple imputation approach (reference) and to estimate cost-effectiveness of CCABG versus OPCAB from data including imputed values for missing data. The study is a post hoc comparison of the complete case analysis with new estimates of cost-effectiveness based on the Gaussian multivariate regression method (in STATA 11 known as mi impute mvn) and chained equations or conditional probability method (the user written STATA program ice). Furthermore, we assessed two different strategies for choosing the imputation model i.e. a data-driven respectively a clinical reasoning selection strategy. To our knowledge no study has attempted to compare these different methods in the context of a health economic evaluation.

**Methods for the economic evaluation**

An economic evaluation was conducted alongside the DOORS-study with patient specific data collected prospectively. Analyses were made according to international guidelines for evaluations alongside randomised trials [13-16]. Calculations of cost were calculated as the unit cost for each type of resource usage multiplied with used amount of resources. The perspective was national health care and cost data comprised hospital costs (cost of initial stay including surgery, intensive care unit and bed department, re-admissions and outpatient visits), costs at general practitioners, specialists and medicine. The analysis followed the intention-to-treat principle i.e. costs and health effects was assigned to the group the patient was randomized to
including additional costs and/or savings for conversion of surgical technique from OPCAB to CCABG or from CCABG to OPCAB. Quality adjusted life years (QALY) was chosen as outcome measure and estimated using EQ-5D at time of admission and 180 days. QALYs were for each patient calculated as the 'area under the curve' [13] using Danish TTO weights [17]. The time horizon was a half year. All calculations were made in Danish Krone (DKK) at price-level 2008 and converted afterwards into British Pounds (GBP) at the exchange rate GBP 100 for DKK 856.64. Detailed assumptions about unit costs and calculated costs for each type of resource are presented elsewhere [12].

The result of the economic evaluation was presented as mean costs and mean QALYs and an incremental cost-effectiveness ratio (ICER) for the complete case analysis, the Gaussian multivariate normal approach and the chained equations approach (table 1). Point estimates of costs and QALYs were calculated as bias corrected non-parametric bootstrapping estimates (1000 replications), and QALYs were adjusted for baseline differences in EQ-5D score [18].

The uncertainty around the incremental costs and QALYs was assessed from 4000 bootstrap replications of the original data set (complete cases) as well as replications of the datasets with imputed values [14]. From these bootstrap replications cost-effectiveness acceptability curves were constructed (see Figure 1).

The problem of missing data in the DOORS-study

A complete case (CC) analysis using only data for those patients in a trial with both observed total costs and outcomes is the most common type of cost-effectiveness analyses [19]. CC analyses involve a loss of information because patients with incomplete data are ignored. Thus, it causes wider confidence intervals
(because of reduced sample sizes) and it reduces statistical power to detect a treatment effect. Furthermore, if missing data represent subgroups of patients in the trial that differ in systematic ways from the rest of the sample, it can severely bias treatment effect. However, a CC analysis may work satisfactorily, if the missingness is “completely at random” (MCAR), meaning that no association exist between the missingness and the outcome (observed or unobserved). A number of studies have shown that under no circumstances should only CC be used in a statistical analysis of cost-effectiveness [19]. In the DOORS-study, complete cases were available for 779 patients, thus there were missing data for 121 patients (61 in the OPCAB and 60 in the CCABG arm). From a total number of 450 in each group, 389 (86%) of the patients randomized to receive OPCAB and 429 (95%) of the patients randomized to receive CCABG, were operated according to protocol. Thirteen patients were excluded from per-protocol treatment between randomisation and operation because of a revision of the decision to operate (n=2), withdrawal of consent (n=4), logistic reasons (n=6), or mortality (n=1). Except for the patient who died before surgery these 12 patients had missing values for self-reported quality of life and costs, but clinical pre-examination data were available. Another 57 patients did not complete EQ-5D at time of follow-up for unknown reasons, and 30 patients had missing time registrations in operating theatre and/or missing registrations of time spent in either intensive care unit or bed department. Drop out due to death amounted to 40 patients (19 OPCAB, 21 CCABG), however these patients were not treated as missing but was assigned the EQ-5D value 0 at the time of death [13] and costs according to their actual patient pathway.

The characteristics of individuals with missing were compared with those having complete data (e.g. patient characteristics, procedures, observed costs and outcomes, reasons for missingness). Data with a missing value was recoded with a missing-value code (for each parameter) and simple descriptive statistics derived to investigate patterns of missingness. Possible clinical reasons as well as administrative reasons for missingness were evaluated and discussed in the research group in order to adopt a suitable analysis.
**Methods to handle missing data**

Imputation methods predict and fill in the missing values based on the observed data. Multiple imputation has been recommended as superior to single imputation methods [16]. The idea of multiple imputation is that instead of filling in missing values to create a single imputed dataset, several imputed data sets are created each containing different imputed values. The analysis is then done on each of the imputed data sets and combined into a single set of results. The major advantage is that it produces standard errors that reflect both within- and between-sample variation. Furthermore, multiple imputation is often better than many simple techniques that can bias results and provide too narrow confidence intervals. For instance, last-value carried forward frequently introduces bias by imputing overly favorable estimates for subjects with missing data due to morbidity, and average-value carried forward may underestimate standard errors [20,21].

The use of multiple imputation techniques in health economic evaluation is still quite new and there are ongoing discussions about which techniques to use, and how these techniques should be implemented in a cost-effectiveness-setting [22,23]. In this study we compared two different approaches to creating multiply imputed datasets: The Gaussian approach is based on the joint multivariate normal distribution of all the variables in the imputation model, including variables to be imputed and variables to be used only for the purpose of imputing other variables. The chained equation approach is based on each conditional density of a variable given other variables. This approach is not as theoretically sound as the multivariate normal approach, but simulation studies have shown that it performs well in practice [24].

The condition for an imputation approach to be consistent is that the missingness is “at random” (MAR), which somewhat counter intuitively differs from MCAR in that association between missingness and outcome is
allowed, if independency can be obtained through conditioning with a selection of additional variables (the predictors or the model). In a less hairy wording observed outcomes may differ systematically from the unobserved, if a model can be constructed (using the observed data) providing unbiased estimates of the unobserved outcomes.

Two important though truly obvious recognitions are that: A given models’ ability to predict the unknown outcomes cannot be formally evaluated. And in fact a prediction working well on the observed data may be seriously biased on the unobserved part, since we are considering a situation with systematic difference between the former and the latter.

We therefore applied a Gaussian and a chained equations approach to multiple imputation under the two different strategies for selecting parameters for our imputation model: The first strategy was to select parameters solely on clinical and administrative reasoning focusing on avoiding biased estimates. We chose to build the imputation model with the relevant variables from the Euroscore, which have been shown to be the best predictor of outcome after bypass surgery [25]. (The following patient-related variables were chosen: age, sex, chronic pulmonary disease, extracardiac arteriopathy, neurological dysfunction, previous cardiac surgery, creatinine > 200 µmol/ L, active endocarditis, critical preoperative state, and the following cardiac-related factors: Unstable angina LV function, recent myocardial infarct, pulmonary hypertension. No operation-related factors were applied because they all got the same treatment i.e. CCABG or OPCAB). We used separate imputation procedures for self-reported outcomes and costs, and because missing data in the imputation is drawn from a multivariate normal distribution we log transformed costs before imputation and back-transformed datasets with imputed values.
The second strategy was to select variables in the imputation model from a data driven approach keeping both significant and insignificant independent variables, and both variables that were associated with outcome and with missingness. We included as many covariates as possible (N=26) to obtain the best fit including center variables (dummies), co-morbidity (e.g. diabetes and cardiovascular disease, base-line data (e.g. age, gender, socio-economic variables and self-reported quality of life), life-style related variable (e.g. smoking, alcohol, BMI), and outcome variables (costs and self-reported outcome). Again we log transformed cost data and performed the imputation in separate procedures for costs and outcomes.

All analyses were conducted in STATA version 11.

**Results**

Table 1 shows the point estimates of costs and outcomes (QALYs) from the complete-case analysis compared with the multiple imputation based analyses. It shows that multiple imputation have a large impact on the point estimate of mean costs, mean QALYs and ICER. Generally, the effect of multiple imputation was an increase in the mean costs in the OPCAB group and a decrease the mean QALY in both arms. This effect of imputation was expected because there were more elderly and more women among the group of patients with missing data especially in the OPCAB group.

A noteworthy result is that the mean QALYs are generally estimated slightly higher with the first imputation strategy (MI and ICE) than the second imputation strategy (MI_d and ICE_d). The choice of imputation model seems to have a larger impact on the ICER than the choice of imputation approach.
Figure 1 shows the cost-effectiveness acceptability curves (CEAC) from the complete-case analysis compared with the multiple imputation based approach. The effect of the multiple imputation on cost-effectiveness acceptability was that it shifted the CEAC upwards.

Assuming a cost-effectiveness threshold for willingness to pay for a QALY of £30.000 the probability of OPCAB being cost-effective was 73-75 % depending on the choice of methods for multiple imputation and the strategy for implementation. Compared to the 88 % likelihood of cost-effectiveness it seems that the complete-case analysis was slightly biased against CCABG.

Figure 1 also shows that there were only smaller differences between the CEAC based on the Gaussian approach and the chained equations approach to multiple imputation. Again the strategy for parameter selection in the imputation model had a larger visual effect on the CEAC curve than the imputation approach.

**Discussion**

The analysis of cost-effectiveness acceptability confirms that OPCAB on average is the most cost-effective procedure in the short term. Addressing the problem of missing data in the DOORS-study by a multiple imputation approach revealed that the complete-case analysis slightly overrated the cost-effectiveness of OPCAB. We found the probability of OPCAB being cost-effective at a threshold value of £30.000 of 73-75 % compared to 88% in the complete-case analysis. The complete case analysis did not allow us to use all the observed data and thus, it did not correctly reflect the cost-effectiveness of OPCAB versus CCABG. This effect of multiple imputation was expected because our initial investigation of the data revealed that missing data were slightly more prevalent among elderly and women (i.e. weaker patients) especially in the OBCAG arm.
Ischemic heart disease is a major cause of morbidity and mortality throughout the world and choosing the most cost-effective strategy for coronary bypass surgery could lead to a much more efficient allocation of scarce financial resources in health care. The results of the DOORS-study are especially important in this respect, because it is the largest trial ever conducted and it is pragmatic in the sense that more elderly patients with multivessel disease are included. Evidence of long term differences in costs and outcomes is of course needed.

The results of our analyses showed how multiple imputation affected point estimates of mean costs, QALYs, ICER, and the position of the CEAC. Especially, this study shows that it is important to perform the multiple imputation, but also how this is conducted. The economic literature on imputation for cost-effectiveness analyses focus on the choice of statistical approach, but little has been written about the strategy for choosing the imputation model. Since, by definition, we do not know what provides the most correct prediction we should advice against purely data driven model building and recommend researchers to perform sensitivity analyses of different strategies for multiple imputation to investigate in what way this affects the ICER.

Although much can be done to avoid missing data, missing data in economic evaluations conducted alongside trials are inevitable [16]. Multiple imputation is therefore attractive in cost-effectiveness analyses, and methods of multiple imputation are especially valuable when the proportion of missing data is high.

Today, many cardiovascular researchers feel more confident with complete case analyses. It is easier to understand and to compare with earlier publications, and in cases where missing data amounts to only a small fraction of data (e.g. below 5 %) the complete case analyses may provide trustworthy results [15]. Statisticians and economists working together with clinical researchers should emphasize the importance of imputation for
assessing cost-effectiveness. More research on the way imputation is done in cost-effectiveness analyses is warranted.
References


Table 1. Point estimates for mean costs, QALY and ICER

<table>
<thead>
<tr>
<th></th>
<th>Costs</th>
<th>QALY</th>
<th>ICER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CABG</td>
<td>OPCAB</td>
<td>Difference</td>
</tr>
<tr>
<td>CC</td>
<td>17.379</td>
<td>16.184</td>
<td>1.195</td>
</tr>
<tr>
<td>MI</td>
<td>17.368</td>
<td>16.852</td>
<td>516</td>
</tr>
<tr>
<td>ICE</td>
<td>17.391</td>
<td>16.843</td>
<td>548</td>
</tr>
<tr>
<td>MI_d</td>
<td>17.414</td>
<td>16.864</td>
<td>550</td>
</tr>
<tr>
<td>ICE_d</td>
<td>17.379</td>
<td>16.895</td>
<td>484</td>
</tr>
</tbody>
</table>

CC Complete-case
MI Multiple imputation approach with imputation strategy based on clinical reasoning
ICE Chained equation approach with imputation strategy based on clinical reasoning
MI_d Multiple imputation approach with imputation strategy based on a data driven model
ICE Chained equation approach with imputation strategy based on a data driven model
Figure 1.

Cost-effectiveness acceptability curves (CEAC)

CC Complete-case
MI Multiple imputation approach with imputation strategy based on clinical reasoning
ICE Chained equation approach with imputation strategy based on clinical reasoning
MI_d Multiple imputation approach with imputation strategy based on a data driven model
ICE_d Chained equation approach with imputation strategy based on a data driven model