**Statistical Editor:**  
The authors have provided lengthy and mostly appropriate responses.  
  
1. Figure 4B: This mislabeled figure has confused the reviewers! I believe this is an estimated hazard function for mortality. Therefore it is not “crude probability of death” but Death (%/yr). I also suggest superimposing the German age-sex survival curve represented by this hazard function as a reference in Figure 4A.

-- Thank you. Fig.4B is relabeled as Death Rate (%/year). We have also added the German population age/sex adjusted line in Figure 4A.

2. With respect to the control chart (SPRT, Figure 3), first, what variables were used for the adjustment, just the logistic EuroSCORE and nothing else? The logistic EuroSCORE switched from EuroSCORE to EuroSCOREII; was one score only used or, in the midst of the series, did the score switch?

-- Logistic Euroscore 1 was used.

3. Logistic EuroSCORE in Table 1 is highly skewed. Provide median and selected percentiles.

-- # .10 .25 .50 .75 .90 .95

# 0.880 1.220 2.070 3.860 7.602 11.350  
  
4. Clinical reviewers seem not concerned by the substantial number of patients with multi-vessel coronary disease treated only with LITA-LAD, or am I misinterpreting Table 1? Was lack of complete revascularization a risk factor for late mortality?

-- Piroze ? / add incomplete revasc in the model for late survival.  
  
5. There is a lot of repeating of Results in Discussion unnecessarily.  
  
**Associate Statistical Editor**: While the authors have significantly improved their work, there are several aspects that require amendment.  
1) Authors have now produced a risk adjusted sequential probability ratio testing using the logistic euroscore to monitor 30-day mortality. The plot shows an extraordinary drift towards performance improvement in the last years. As Logistic EuroSCORE is known to have a significant calibration drift, authors should present the E: O ratios across eras to reassure that the positive trend reflect a true effect and not risk inflation due to calibration drift. As STS is a more stable score, have they tried with this score? Do they reach the same conclusions? Please comment.

O/E ratio per era – 1, 0.67, 0.72

2) There are too many figures which are not necessary. Authors are not meant to provide a plot for each stratifying variable. Please remove. The only plot required would have been the comparison between their cohort and the general population but what authors have produced is very confusing.

Would suggest putting some KM plots in the supplement. – era + age figure / diabetes mellitus goes in supplement.

Figure 4 left shows a KM where mortality at 15y is ~70% while in the other plot (right, cohort vs matched German population) authors used cumulative incidence but ultimately numbers should be the same which the survival at 15y in their cohort is 25% (or 2.5%??). Authors have not realised that the scale is 0-1 and not 0-100, and the gap between the two curves is 70% AND NOT 0.7% AS THEY SUGGEST IN THE RESULTS SECTION   
Authors need to present ONE plot only, and make sure they get the number right. From the plot they present, it seems that the cohort has a remarkable increase in mortality which does not really support the long term safety of the procedure! Please look into this and come up with consistent results.

-- present the plot with overall survival in our group and German population.  
  
3) In MV model age was used as categorial variable, why? this caused that age could not be included in the last eras model for instance. Please run MV forcing age as a continuous variable.

-- repeat the models with Age as continuous variable. – redo Age as continuous and incomplete revasc. ---

4) Authors suggest that they have performed a phone call/note follow-up in 90% of cases. However long term is reported only in term of survival. What about MI? REHOSPITALIZATION FOR HF? STROKE?  
  
Reviewer #1: Twenty-year outcomes of minimally invasive direct coronary artery bypass surgery: The Leipzig experience.  
  
In response to the suggestions of the prior reviews the authors have made significant changes to the manuscript. This seems particularly the case regarding changes to the statistical methods in response to comments from the Associate Statistical Editor. Overall it is improved.  
  
I have several additional questions and comments:  
  
Although there was a request from one of the reviewers to increase the amount of information in the figure legends, in the opinion of this reviewer, the revised figure legends are now too lengthy and too descriptive. The reader should be able to clearly understand from the Figure legend what the figure is representing and the outcomes it is demonstrating. I do not think it is necessary to include a full description of the results instead of in the body of the text, nor should it essentially duplicate that which is the body of the text.  
  
The discussion has been shortened and improved, however, it remains longer than necessary. For instance, after noting the trend in logistic EuroSCORE and referencing Table 1, I do not see the need to list all of the individual variables which are more frequent and thus responsible for the higher risk profile in the later time periods.  
  
I believe the authors are correct when they state within the limitations that their program is in fact regarded as a "…high-volume center known for its expertise in minimally invasive cardiac surgery." I will, therefore, again suggest that the still lengthy discussion and explanation which compares the results of this analysis with those from the Italian group in Brescia (Reference 20) seems excessive and unnecessary.  
  
Specific to that, however, I would note that the authors do suggest one difference is the higher incidence of patients with left ventricular dysfunction in their study which they have defined rather liberally as an EF < 60%. I will ask how the authors decided that LV dysfunction would be defined as an ejection fraction less than 60%. – EF < 40%.  
  
It is stated twice in the discussion that the study is the largest in the literature. Once would be sufficient.  
  
With regard to revised figure 4B, the upper interval around the survival curve for the study population (blue line) extends beyond 1. What does this mean?  
  
Reviewer #2: Excellent responses. The MS is much improved.  
  
Reviewer #3: The authors have appropriate responded to my comments/queries and I believe have substantially improved their manuscript.

-- age as cont/ lvef < 40%/incomplete revasc. = add in all models.