Evaluation of cardiac surgery mortality rates: 30-day mortality or longer follow-up?1

Sabrina Siregar*, Rolf H.H. Groenwoldb, Bas A.J.M. de Molc, Ron G.H. Speekenbrinkd,
Michel I.M. Versteeghe*, George J. Brandon Bravo Bruinsmaf, Michiel L. Botsb,
Yolanda van der Graa, Lex A. van Herwerdena

* Corresponding author. Department of Cardio-Thoracic Surgery, University Medical Center Utrecht, Heidelberglaan 100, E03.511, P.O. Box 85500, 3508 GA Utrecht, Netherlands. Tel: +31-88-7556179; fax: +31-88-7555058; e-mail: s.siregar@umcutrecht.nl (S. Siregar).

Received 8 October 2012; received in revised form 20 December 2012; accepted 8 January 2013

Abstract

OBJECTIVES: The aim of our study was to investigate early mortality after cardiac surgery and to determine the most adequate follow-up period for the evaluation of mortality rates.

METHODS: Information on all adult cardiac surgery procedures in 10 of 16 cardiothoracic centres in Netherlands from 2007 until 2010 was extracted from the database of Netherlands Association for Cardio-Thoracic Surgery (n = 33 094). Survival up to 1 year after surgery was obtained from the national death registry. Survival analysis was performed using Kaplan–Meier and Cox regression analysis. Benchmarking was performed using logistic regression with mortality rates at different time points as dependent variables, the logistic EuroSCORE as covariate and a random intercept per centre.

RESULTS: In-hospital mortality was 2.94% (n = 972), 30-day mortality 3.02% (n = 998), operative mortality 3.57% (n = 1181), 60-day mortality 3.84% (n = 1271), 6-month mortality 5.16% (n = 1707) and 1-year mortality 6.20% (n = 2052). The survival curves showed a steep initial decline followed by stabilization after ~60–120 days, depending on the intervention performed, e.g. 60 days for isolated coronary artery bypass grafting (CABG) and 120 days for combined CABG and valve surgery. Benchmark results were affected by the choice of the follow-up period: four hospitals changed outlier status when the follow-up was increased from 30 days to 1 year. In the isolated CABG subgroup, benchmark results were unaffected: no outliers were found using either 30-day or 1-year follow-up.

CONCLUSIONS: The course of early mortality after cardiac surgery differs across interventions and continues up to ~120 days. Thirty-day mortality reflects only a part of early mortality after cardiac surgery and should only be used for benchmarking of isolated CABG procedures. The follow-up should be prolonged to capture early mortality of all types of interventions.

Keywords: Cardiac surgery • Outcomes • Statistics • Risk analysis/modelling • Health policy • Database

BACKGROUND

Mortality is the most commonly used outcome measure in cardiac surgery. The fact that it is unambiguous and relatively easy to determine makes it an appealing measure for outcome evaluation. There is no consensus on the optimal period during which to assess mortality related to cardiac surgery. Often in-hospital or 30-day mortalities are used, but some have opted for longer follow-up periods varying from 60 days up to 6 months [1–5]. Previous studies comparing these outcome measures led to varying conclusions. Where some studies conclude that in-hospital and 30-day mortality are nearly identical, others show that evidently lower rates are measured when in-hospital mortality is compared with 30-day mortality [3, 6]. If differences between mortality measures exist, the results of outcome evaluation or benchmarking might depend on which mortality measure is compared across hospitals, i.e. after which time interval mortality is measured. There have been studies implying that the use of 30- or 180-day mortality after coronary artery bypass grafting (CABG) would not alter benchmarking results [7]. However, the topic remains frequently debated whenever outcomes are evaluated. To our knowledge, our study is the first to investigate the impact of the use of different outcome measures on benchmarking.

1Presented at the 26th Annual Meeting of the European Association for Cardio-Thoracic Surgery, Barcelona, Spain, 27–31 October 2012.
using clinical data. The aim of our study was to investigate early mortality after cardiac surgery and determine the most adequate follow-up period for the evaluation of mortality rates.

**METHODS**

**Data**

Information was extracted from the database of Netherlands Association for Cardio-Thoracic Surgery (NVT). All records of adult cardiac surgery in 10 of 16 cardiothoracic centres in Netherlands from 1 January 2007 until 31 December 2010 were used, which comprised 34,234 surgical procedures. There were 213 (0.6%) records with one or more missing EuroSCORE variables, which were excluded, leaving 34,021 complete cases for further analyses. The dataset consisted of demographic characteristics, details on the intervention, in-hospital mortality and risk factors for mortality after cardiac surgery, notably EuroSCORE variables [8].

**Patient follow-up**

Survival status was obtained from the national death registry by linkage of data. Linkage was facilitated by Statistics Netherlands [9]. All 10 centres consented to the linkage. All deaths occurring up to 31 December 2011 were extracted, including cause and place of death. Linkage between the NVT data and the death registry was performed using matching based on date of birth, sex and postal code at the time of surgery. The sensitivity of this matching procedure was 97.3% (927 patients could not be matched). This means that a minimum follow-up of 1 year could be performed for in total 33,094 interventions (among which 620 reoperations performed in patients previously included in the database). All analyses were performed at an intervention level, meaning 1 patient could be counted multiple times in case of reoperations.

**Early mortality measures**

The following early mortality measures were calculated: in-hospital mortality (mortality in-hospital where cardiac surgery was performed and in same admission), 30-day mortality (mortality within 30 days after cardiac surgery, regardless of place of death), operative mortality (in-hospital mortality or mortality within 30 days after cardiac surgery), 60-, 90-, 120-, 180-day and 1-year mortalities (mortality within 60-90-, 120, 180 days and 1 year after cardiac surgery, regardless of place of death). Mortality at fixed time intervals includes all mortality up to that point, including all causes and irrespective of location.

**Survival and hazards**

The risk of mortality after surgery at any given time can be expressed as the instantaneous hazard. It can be calculated by dividing the number of deaths by the number at risk at any time during the follow-up and thus represents the risk of mortality at that moment. Hazards and survival functions are different ways to describe time-to-event data (in this case time-to-death), but in fact give the same information: the survival at any point in time is 1 minus the cumulative hazard up to that point. The instantaneous hazard after cardiac surgery varies with time. Survival functions are calculated using all deaths as events, including all causes and irrespective of location.

**Risk adjustment and benchmarking**

For the benchmarking procedure, risk adjustment was performed using the logistic EuroSCORE. This model is the most commonly used risk-adjustment method in Europe and its definitions are used in the NVT database. The logistic EuroSCORE was calculated for each patient. Subsequently, benchmarking was performed using each early mortality measure. A random effects model was fitted with one of the mortality measures as the outcome variable and including the logistic EuroSCORE as covariate. A random effects model accounts for within-hospital variability and between-hospital variability and is the preferred type of regression model used for comparison between centres [10, 11]. This regression model thus assumes that mortality is partly explained by patient characteristics (i.e. disease severity quantified by the EuroSCORE) and partly by a centre effect, which is specific to each centre and can be compared across centres [12, 13].

**Analyses**

All abovementioned early mortality measures were calculated. Non-parametric survival analysis was performed using the Kaplan–Meier method [14]. The survival rate in the age-matched general population was calculated and compared with the survival rate in our study population. Survival functions were calculated for all cardiac surgery and for different strata of preoperative risk, quantified by quartiles of the logistic EuroSCORE values. In addition, survival functions stratified by type of intervention were calculated. Also, the survival function for all cardiac surgery was calculated using only cardiac mortality.

Risk-adjusted survival functions were calculated using the Cox Proportional Hazard method with the logistic EuroSCORE as a covariate [15]. Survival functions corrected for logistic EuroSCORE were calculated while stratifying for type of intervention and centre. In addition, time-dependency of the effect of logistic EuroSCORE and the effect of intervention type were investigated by testing if the coefficients were constant in time (slope = 0), indicating proportional hazards.

A random effects model with one fixed intercept and a random intercept for each centre was modelled. The random intercepts were compared with the overall random intercept value of 0. Statistical uncertainty was addressed by estimating 95% confidence intervals (CIs) of the random intercepts for all centres using the posterior variances [11]. Centres with a CI of the random intercept that does not cover 0 are identified as statistical outliers. Random intercepts above 0 reflect higher-than-expected mortality rates; those <0, lower-than-expected mortality. Benchmarking using regression analysis was repeated for each of the eight mortality measures. All analyses were performed in R version 2.12 [16]. The code is available on request.

**RESULTS**

**Overall mortality rates and survival**

In total data on 33,094 interventions were extracted from the NVT national database. The study population is described in
Table 1. The total follow-up time after intervention was 90 386.6 years and the mean follow-up time was 996.9 days. Early mortality rates using the different measures are presented in Fig. 1. Mortality after discharge from the primary hospital was doubled after 1 year: from 972 deaths (2.94%) to 2052 deaths (6.20%). In-hospital and 30-day mortalities were nearly identical. However, in Table 2 the difference between these outcome measures is shown. Approximately 20% of all deaths during admission occur after 30 days. The other way around holds true as well: 20% of all deaths within 30 days occur at home or at another care facility.

The Kaplan–Meier survival analysis of all cardiac surgery is shown in Fig. 1. The survival curves of the cardiac surgery population and that of the general Dutch population run parallel to each other from \( \approx 120 \) days onwards. The mortality rate in the remainder of the first year is 0.065 (95% CI of 0.060–0.070) deaths per 1000 person-days and is comparable with the mortality rate in the age-matched general population of 0.06 deaths per 1000 person-days. The hazard function in Panel B seems to stabilize after the same period. Analyses using only cardiac mortality yielded similar results.

**EuroSCORE and survival**

Figure 2 shows the survival curves for each quartile of EuroSCORE (with inter-quartile logistic EuroSCORES of 1.94/3.74/7.48%) and the accompanying hazard functions. This figure shows that the risk of dying is higher with high logistic EuroSCOREs. This holds true for the whole follow-up period of 1 year. In the low EuroSCORE stratum, most mortality occurs in the first 60 days postoperatively, whereas in the stratum with the highest EuroSCORE, most mortality occurs in the first 120 days. This means that the duration of the early phase of the hazard after cardiac surgery depends on the preoperative risk. The effect of the EuroSCORE appeared to be time-dependent (slope of coefficient −0.212, \( P < 0.0001 \)). This means that the effect of

---

**Table 1: Characteristics of study population**

<table>
<thead>
<tr>
<th>Risk factors for mortality after cardiac surgery</th>
<th>n %, ( N = 33,094 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (continuous)</td>
<td>65.8 (±11.3)</td>
</tr>
<tr>
<td>Female</td>
<td>9911 (29.9)</td>
</tr>
<tr>
<td>Serum creatinine &gt;200 μmol/l</td>
<td>4043 (12.2)</td>
</tr>
<tr>
<td>Extracardiac arteriopathy</td>
<td>3731 (11.3)</td>
</tr>
<tr>
<td>Neurological dysfunction</td>
<td>991 (3.0)</td>
</tr>
<tr>
<td>Previous cardiac surgery</td>
<td>2354 (7.1)</td>
</tr>
<tr>
<td>Recent myocardial infarct</td>
<td>4098 (12.4)</td>
</tr>
<tr>
<td>LVEF 30–50%</td>
<td>5043 (15.2)</td>
</tr>
<tr>
<td>LVEF &lt;30%</td>
<td>1699 (5.1)</td>
</tr>
<tr>
<td>Systolic pulmonary pressure &gt;60 mmHg</td>
<td>800 (2.4)</td>
</tr>
<tr>
<td>Active endocarditis</td>
<td>538 (1.6)</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>1894 (5.7)</td>
</tr>
<tr>
<td>Emergency operation</td>
<td>1899 (5.7)</td>
</tr>
<tr>
<td>Critical preoperative state</td>
<td>1352 (4.1)</td>
</tr>
<tr>
<td>Ventricular septal rupture</td>
<td>61 (0.2)</td>
</tr>
<tr>
<td>Other than isolated coronary surgery</td>
<td>15,324 (46.3)</td>
</tr>
<tr>
<td>Thoracic aortic surgery</td>
<td>1682 (5.1)</td>
</tr>
<tr>
<td>Logistic EuroSCORE</td>
<td>mean 6.8 (±9.3), median 3.7</td>
</tr>
</tbody>
</table>

**Types of intervention**

| CABG                                          | 22,696 (68.6) |
| Isolated CABG                                 | 17,711 (53.5) |
| Valve                                         | 12,262 (37.1) |
| Isolated valve                                | 6653 (20.1)   |
| Aortic                                       | 4128 (12.5)   |
| Mitral                                       | 1466 (4.4)    |
| Double valve                                  | 861           |
| Other                                         | 198           |
| CABG and valve (and other cardiac surgery)    | 4500 (13.6)   |
| Aortic                                       | 2746 (8.3)    |
| Mitral                                       | 1198 (3.6)    |
| Double valve                                  | 476           |
| Other                                         | 80            |
| Aortic surgery                                | 1685 (5.1)    |

LVEF: left ventricular ejection fraction.

Figure 1: Kaplan–Meier survival curve with 95% CI after cardiac surgery. The green line represents the survival rate of the age-matched general population in The Netherlands. The survival rate of the cardiac surgery population equals that of the general population from approximately 120 days after surgery onwards. The hazard (risk of mortality) after cardiac continues to decline well after 30 days postoperatively. The constant phase of the hazard seems to start after \( \approx 120 \) days.
the logistic EuroSCORE (i.e. the preoperative risk) on mortality decreases with time.

Survival across types of interventions and across centres

Figure 3 shows the risk-adjusted survival and hazard functions, stratified in the following intervention groups: isolated CABG, isolated valve, CABG and valve and other cardiac surgery. The curves all correspond to a patient with the median logistic EuroSCORE value of 3.74%. Stabilization of hazards is seen after a varying period of time. The hazard in the isolated CABG subgroup appears to reach the constant phase much earlier than the other intervention groups, approximately after 60 days. For the isolated valve subgroup and for the CABG and valve group subgroup, this takes \( \approx 90 \) and \( \approx 120 \) days, respectively. The effect of the intervention group appeared to be time-dependent (slopes of coefficients \(-0.04, -0.06, -0.12, P = 0.01, P < 0.001, P < 0.001\)). This means that the effect of the type of intervention decreased with time.

Figure 4 shows the risk-adjusted survival and hazard functions for the 10 centres. The curves all correspond to a patient with the median logistic EuroSCORE value of 3.74%. Again, stabilization of hazards is seen after a varying period of time, even when risk adjustment is performed. For example, in the centre corresponding with the dark blue line, the hazard appears to stabilize earlier than in the other centres. Overall hazards reached the constant phase at \( \approx 120 \) days.

Benchmarking using different outcome measures

The effect of using different outcome measures on the benchmarking procedure is shown in Fig. 5. When in-hospital mortality (blue symbols) is used as outcome, one low mortality outlier (Centre A) and two high mortality outliers are found (Centres I and J). However, by using 30-day mortality as outcome measure, two other centres are identified as outliers as well: Centre B as a low mortality outlier and Centre E as a high mortality outlier. Prolonging follow-up from 30 days to 1 year leads to changes in outlier status in four hospitals (Centres B, C, H and J). When the same is done for a subset of isolated CABG procedures, benchmarking results remain unchanged with the different follow-up periods. This is shown in Fig. 6.

DISCUSSION

Main findings

We used survival status after all adult cardiac surgery procedures in Netherlands from 2007 to 2010 to study early mortality after cardiac surgery. The differences between mortality rates during hospital stay, after 30 days or after longer intervals were assessed and benchmarking was performed using these different outcomes. The slope of the survival function after cardiac surgery continues to decline many days after the usual 30-day cut-off point for evaluation. The decline of the slopes depends on the...
performed intervention: isolated CABG procedures seem to reach a stable phase after \( \approx 60 \) days, whereas valvular or combined interventions maintain a higher hazard for a longer period. Similarly, the decline of the slopes also depends on the preoperative risk of mortality, as measured with the logistic EuroSCORE. Benchmarking using in-hospital mortality, 30-day mortality or longer fixed-period mortality rates yield different results. Average-mortality centres could change into outliers when the follow-up is extended up to 1 year and vice versa. From \( \approx 120 \) days onwards the hazards had reached a constant phase for all the types of interventions. Benchmarking of different types of interventions should therefore not be performed before this period. When the follow-up is shorter, it is recommended that benchmarking is only performed in isolated CABG procedures.

### Differences in outcome measures

Our results show that mortality rates can vary largely depending on the cut-off point used for the follow-up. Mortality at discharge or at 30 days is more than doubled after 1 year. Previous studies found similar results. Edwards and Taylor [17] studied over 80,000 patients in the UK Heart Valve Registry and found almost a doubling of the mortality rate after 1 year, when compared with 30 days.
Benchmarking results are unaffected by the choice of the follow-up period. Interventions from 2007 until 2010 were included. Types of end points and are not interchangeable.

In-hospital mortality and 30-day mortality measure two different types of end points and are not interchangeable. Patients who die in the hospital for a long time and eventually die. The latter type of mortality is more likely to be influenced by preoperative comorbidities than the former [2]. Thus, in-hospital mortality and 30-day mortality measure two different types of end points and are not interchangeable.

Hospital or fixed interval mortality?

Hospital mortality rates depend on the postoperative transfer policy of patients to other health care facilities or back to the hospital. A large study comparing hospital mortality, 30-day mortality and operative mortality rates in CABG reported similarly [6]. The authors concluded that because the numbers are nearly identical, the more convenient outcome in-hospital mortality could be used for outcomes evaluation. However, although the in-hospital and the 30-day mortalities are equivalent in numbers, they do not refer to the same patients: 20% of the patients counted in each measure are not included in the other measure. This difference is relevant because patients who die within 30 days (either in the hospital or elsewhere) are likely to be different from those who remain in the hospital for a long time and eventually die. The latter type of mortality is more likely to be influenced by preoperative comorbidities than the former [2]. Thus, in-hospital mortality and 30-day mortality measure two different types of end points and are not interchangeable.

The effect of case mix on survival and benchmarking

The duration of follow-up (i.e. the use of different outcome measurements) has clear consequences on the benchmarking results. In our empirical data, benchmarking of in-hospital mortality yields other outliers than that of 30-day or 1-year mortality. Centre H is initially benchmarked as an average centre, but becomes an outlier when the follow-up is extended to ≥60 days. The opposite occurs in Centre J. The hospital is initially identified as a high-mortality outlier, but converts its outlier status into average after 120 days. Changing positions with relation to the benchmark reflects the crossing of hazard curves (i.e. have not reached a steady state yet). Our results show that survival differs for each type of intervention. As a result, the total hazard curve of a hospital depends on the type of interventions performed. Thus, the underlying mechanism causing the observed changes in outlier status is probably the difference in the performed types of interventions. When benchmarking is performed only with isolated CABG procedures, the results are unaffected by the choice of the follow-up period. This is best illustrated using the following example: a hospital with mainly isolated CABG procedures will have 60- and 90-day mortality rates that will be nearly comparable. After all, the hazard for isolated CABG has nearly reached a steady state after 60 days, meaning mortality rates will not rise much after 60 days in both hospitals. In contrast, a hospital where mainly combined CABG and valve procedures are performed will have a 60-day mortality rate that will be considerably higher than the 90-day mortality rate. As shown in Fig. 3, this is due to the fact that the hazard is still on the steep decline between 60 and 90 days. For this hospital, an evaluation of mortality rates across centres is therefore much more beneficial after 60 days than after 90 days. When the goal is to evaluate early mortality, a follow-up of 90 days or even longer is clearly more adequate in this example to ensure early mortality is captured as completely as possible. If only isolated CABG procedures are compared, most mortality already occurs in the first 30 days and the difference with the longer follow-up periods is expected to be small. This is confirmed in Fig. 6, where benchmarking results are similar using a follow-up of 30 days and longer periods. A Health Services Research study following 5000 CABG patients for 6 months...
concluded similarly. When observed minus expected mortality rates at 30 days and at 180 days were compared, the ranking lists composed using these two outcomes hardly differed [7].

What is the adequate follow-up interval for benchmarking?

Having shown the differences between the mortality measures and the effects on benchmarking, the next question is to decide which measure to choose. Considering the arguments mentioned above, it would be logical to take the longest follow-up possible before benchmarking is performed. However, a long follow-up has several downsides. First, the follow-up of patients after discharge is time-consuming and requires effort and money. This problem should not be underestimated, since an incomplete follow-up might lead to biased results [20]. Secondly, the question is what it is that needs to be measured; different mortality rates might reflect other processes. For instance, patient compliance to medication, quality of home care, extent of involvement of the cardiologist and many other factors have an increasing effect on the risk of mortality after discharge. On the contrary, the effect of the initial care provided around the intervention in the hospital is likely to decrease in time. Therefore, it seems less adequate to measure mortality after 1 year or longer, when it is the process surrounding the cardiac surgery in which one is interested. In 1986, Blackstone et al. suggested that the hazard can be subdivided into an early, constant and a late phase [21, 22]. In benchmarking of outcomes, the early phase reflects the part of the process of care that we want to evaluate.

Sergeant and Blackstone found an early phase hazard that lasted for 6 months after CABG, suggesting that the follow-up should be extended to a half year after surgery [4]. Other studies evaluating early mortality after CABG found that early mortality occurred up to ∼60 days [2]. The authors expected this interval to be even longer in procedures performed in more recent years. We found similar results for isolated CABG procedures. For other interventions stabilization of hazards was seen after ∼120 days. Based on these findings, it seems advisable to extend follow-up to a period beyond the commonly used 30 days. In case longer follow-up is not attainable, evaluation of mortality rates should only be performed within specific procedures.

Possible limitations

Ten of the sixteen centres performing cardio-thoracic surgery in Netherlands consented to the linkage of data with the national death registry. This resulted in a comprehensive multi-centre database of all types of cardiac surgery, including a follow-up of 1 year or longer. However, this also means that approximately a third of all cardiac surgery procedures in Netherlands (in the non-participating centres) from 2007 to 2010 were not included in our analyses. Differences between the population treated by the participating and non-participating centres could theoretically affect the generalizability of the results. However, a comparison of risk factors showed no significant differences between our study population and the six other cardiac surgery centres (results not shown, available on request). Results are therefore likely to be generalizable to other populations.

In addition, the matching performed to establish a linkage to the national death registry had a sensitivity of 97.3%. Unmatched individuals were removed from further analyses. Baseline characteristics of these patients were comparable with the matched patients, with no significant difference in in-hospital mortality ($P = 0.494$) and logistic EuroSCORE ($P = 0.174$). Therefore, we assume that this constituted missing completely at random and therefore removal is unlikely to have biased our results.

Analyses were performed at an intervention level, meaning that the death of one patient was counted twice in case of a second heart operation. The NVT database does not contain person-identifying variables and analyses at a patient level are therefore not possible. Although in this study, interventions could be linked to individual patients using the national death registry, the NVT database would otherwise not have had this option. To maintain the consistency of methods, we chose to perform survival analyses at an intervention level as well. In total, 620 reoperations were performed. Considering our large study population, we assume the influence on our results was minimal.

CONCLUSION

The course of early mortality after cardiac surgery differs across interventions and continues up to ∼120 days. Thirty-day mortality reflects only a part of early mortality after cardiac surgery and should only be used for benchmarking of isolated CABG procedures. To capture early mortality of all types of interventions, follow-up must be prolonged.

Funding

The Department of Cardio-Thoracic Surgery University Medical Center Utrecht has received financial support from the
REFERENCES


APPENDIX. CONFERENCE DISCUSSION

Dr B. Bidstrup (Brisbane, Australia): There is a lot of work in this, and it sets the standard for getting better data about outcomes. I wonder whether we are now starting to reach the limit of mortality as Bruce Keogh mentioned the other day. Mortality, especially for first-time coronary bypass surgery, is getting lower and lower, and whether it is still a good metric to use remains to be seen.

I think there are a couple of interesting things that come out of this. I think we are starting to see the limits of EuroSCORE, even EuroSCORE II. In the figures that you show in the first benchmarking graph of all-comers, there can be a lot of reasons for Center E, I, and J being outliers. There are a lot of other techniques that can be used to see whether they truly are outliers; they may have a disproportionate volume of really sick cases. I think this is starting to stretch the limits of EuroSCORE. So to address your issue about benchmarking, I think we need to take a leaf out of the STS book and probably out of the UK Registry book as well, such that we need to have benchmarking models for specific procedures.

You have over 50% of coronary bypass, so that is a good procedure to use as a benchmark. If people are doing coronaries badly, then they are probably likely to do a whole lot of other things not as well either. But we need to have a long hard think about how to do the rest of it and can we truly lump them together? I do not think we can, and I wondered whether you would comment on that.

There is one other comment I would make. I notice that there was a bit of double counting. You did mention in the manuscript, that patients could die more than once. Did you have a look in a sensitivity analysis to see whether that did make a difference? I wonder whether it just might have made a difference in one or two of those graphs.

Dr Siregar: I would just like to comment on your first remark about mortality measures as being not the only outcome measure. I fully agree with you. There are many outcome measures that could add to the measurement of quality or safety, such as morbidity. And this falls outside of the scope of this study, but you are absolutely right.

As to your second point, we did the survival analysis using this database, and this database is a procedural database not a patient database. So this means that some patients that were treated twice in the five years were twice in the database, and if they died within that one year and were treated twice in one year, then they were exactly counted as two times mortality. But this was a very small proportion, less than, from the top of my head, approxi-

Dr Kappetein (Rotterdam, Netherlands): Suppose that you used an STS score, which we know is a better predictor especially in the high-risk patients, would this optimize your results or present a different view of the benchmarking results?

The variability is, of course, in the high-risk patients with aortic valve disease, mitral valve disease, aortic disease, etc., not so much in the coronary cohort. If the EuroSCORE is not able to correct for high-risk patients, benchmarking becomes less reliable.

Dr Siregar: I do not think that the problem is in the correction for case mix or in the correction for preoperative risk. I believe the problem lies in the fact that the main part of the mortality is not only in the first 30 days, but it continues far beyond the 30 days. So even -

Dr Kappetein: Well, that is actually an argument to have a better categorization of your patient population in low and high-risk.

Dr Siregar: Well, even if you did that, if most of your people die after 30 days then you would still not catch them if you take 30-day mortality.

Dr Kappetein: No, of course. However, if you collected more variables, you could probably better predict the outcome than if you have fewer variables as in the EuroSCORE I.

Dr Siregar: Yes. That is always true. Yes.

Dr Kappetein: So, therefore, I think it would be a better benchmarking tool if you would use that score.

Dr Siregar: If you used a better score in addition to a longer follow-up, that would be perfect.
Dr Bidstrup: If I could just make one last comment again, then. I think what we are doing here is actually very important. We are now plotting out the natural history of disease because coronary artery disease nowadays very rarely goes untreated; it is rare that it is untreated in some way or another. But now we are seeing the natural history of aortic valve disease, most of which gets some sort of procedure done to it. We are actually at least out to one year which means that, again, it begs the question, should we be categorizing these things or should we be making them a continuous variable? And then that fits in again with comments that Dr Mack made earlier about the dynamic modelling, etc.

Dr E. Daeter (Nieuwegein, Netherlands): I have a little question because I think the opposite. It is truly sure that EuroSCORE actually predicts well, that is shown by these figures, that the decline until 30 days is accurate. At 30 days you show in your graph that the real mortality is at the level, and it does not change further on. So in that sense, the 30-day mortality is well predicted by this EuroSCORE I think.

Dr Kappetein: Yes, of course, in certain patient groups you will have a better prediction than in other patient groups. But, especially in the high-risk patient groups where there is a lot of variability between patients, EuroSCORE is not predicting very well.

Dr Daeter: No.

Dr Kappetein: So, therefore, you need another score especially in those patients. So the points in your conclusion are excellent. It is that I have some doubts about whether you can use this score for benchmarking.