Review

Does stentless aortic valve implantation increase perioperative risk?
A critical appraisal of the literature and risk of bias analysis

Bari Murtuza a,c,* , John R. Pepper a, Catherine Jones c , Petros Nihoyannopoulos d ,
Ara Darzi c , Thanos Athanasiou b,c

a Department of Cardiothoracic Surgery, Royal Brompton Hospital, Faculty of Medicine, Imperial College, UK
b Department of Cardiothoracic Surgery, St Mary’s Hospital, Faculty of Medicine, Imperial College, London, UK
c Department of Surgical Oncology & Technology, St Mary’s Hospital, Faculty of Medicine, Imperial College, London, UK
d Department of Echocardiography, Hammersmith Hospital London, UK

Received 17 May 2010; received in revised form 25 July 2010; accepted 5 August 2010; Available online 17 September 2010

Summary

Stentless aortic valve replacement has potential benefits in terms of valve hemodynamics and clinical outcomes, although these may be offset by greater technical complexity of implantation with longer cardiopulmonary bypass and cross-clamp times compared with stented valves. Meta-analyses of the small number of published randomized trials have been limited by their lack of critical synthesis of the literature, including evaluation of the Risk of Bias. Our objective was to determine whether stentless aortic valves increase perioperative risk of mortality. We also examined secondary clinical outcomes of neurological, renal and respiratory complications as well as hemodynamic changes reported by studies following implantation of the two types of aortic prostheses. The methodology used to answer this question was a rigorous meta-analysis of randomized controlled trials, using bias-assessment techniques designed to address limitations of conventional meta-analysis. Our findings show that many of the existing randomized trials have a high or uncertain risk of bias. Analysis of studies with low risk of bias reveals that stentless valves do not increase perioperative risk in terms of 30-day mortality and morbidity though neither do they exhibit benefits in hemodynamics or clinical outcomes compared with stented valves. Larger, more stringent randomized studies would be required to identify any robust clinical difference.

© 2010 European Association for Cardio-Thoracic Surgery. Published by Elsevier B.V. All rights reserved.

Keywords: Stentless; Aortic valve replacement; Risk of bias

1. Introduction

Stentless (SL-) aortic valve replacement (AVR) is an alternative to stented (ST-) AVR with theoretical benefits in terms of valve performance and clinical outcomes. These benefits, however, must be balanced against possible increased perioperative risks, primarily attributable to the more complex implantation technique, often requiring longer cardiopulmonary bypass (CPB) and cross-clamp (CC) times [1]. The first true SL valve insertion was reported in by Ross in 1962 using an aortic homograft [2]. Binet reported the use of SL xenografts in 1965 [3] and David subsequently reported a comparative series of SL versus ST aortic valves in 44 patients [4].

A number of randomized and non-randomized studies have compared SL-AVR with stented AVR, though with conflicting results. Further, many of the randomized studies have small numbers of patients, limiting their power and the significance of the inferences, which may be drawn from them, especially for rare clinical outcomes. In an attempt to address limitations of small individual trials, various meta-analytical reports have been published [5,6]. These latter analyses, however, have been inconclusive and, importantly, have lacked a critical assessment of the studies included. This aspect of evidence assessment and synthesis is very important in current surgical practice, especially when practice guidelines are lacking and performance assessment of index operations has been very dominant in decision making.
An important new development in terms of performing a meta-analysis of randomized control trials (RCTs) is the introduction of the Risk of Bias (RoB) evaluation [7]. This relates to the external validity of a study. Differences in RoB may help to account for heterogeneity of results for a primary outcome of interest (POI) across studies included. Bias, in turn, may be defined as a systematic error or deviation from the truth in results or inferences [7]. RCTs may be performed in a stringent manner, yet still be at risk of bias. The Cochrane Collaboration has devised the RoB analysis as a domain-based evaluation tool designed to examine six parameters: adequate sequence generation; allocation concealment; blinding; incomplete outcome data assessment; freedom from selective reporting; and freedom from other bias. As such, this tool differs from the commonly used Jadad score, which places greater emphasis on the reporting of an RCT and has been used to evaluate surgical randomized studies [8]. Other similar scores have also been developed [9]. It is recommended that studies with a high risk of bias or with an unclear RoB be given lesser weight in the overall meta-analysis.

The aim of the present study was therefore to ascertain whether SL-AVR does indeed increase perioperative risk in terms of clinical outcomes of early valve-related mortality and morbidity by performing a meta-analysis of RCTs with a critical appraisal of the published data, including RoB analysis. We reanalyzed the data using only the highest quality RCTs as defined by their Jadad score and low RoB for POI, and performed sensitivity analysis focusing on subgroups such as the elderly population and those with predominant aortic stenosis (AS) to examine for potential sources of heterogeneity.

2. Patients and methods

2.1. Literature search

All RCTs of SL- versus ST-AVR were identified from the existing literature up until April 2010. A MEDLINE literature search was performed using MeSH terms: 'stentless', 'aortic', 'randomized' and 'homograft'. For the purpose of this work, stentless AVR insertion was defined as implantation of a SL bioprosthesis (xenograft) or allograft. A second-level search through reference lists of the included studies was performed to ensure that no reports were omitted. In addition, the ‘related articles’ function in PubMed was used as a further check of rigor. In cases where multiple studies have been published by the same institution, the larger, more recent or more informative study was included.

2.2. Data extraction

All data extraction was performed from the 18 included studies independently by two reviewers (BM, TA). The following study characteristics were extracted: author, study design, geographical location, period of study, year of publication, number of patients in each group, mean age and male gender percentage. Extracted intraoperative variables included: CC and CPB times; myocardial protection strategy; implantation technique; and type of prosthesis. Primary outcome data included 30-day mortality, 1 month to 1 year valve-related mortality, neurological complications, renal failure and pulmonary complications. We also extracted data for postoperative cardiac-related morbidity: myocardial infarction (MI); cardiac failure/low cardiac output; insertion of intra-aortic balloon pump (IABP); insertion of permanent pacemaker (PPM); and atrial fibrillation (AF). Other morbidity outcome data extracted related to incidence of prosthetic valve endocarditis (PVE), intensive care unit (ICU) and total length of stay (LOS) in hospital. The study was performed according to guidelines for the reporting of meta-analytical data in epidemiological studies (MOOSE statement) [10].

2.3. Appraisal of literature, Jadad and RoB scores

We first assessed the overall comparability of studies in a qualitative manner and documented the exclusion criteria for all included studies. Second, we used two complementary and overlapping approaches to determine the quality of studies. The first was to score each study according to the Jadad Scale, which is a semi-quantitative appraisal of the methodological quality of an RCT [8]. Studies score 1 point up to a maximum of 5 for each of the following criteria: study randomized; study double-blind; withdrawals and dropouts described; description of appropriate method of randomization; and description of blinding protocol. The second method for study appraisal was by determination of the RoB. This is a tool, which has been recently introduced, is a useful means for assessing the rigor of a randomized study, and has been used to evaluate surgical randomized studies [11]. Six criteria are used for this score: adequate sequence generation; allocation concealment; blinding; incomplete outcome data assessed; freedom from selective reporting; and freedom from other bias. Each criterion may be marked as yes/no/unclear. We, in turn, converted this into a score recording +1: yes; −1: no; and 0: unclear. Finally, we computed a composite quality score (CQS) taking into account the Jadad score and the RoB-derived score. We defined a threshold of rigor based on the six bias domains with a score of >3. For determination of the RoB score and CQS, differences in grading between the two reviewers were resolved by consensus.

2.4. Statistical analysis

All published RCTs were included in the analysis. Studies included defined at least one of the outcomes of interest and comprised a previously unreported patient group. All non-comparative studies were excluded as well as those in which outcomes of interest were not documented. Where a value of zero was reported for a particular outcome of interest for both SL- and ST-AVR, this outcome measure was excluded from meta-analysis, as was outcome data where the mean and standard deviation were not stated or possible to calculate. Meta-analysis was performed using the odds ratio (OR) or weighted mean difference (WMD) as the summary statistics for binary or continuous variables, respectively. The analysis was performed according to the recommendations of the Cochrane Collaboration and the Quality of Reporting of Meta-analyses guidelines [12]. An OR or WMD < 1 favors SL-
AVR (defined as the treatment group; ST-AVR defined as the control group). A random-effects model was used as this model assumes variation between studies and is preferable for surgical data as selection criteria and risk profiles for patients differ between centers.

Figures tabulating data in this report show squares indicating point estimates of treatment effect (OR or WMD) with 95% confidence intervals (CIs) indicated by horizontal bars. The size of squares represents the weight attributed to each study and the diamond represents the overall OR or WMD from the pooled studies listed with 95% CIs and is significant \( p \leq 0.05 \) if the diamond does not cross the central vertical line (i.e., outside the 95% CI). A sensitivity analysis for quantitative assessment of heterogeneity was performed. We reanalyzed studies of high quality as defined by the Jadad, RoB or CQS as defined above for the POI, 30-day mortality. We also performed subgroup analysis for studies including only patients with a mean age of \( \geq 70 \) years and those including patients with predominant AS. All data were analyzed using Statistical Package for Social Sciences (SPSS) version 12.0 for Windows (SPSS Inc., Chicago, IL, USA) and Review Manager version 5.0 (The Cochrane Collaboration, Software Update, Oxford, UK).

3. Results

3.1. Studies included in the meta-analysis

A total of 645 published reports were identified and from these, 18 RCTs were included in this work, published between 1994 and 2008 \([13—30]\). The total number of patients in these 18 studies was 1410 with 707 in the SL and 703 in the ST group. Characteristics of the included studies are shown in Table 1. We did not exclude studies reporting outcomes for patients undergoing redo-AVR or those reporting concomitant procedures such as coronary artery bypass at the time of AVR. In performing data extraction for outcomes of interest, there was 100% agreement between the two reviewers as well as complete agreement for scoring according to the Jadad Scale, RoB assessment and CQS determination.

Only four studies included more than 150 patients in total \([13,15,21,25]\) with the largest being that of Lehmann et al. \([21]\). Several studies reported 12-month follow-up for patients \([15,16,18—20,23,25,28]\) though only five reported greater than 18-months follow-up \([17,21,24,26,30]\); the 5-year follow-up data for the studies of Doss et al. and Cohen et al. have been recently published as updates \([31,32]\) to the original studies included in the present work \([16,18]\). With respect to differences in clinical characteristics between included studies, six studies excluded patients with previous cardiac surgery \([13,16,18,19,21,25]\), six excluded patients, who were unstable or emergencies \([13,14,16—19]\), while only that of Totaro et al. was explicit in its exclusion of patients with poor left ventricular (LV) function \([29]\). Notably, six studies excluded patients with aortic root dilatation \([13,16,18,24,25,30]\); Lehmann et al. excluded patients with a known aortic connective tissue disorder \([21]\). All but four studies described details of the surgical implantation technique (Table 1). The majority of studies employed the subcoronary technique for SL-AVR and a supra-annular technique for stented valve implantation. The study by Narang et al. used an inclusion cylinder technique for SL-AVRs \([24]\).

3.2. Jadad, RoB and composite quality scores

In performing Jadad scoring of studies, all studies scored 1 for being RCTs and all had appropriate randomization methods and, therefore, no point was deducted. By definition, no study could be defined as double-blind and, hence, this criterion and method of blinding could not be scored. The maximum score attainable was therefore 3, with only five studies \([13,15,19,25,27]\) with this score (Table 2). Two studies scored 2 \([16,21]\) and the remainder scored 1. Only three studies were matched for Parsonnet or EuroSCORE \([13,19,30]\) though the majority did match patients for the New York Heart Association (NYHA) status and/or LV ejection fraction (LVEF; Table 2). The assessment for RoB is shown in Fig. 1, and the methodological summary graph derived from these data in turn shows that few studies overall score adequately for the criteria in the RoB analysis (Fig. 2).

Finally, only nine studies scored \( \geq 3 \) in the RoB score \([13—16,18,19,21,25,28]\) and only eight scored \( \geq 5 \) for the CQS \([13,15,16,19,21,25,27,30]\).

3.3. Meta-analysis of primary and secondary outcome events

All meta-analytical data were calculated using random-effects models. In terms of 30-day mortality, 14 studies reported data with a total of 1124 patients. We did not find a significant difference in 30-day mortality between the two groups (Fig. 3): OR = 1.37, \( p = 0.39 \) (95% CI: 0.67—2.81) and \( \chi^2 \) of 2.88 (\( p = 0.94 \)). The corresponding incidence of mortality across studies was 3.2% (17/536 patients) and 2.38% (14/588 patients) for SL and ST groups, respectively. There was also no significant difference for valve-related mortality between 1 month and 1 year: OR = 1.21, \( p = 0.62 \) (95% CI: 0.57—2.55) and \( \chi^2 \) of 0.6 (\( p = 0.56 \)). Table 3. Neither was there a difference for valve failure rate within the first year, though only four studies reported this outcome clearly (309 patients: SL 150, ST 159) with OR = 0.77, \( p = 0.73 \) (95% CI: 0.17—3.54) and \( \chi^2 \) of 0.05 (\( p = 0.83 \)). Similarly, the ICU stay and total LOS were not significantly different between groups (Table 3).

Furthermore, we did not find any significant difference for cardiac-related morbidity (MI, PPM rate, AF, IABP) although there appeared to be a tendency for a lower rate of postoperative MI and IABP insertion in the ST group (Table 3). The other outcomes for postoperative renal failure and incidence of PVE were also similar for SL and ST groups (Table 3). Although not significant, there was some tendency for a lower pulmonary complication rate in the SL group, though only 137 patients (SL 67, ST 70) were included across three studies reporting this outcome: OR = 0.33, \( p = 0.35 \) (95% CI: 0.03—3.31) and \( \chi^2 \) of 0.0 (\( p = 0.97 \)). The only significant differences found in the overall meta-analysis were for CC and CPB times which were shorter in the ST group: for CC times, WMD = 26.57, \( p < 0.00001 \) (CI: 18.91—34.23), \( \chi^2 \) of 182.57 (\( p < 0.00001 \)); for CPB times, WMD = 23.38, \( p < 0.00001 \) (CI: 20.42—36.33), \( \chi^2 \) of 96.61 (\( p < 0.00001 \)).
<table>
<thead>
<tr>
<th>Ref.</th>
<th>First author</th>
<th>Geographic location</th>
<th>Period of study</th>
<th>Follow-up (months)</th>
<th>Implantation technique</th>
<th>Excl. criteria</th>
<th>Total</th>
<th>Mean age (years)</th>
<th>% Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>[22]</td>
<td>Maselli et al. (1999)</td>
<td>Novara, Italy</td>
<td>1/1996—12/1996</td>
<td>8</td>
<td>n/s</td>
<td>L</td>
<td>40</td>
<td>74.5 ± 5.1</td>
<td>43.3</td>
</tr>
<tr>
<td>[29]</td>
<td>Totaro et al. (2005)</td>
<td>Swansea, UK</td>
<td>n/s</td>
<td>18.5 ± 12 SL; 14.5 ± 10 ST</td>
<td>SL: inclusion cylinder; ST: n/s</td>
<td>K, G, I, B</td>
<td>62</td>
<td>55.0 ± 8.0</td>
<td>73.3</td>
</tr>
<tr>
<td>[26]</td>
<td>Santini et al. (1998)</td>
<td>Verona, Italy</td>
<td>1/1993—12/1996</td>
<td>18.5 ± 3</td>
<td>SL: inclusion cylinder; ST: n/s</td>
<td>77</td>
<td>76.0 ± 2.0</td>
<td>32.4</td>
<td></td>
</tr>
</tbody>
</table>

Exclusion criteria: A: emergency/unstable; B: +other cardiac/aortic procedure; C: coronary artery disease/previous MI; D: poor LV; E: calcified aorta; F: redo; G: acute endocarditis; H: TIA/CVA; I: renal failure/dialysis; J: associated MV or TV disease; K: aortic root dilatation; L: <20 and/or >80 years of age; M: drug abuse/excess alcohol; N: aortic tissue disorder; O: active malignancy; and P: predominant aortic regurgitation.

n/s: not specified.

*/ ***This study included three stentless groups (10 patients in each) with mean ages within groups of: 70.4 ± 4.3; 74.5 ± 6.8; and 71.1 ± 3.1 years.
3.4. Sensitivity analysis

We reanalyzed the data for 30-day mortality as the POI using only the highest quality RCTs as defined by the RoB score and the CQS. Use of either the RoB score alone or the CQS resulted in lower heterogeneity with $\chi^2 = 1.90$ and 0.32, respectively, and the 30-day mortality, while non-significant, perhaps exhibited a slight tendency to be lower in the ST group (Fig. 3). In performing a sensitivity analysis including only studies with AS as the predominant valve pathology, or in which the mean age of patients in both SL and ST groups was $\geq 70$ years, there was again a slight tendency to favor the ST group in terms of 30-day mortality, though this was non-significant (Table 3).

4. Discussion

This study was designed to answer whether the more complex technique of SL-AVR increases perioperative risk compared with stented AVR, and how this relates to theoretical benefits for SL valves of improved valve performance, defined by postoperative hemodynamics, valve durability and functional cardiovascular status. We performed an overall meta-analysis for 30-day mortality as the primary outcome and performed a sensitivity analysis for sources of possible heterogeneity by including only the highest quality RCTs as defined by their RoBs and CQSs. We included 18 RCTs, with a total of 1410 patients. There was no difference in the 30-day mortality for SL- versus ST-AVR in accordance with previous reports [5,6]. More importantly, this lack of significant difference in the POI persisted when only the highest quality RCTs were included. Further analysis of 30-day mortality in subgroups consisting of studies, which included predominantly patients with AS or the elderly, did not suggest any significant difference in outcome between the SL and ST groups. Overall valve-related mortality from 1 month to 1 year also did not reveal a significant difference between SL- and ST-AVRs. In terms of morbidity, we examined a number of other outcomes and found no difference between groups for postoperative incidence of cardiac-related morbidity (MI, incidence of AF, need for PPM or IABP), renal failure, and neurological or pulmonary complications. Concerning potential resource use, there were no significant differences for ICU or LOS. The only consistent significant differences between the study groups were found for the aortic CC and CPB times.

4.1. Valve performance: valve durability and hemodynamics

It was thought that the stent apparatus of conventional bioprosthetic valves had two important but related consequences: first, turbulent flow results in less than optimal hemodynamics, mass regression and, perhaps, clinical outcomes; second, that stresses are greater at the strut–leaflet junction, resulting in greater mechanical stresses and reduced durability of stented valves. Concerning durability, we found no significant difference in valve failure rates at 1 year for SL- versus ST-AVR, though only four studies clearly reported this data with a total of 309 patients [17,18,24,25]. The technique of aortic valve implantation might also influence valve durability and performance, though many of the randomized studies included in our analysis used similar methods for SL and ST AVR: subcoronary and supra-annular implantation, respectively, and this could be a factor contributing to the lack of difference in valve performance that we observed.

With respect to LV mass regression, initial studies appeared to suggest that this was more rapid and complete in patients following SL-AVR compared with ST-AVR [22]. A small number of the included RCTs here reported detailed hemodynamic data at serial time-points in terms of change in LVEF or LV mass index (LVMI) [13,14,18,19,23,25,27,28]. A qualitative examination of these data did not show any clear differences between SL and ST valves across the time-points included (Fig. 4) and it may be that the lack of differences in hemodynamics, contrary to what one might expect in

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Randomized</th>
<th>Rand protocol</th>
<th>Withdrawals/dropouts</th>
<th>Jadad score</th>
<th>Matched for Parsonnet/EuroSCORE</th>
<th>Matched for NYHA a/o LVEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>[28]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[14]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[18]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[30]</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>n/s</td>
</tr>
<tr>
<td>[27]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>[20]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>n/s</td>
</tr>
<tr>
<td>[21]</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>n/s</td>
</tr>
<tr>
<td>[15]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[19]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>[17]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[22]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[13]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>[25]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[23]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[16]</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[29]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[26]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>[24]</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
theoretical terms, underlies, at least in part, the lack of difference in terms of clinical outcomes, even when analyzing only the highest quality studies. Further, when we performed a subgroup analysis including only RCTs with predominant AS patient groups, there was still no significant difference between groups in 30-day mortality, despite the lower overall heterogeneity. There was insufficient data from the RCTs available to perform a meta-analysis for 30-day mortality in patients, who had undergone AVR for predominant aortic regurgitation. In a non-randomized study, however, Collinson et al. examined the effect of SL- versus ST-AVR on LV function in patients with AR and LV impairment [33]. SL-AVR was found to result in significant benefits in terms of rapidity and completeness of LV mass regression and LV function, though there were no perioperative deaths in either group.

While several studies included here reported hemodynamic data, only two reported using magnetic resonance imaging. Theoretical terms, underlies, at least in part, the lack of difference in terms of clinical outcomes, even when analyzing only the highest quality studies. Further, when we performed a subgroup analysis including only RCTs with predominant AS patient groups, there was still no significant difference between groups in 30-day mortality, despite the lower overall heterogeneity. There was insufficient data from the RCTs available to perform a meta-analysis for 30-day mortality in patients, who had undergone AVR for predominant aortic regurgitation. In a non-randomized study, however, Collinson et al. examined the effect of SL- versus ST-AVR on LV function in patients with AR and LV impairment [33]. SL-AVR was found to result in significant benefits in terms of rapidity and completeness of LV mass regression and LV function, though there were no perioperative deaths in either group.

While several studies included here reported hemodynamic data, only two reported using magnetic resonance imaging.
imaging (MRI) in their assessment. MRI has disadvantages of cost and time, though is an alternative and perhaps more accurate method for performing LVM determinations and calculations of LVEF. In particular, MRI is a three-dimensional (3-D) technique, whereas echocardiography as used in the studies included in our analysis was a two-dimensional (2-D) study. Sensky et al. did not find a significant difference in LVM regression between groups at 6 months [27]. This is in contrast to the study by Williams et al., who did find a difference favoring SL valves at 3 months [28].

4.2. Valve performance: functional status and LV function

In addition to proposed benefits in terms of LVM regression, a secondary effect of SL-AVR was thought to be improved LV function and in clinical indices of functional cardiovascular status. In the present study, there were insufficient data for meta-analysis relating to functional cardiovascular status in terms of NYHA and Canadian Cardiovascular Society (CCS) class. Further, only the study by Cohen et al. examined functional status in patients using the Duke Activity Status Index score and found no difference between SL- and ST-AVR groups at 12 months following surgery [16].

As a consequence of more rapid LV mass regression, one might expect patients with initially poor LV function to benefit more following SL-AVR compared with ST-AVR. The recent review by Gulbins and Reichenspurner suggested that SL valves might offer an advantage for patients with poor LV function or those with a small aortic annulus [34]. In our study, 14/18 studies matched patients in the two arms for NYHA status and/or LVEF (Table 2). Only that of Totaro et al. explicitly excluded patients with poor LV function [29]. Most studies reporting pre- and postoperative LVEF out to 6 months did not show evident differences from baseline function for the SL- and ST-AVR groups. Studies of Szafranek and Sensky suggest that LVEF improvement in the SL groups were perhaps greater than seen in the ST group, though in the Szafranek study, this could not apparently be accounted for by the change in LVM [27,28]. Related to possible operative differences between studies that might influence LV function, only 12 of the studies commented upon intraoperative strategies for myocardial protection, with either blood or crystalloid cardioplegia and varying degrees of hypothermia being employed [13,14,16–18,21,22,26–30]. These differences may account for some of the observed inter-study heterogeneity. The majority of studies reporting SL valve implantation used a subcoronary technique.

An indirect reflection of changes in LV function and hemodynamic indices may be seen through variations in plasma levels of natriuretic peptides. These are useful supplementary data aiding in the overall assessment of patient outcomes, though not reliable indicators in themselves. Elevated levels of brain natriuretic peptide (BNP) are well known to correlate with heart failure [35]. Only one
though they did find a significant difference in favor of SL-AVR
Nooten et al. found no significant difference in 30-day
and to a higher rate of IABP insertion. In another study, Van
full-root implantation technique for SL valves in the elderly
group [39]. This was partly attributed to the higher use of the
ST-AVR and found a higher perioperative mortality in the SL
that of Westaby et al., who analyzed multicenter data
years in each group, and subgroup analysis of these did not
present meta-analysis suggested a trend towards higher
of MI, low cardiac output and IABP insertion in the SL
group. It is not clear as to why this might be, particularly in
view of the suggested findings of more rapid reduction in LVMI
in patients with SL aortic valves as alluded to above. One of
the included RCTs examined changes in coronary flow in their
patients [14]. It has been suggested that coronary flow
reserve (CFR) may be relevant to long-term survival in
patients with aortic stenosis following AVR [36]. The
investigators used MRI to measure coronary flow as well as
CFR under adenosine stress and found that only the SL-AVR
group exhibited normalization of CFS following surgery. This,
however, could not be related to overt mortality as there
were no deaths or serious adverse clinical events in either the
SL or ST groups at 6-month follow-up. In a non-randomized
comparative study, Collinson et al. also suggest reduced
subendocardial ischemia in a SL-AVR group compared with ST-
AVR, based on electrocardiographic criteria of reduction in
QRS and QT intervals and reversal of LV strain pattern [37].
We found no difference in PPM insertion rates between SL-
or ST-AVRs, though a non-randomized comparative study has
reported a significantly higher incidence of PPM insertion in
patients following SL-AVR compared with ST-AVR [38]. In their
analysis, multivariable predictors of PPM insertion were:
preoperative MI; poor LVEF; left atrial enlargement; left
bundle branch block; CPB time > 100 min; CC time > 70 min;
and prosthetic valve size ≤ 21 mm. The incidence of PVE was
not significantly different in our meta-analysis, though it
showed a tendency to be lower in the ST-AVR group (Table 3).
4.4. Other outcomes: elderly patients and longer-term
survival
Eight of the RCTs included patients with a mean age ≥ 70
years in each group, and subgroup analysis of these did not
show a significant difference in 30-day mortality between
groups. One of the largest non-randomized studies to
examine the influence of SL-AVR on perioperative risk is
that of Westaby et al., who analyzed multicenter data
comparing 583 patients with SL-AVR with 1260 patients with
ST-AVR and found a higher perioperative mortality in the SL
group [39]. This was partly attributed to the higher use of the
full-root implantation technique for SL valves in the elderly
and to a higher rate of IABP insertion. In another study, Van
Nooten et al. found no significant difference in 30-day
mortality among elderly patients with SL- versus ST-AVR,
though they did find a significant difference in favor of SL-AVR
at mid-term follow-up at approximately 5 years [40]. The
same center have published a more recent report, however,
with a slightly larger number of patients and found that this
survival advantage of SL aortic valves is apparently lost at 8
years of follow-up [41].
In terms of survival beyond 30 days, we found no
difference in valve-related mortality from 1 month to 1
year between groups. For mid- and late-term outcomes, very
few RCTs report these data. That of Williams et al. report
only hemodynamic data to 36 months, while
Lehmann et al. report the longest follow-up with a mean
follow-up duration of 94.2 ± 9.6 months [21,30]; Santini et
al. report Kaplan–Meier survival curves to 36 months [26].
More recently, two RCTs have reported mid- and long-term
follow-up data: Cohen et al. report that freedom from
valve-related morbidity at 12 years was significantly lower
in the ST-AVR group [31]. There was a trend for survival to
be better at 12 years in the SL group, though this was not
found to be statistically different. Similarly, Risteski et al.
recently reported their 5-year follow-up data for an RCT
and again found no difference in survival between SL- and
ST-AVR [32].
4.5. Appraisal of studies and RoB analysis
Recent reviews have attempted to examine whether SL
aortic valves have benefits compared with stented ones
[6,34,39]. Cheng et al. surmise that although there may be
early hemodynamic benefits in patients with SL-AVRs, there
do not appear to be benefits in terms of morbidity and
mortality, particularly in the short term [6]. A major
limitation of this study, however, and that of Kunadian et
al. [5] is the lack of critical appraisal of the studies included
in the meta-analyses; in particular, no assessment was
performed for either study quality or for RoB. This limits the
inferences that may be drawn from these studies as it is now
accepted that the standard of rigor required of such analyses
included an assessment of RoB [7].
In the present evaluation, there were marked differences
between studies in terms of the stated exclusion criteria. As
an additional, more critical analysis of the quality of studies
included in our study, we performed a RoB assessment
according to recent, new guidelines reported by the
Cochrane Collaboration [7]. Further, few studies were of
high quality as determined by Jadad or RoB scores or a
composite score devised by adding the Jadad and RoB-
derived scores (CQS).
We found that <25% of studies reported blinding
procedures. While almost 75% of the studies did address
incomplete outcome data and appeared free of selective
reporting, <50% of the RCTs were free of other sources of bias
or adequately reported sequence generation or allocation
concealment. Reanalyzing the primary outcome of 30-day
mortality using only studies with low RoB, we found the
treatment effects of the included studies were closer
further that the overall effect had greater
CIs as previously reported in the study by Hartling et al.,
validating the RoB tool [42]. For studies with a RoB score ≥ 3
(n = 9), the OR = 1.22 (788 patients; CI 0.53–2.80; p = 0.65).
Analyzing studies with a CQS ≥ 5, 670 patients were included
in eight studies with OR = 1.41 (CI 0.59–3.36; p = 0.44). In
terms of sensitivity analysis, the $\chi^2$ value confirmed lower heterogeneity when including only higher quality studies. The evaluation strategy discussed here was to report the POI across all studies and then reanalyze the data using only studies with a low RoB for the same primary outcome as is recommended [7]. We therefore defined a threshold based on the six bias domains with a score of $\geq 5$. This sensitivity analysis allows one to examine sources of heterogeneity between studies. A validation of the RoB tool has been performed by Hartling et al., who have confirmed that studies with high or unclear RoB tend to have higher treatment effects and that the RoB is more accurate than Jadad scores for analyzing RoB [42]. Further, Gurusamy et al. have described the use of RoB analysis as applied to RCTs in surgery [11]. In particular, the lack of blinding in many surgical studies results in bias, though this can be minimized by using outcomes not easily influenced by the patient, investigators or care team, such as all-cause mortality. In the present context, other sources of bias relate to baseline imbalance in terms of selection criteria and an imbalance in known prognostic factors for a treatment, which may thus raise questions about the effect estimate for a treatment [11]. Finally, other sources of bias in surgical trials of devices such as prosthetic valves may refer to sources of funding as well as so-called differential expertise bias relating to experience of individual surgeons [43].

4.6. Study limitations

All studies included in the present analysis were RCTs and as such represent a suitable design to eliminate for confounding factors in terms of intra-study variability, despite the relatively small numbers of patients included within individual reports and the small overall number of published RCTs. One limitation of meta-analysis of the cohorts of studies included here, however, is the marked inter-study variability in terms of exclusion criteria and this represents a major source of study heterogeneity. There were also some key differences in various clinical characteristics between studies such as proportion of patients with mean age $\geq 70$ years, proportion of those undergoing emergent procedures, redo-procedures and those with poor LV function as well as between the valve types compared within and across studies. Preoperative risk factors such as diabetes, smoking, chronic obstructive pulmonary disease, MI, peripheral vascular disease and renal failure were stated by only a small number of the included RCTs [16, 17, 21, 28]. None of the included RCTs provided detailed data relating directly to cost or resource use, although we found no significant differences in ICU or total LOS between groups in the overall meta-analysis.

There may also be an element of publication bias inherent in this type of study as positive findings tend to favor publication. A further potential source of bias might occur within an individual study where more than one surgeon was involved in the trial and SL-AVR was performed by the more experienced surgeons where randomization occurred prior to surgery. This type of bias, however, may only be presumed as no report explicitly stated this. Related to this surgeon-specific bias is longitudinal bias over time within a RCT attributable to the learning curve of less-experienced operators implanting SL-AVRs. Again, this may only be hypothesized as these data were not explicit.

As the number of RCTs of highest quality was even fewer in number, this is a particular limitation of the sensitivity analyses performed and means that one must make inferences with appropriate reservation. This also applies to the small subgroup analyses for the elderly and for RCTs reporting patients with AS as the predominant indication for AVR. Finally, the lack of mid- to long-term follow-up data for RCTs of SL- versus ST-AVR means that definitive conclusions regarding the clinical outcomes of patients are somewhat limited at the present time as reported here.

4.7. Implications of this work and conclusions

The present overall meta-analysis did not find any significant differences in terms of clinical outcomes or valve performance with SL-AVR. More importantly, these findings were robust upon performing sensitivity analysis, excluding studies with a high/unclear risk of bias. Overall meta-analysis for secondary outcomes including cardiac-related morbidity, pulmonary and renal complications similarly did not show any significant differences between groups, despite longer CPB and CC times as found here and reported by others for the SL group [5, 6].

It had been hoped that better hemodynamics of SL valves would translate into better clinical outcomes for patients and, ultimately, this must be a major criterion for success. While we have not been able to find a statistically significant difference in perioperative risk between SL- and ST-AVRs as defined by 30-day mortality or other measures of patient morbidity, no clear benefit has been demonstrated in this analysis for SL aortic valves. Further, the CIs of the data presented here mean that clinically relevant differences between the two valve types cannot be excluded. It should be noted, however, that SL aortic valves have been important in highlighting the resolution of LV hypertrophy following AVR, which few authors had previously considered. SL valves have also spawned two recent developments: improvement in the profile and effective orifice area of stented valves [44]; and facilitation of transcatheter aortic valve implantation [45]. While adequately powered randomized trials comparing SL versus ST AVRs may be difficult to achieve, given the rates of structural valve degeneration for tissue prostheses, larger RCTs with more stringent inclusion and exclusion criteria with a low risk of bias and longer clinical and hemodynamic follow-up data are needed to address the sources of heterogeneity identified in this work and therefore improve the probability of demonstrating a true difference between SL- and ST-AVR if it exists.

References


