Outcome of Mitral Valve Repair or Replacement: A Comparison by Propensity Score Analysis

Robert R. Moss, MD; Karin H. Humphries, PhD; Min Gao, PhD; Christopher R. Thompson, MD; James G. Abel, MD; Guy Fradet, MD; Brad I. Munt, MD*

Background—There are no randomized trials comparing outcomes after mitral valve (MV) repair and replacement. Propensity scoring is a powerful tool that has the potential to reduce selection bias in nonrandomized studies.

Methods—From the BC Cardiac Registries, 2,060 patients presented for MV surgery, with or without CABG between 1991 and 2000. We then identified 322 MV repairs who were then matched by propensity score to an equal number of MV replacement patients. We compared survival and freedom from re-operation outcomes using Cox proportional hazards model analysis. Multivariable analysis was then used to compare outcomes in 358 MV repair patients with 352 MV replacement patients who had undergone chordal sparing surgery.

Results—The comparison groups generated using propensity scores were well balanced with respect to all collected baseline risk factors. Median follow-up time was 3.4 years. Patients undergoing MV repair had significantly improved survival (RR 0.46; 95% CI, 0.28 to 0.75) but a trend toward more re-operations (RR 2.11; 95% CI, 1.00 to 4.47) compared with patients undergoing replacement. Mitral valve repair patients still had better survival (RR 0.52; 95% CI, 0.32 to 0.85) compared with MV replacement patients who had undergone chordal sparing surgery.

Conclusion—We used propensity score methods to reduce selection bias in a population-based cohort of patients undergoing MV repair/replacement. Repair was associated with better survival, but a trend to increased re-operation.

(Circulation. 2003;108[suppl II]:II-90-II-97.)

Key Words: mitral valve ■ prosthesis ■ valvuloplasty ■ surgery

The last decade has seen a significant shift in the surgical management of mitral regurgitation. Mitral valve (MV) repair is now a practical alternative to mitral valve replacement in an increasing number of patients with myxomatous and other etiologies of mitral valve disease. Technical advances in operative procedure1–3 mean that valves with advanced myxomatous disease previously considered not amenable to repair are now frequently repaired.4,5 Increasingly repair of the mitral valve is being undertaken in patients with severe impairment of left ventricular systolic function and in those where ischemia is thought to be the cause of the mitral regurgitation.6 Nonetheless, it is not clear that mitral repair is superior to replacement in all clinical settings.7 There are no prospective randomized trials comparing outcomes after mitral valve (MV) repair versus replacement, and it seems unlikely that such a trial will ever be undertaken.

Proponents of an aggressive repair strategy point out that mitral valve repair better preserves mitral valve architecture, which has been clearly shown in animal models to be an important determinant of postoperative left ventricular function.8,9 Flow characteristics of a repaired valve are likely to be superior to those of a prosthesis, a factor that could affect the functional result and influence long-term determinants of outcome such as atrial and ventricular remodeling. Finally, mitral repair may offer lower embolic risk10,11 and may free the patient from the burden of long-term anticoagulation with the attendant hemorrhagic risk that is associated with the use of mechanical mitral valve prostheses.12,13 On the other hand, mitral repair is technically more complex,14 a factor that is likely to increase time on cardiopulmonary bypass, especially in the case unsatisfactory early repair. Late failure of the repair often results in reoperation that carries an associated morbidity and mortality.14 The mechanism of ischemic mitral regurgitation is incompletely understood so it is not surprising that a variety of techniques for repair have been proposed, and that failure of the repair in the ischemic setting is more frequent than with myxomatous mitral disease.6 It is not clear that repair is always superior in this group.7

Using traditional multivariate adjustment methods, several nonrandomized studies suggest that patients undergoing repair have improved survival,10,15 better postoperative left ventricular function,15 and lower postoperative mortality15 compared with patients undergoing replacement, although this experience has not been universal.16 Traditional methods
of covariable adjustment, however, may not adequately adjust for the bias related to choice of therapy. Propensity scoring is a powerful tool for retrospectively matching covariables in nonrandomized studies that has the potential to reduce selection bias. Propensity scoring, therefore, enabled us to compare mitral repair patients to those undergoing mitral replacement with similar baseline characteristics.

The importance of preserving the subvalvular architecture in mitral valve replacement surgery and its salutary effect on left ventricular systolic function in patients is now generally accepted. We also wished to investigate the influence that chordal sparing might have on the comparison in outcome between the repair and replacement groups.

### Patients and Methods

We compared overall survival, survival from 30 days post surgery and freedom from reoperation in 322 patients undergoing MV repair and an equal number of patients undergoing MV replacement selected using propensity score methods to reduce the bias related to choice of surgical procedure.

#### Patients

We identified 2,060 patients presenting for MV surgery, with or without coronary artery bypass surgery between 1991 and 2000 from the BC Cardiac Registries, Vancouver, BC, Canada. This cohort excluded patients who had had previous cardiothoracic surgery. Those undergoing concomitant aortic valve surgery were likewise excluded, whereas those undergoing tricuspid repair surgery were included. After exclusions for missing mitral valve diagnosis (N=21) and missing covariables required for the propensity score (N=58), the final study cohort comprised 1,981 patients: 338 patients undergoing MV repair and 1,643 patients undergoing MV replacement. The derivation of the final study cohort is shown in Figure 1. The study was undertaken with the approval of the local institutional review board.

#### Data Sources

BC Cardiac Registries prospectively capture detailed demographic, clinical, and procedural information on all cardiac procedures performed on adults in British Columbia. Data collection utilizes standardized forms and definitions.

#### Surgery

Sufficient information was included in the registry to allow categorization of the types of repair in 228 of the 322 MV repair patients. Of the patients undergoing a repair, 215 (95%) had a mitral ring replacement, 120 (53%) had a quadratic posterior leaflet resection, and 7 (3%) had a triangular anterior leaflet resection, whereas 29 of the posterior leaflet repairs had a leaflet advancement procedure. Twenty percent (9%) had a suture repair of a congenital mitral cleft, torn leaflet, or mitral valve perforation. A chordal procedure, either chordal shortening, transfer, or the insertion of artificial chordae was performed on adults in British Columbia. Data collection utilizes standardized forms and definitions.

### Statistical Analyses

Logistic regression was used to model the probability of being assigned to MV repair over MV replacement, based on observed baseline characteristics (Tables 1 and 2). A stepwise approach was used with P≤0.25 as the limit for selecting variables for entry into the model. The estimated probability of the final model, called the propensity score, was calculated for each patient. For every MV repair, matching patients with the closest propensity score were identified from the larger pool of MV replacements. The maximum allowable difference for a matching was <0.1 in the propensity score. Specifically, each MV repair was matched to one replacement with the closest propensity score. When two or more replacement patients had the same propensity score match, the match for the analysis was chosen randomly. Matched subjects were removed from the pool and the next MV repair and its matched replacement were selected until no further matches could be identified. For further details on propensity score methodology, see the Appendix.
were able to identify 352 patients from the original MV replacement cohort who had either a Khonsari type I or II chordal MV repair and replacement. The proportion of the total MV replacement cohort would have on comparison of the outcomes for Chordal Sparing was determined from the British Columbia cardiac registries. suggesting that loss to follow-up is minimal. Redo mitral surgery Statistics database resulted in matches for 95.7% to 99.8% of cases, Previous linkages between the BC Cardiac Registries and the Vital Statistics database, using the personal health number, surname, first name, middle health number. Probabilistic linkage with the Vital Statistics data-


All residents of British Columbia, Canada, have a unique personal health number. Probabilistic linkage with the Vital Statistics database, using the personal health number, surname, first name, middle initial, date of birth, and gender, provided mortality information. Previous linkages between the BC Cardiac Registries and the Vital Statistics database resulted in matches for 95.7% to 99.8% of cases, suggesting that loss to follow-up is minimal. Redo mitral surgery was determined from the British Columbia cardiac registries.

**Endpoints**

All residents of British Columbia, Canada, have a unique personal health number. Probabilistic linkage with the Vital Statistics database, using the personal health number, surname, first name, middle initial, date of birth, and gender, provided mortality information. Previous linkages between the BC Cardiac Registries and the Vital Statistics database resulted in matches for 95.7% to 99.8% of cases, suggesting that loss to follow-up is minimal. Redo mitral surgery was determined from the British Columbia cardiac registries.

**Chordal Sparing**

We wished to look at the effect that chordal sparing in the MV replacement cohort would have on comparison of the outcomes for MV repair and replacement. The proportion of the total MV replacement cohort who had either partial or complete chordal sparing, although we believe the true proportion of the total MV replacement cohort to be higher. Unfortunately, this left a reduced cohort of MV replacement patients available to be propensity matched to the MV repair patients. With such a reduced set of MV replacement patients to choose from, propensity matching would produce a significantly reduced and highly selected cohort. The results of an analysis based on a cohort such as this would be unlikely to be able to be applied to a biologically real population. We therefore used multivariable analysis to compare the major outcome measures of survival, freedoms from redo and 30-day post-surgery survival and freedom from redo. The Cox proportional hazards model was used with adjustment for clinically significant covariables. The significant covariables differed slightly depending on the end-point used. The MV replacements with chordal sparing were compared with a group of 358 MV repair patients. The number of patients in this group is higher than that of the repair group used in the propensity analysis due to our ability to use patients with some missing covariables as well as patients who we could not adequately match to a replacement patient.

**Results**

As illustrated in Table 2, significant baseline differences were observed between the two groups before propensity score matching (all). After matching by propensity score (matched), however, the MV repair and replacement cohorts were well balanced with respect to baseline variables (Table 2). Potentially important covariables investigated in propen-
TABLE 3. Hazard Ratios and 95% Confidence Intervals for Covariables Used in Cox Proportional Hazard Risk Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Survival</th>
<th>Redo Free Survival</th>
<th>Freedom From Post 30-Day Redo</th>
<th>Post 30-Day Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Propensity score</td>
<td>1.09</td>
<td>0.14–8.73</td>
<td>0.61</td>
<td>0.09–1.84</td>
</tr>
<tr>
<td>Age (per 10 years)</td>
<td>1.36</td>
<td>0.99–1.85</td>
<td>1.38</td>
<td>1.14–1.66</td>
</tr>
<tr>
<td>Sex (female)</td>
<td>0.03</td>
<td>0.01–0.61</td>
<td>1.66</td>
<td>1.09–2.53</td>
</tr>
<tr>
<td>Age*Sex</td>
<td>1.91</td>
<td>1.21–3.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Etiology</td>
<td>(degenerative + myxomatous)</td>
<td>0.49</td>
<td>0.29–0.85</td>
<td>-</td>
</tr>
<tr>
<td>Etiology</td>
<td>(congenital)</td>
<td>16.03</td>
<td>1.93–132.96</td>
<td>6.98</td>
</tr>
<tr>
<td>Ejection fraction (&lt;35%)</td>
<td>2.86</td>
<td>1.54–5.32</td>
<td>2.56</td>
<td>1.52–4.30</td>
</tr>
<tr>
<td>Ejection fraction (35–50%)</td>
<td>1.86</td>
<td>1.02–3.39</td>
<td>1.98</td>
<td>1.23–3.21</td>
</tr>
<tr>
<td>Emergency procedure†</td>
<td>-</td>
<td>-</td>
<td>2.70</td>
<td>0.88–8.31</td>
</tr>
<tr>
<td>CHF (current or history)</td>
<td>2.79</td>
<td>1.49–5.22</td>
<td>2.05</td>
<td>1.24–3.39</td>
</tr>
<tr>
<td>Active endocarditis</td>
<td>6.67</td>
<td>1.27–34.97</td>
<td>9.10</td>
<td>2.39–34.72</td>
</tr>
<tr>
<td>PVD</td>
<td>2.17</td>
<td>1.08–4.35</td>
<td>1.95</td>
<td>1.02–3.74</td>
</tr>
<tr>
<td>Dialysis or elevated</td>
<td>2.14</td>
<td>1.23–3.73</td>
<td>2.27</td>
<td>1.39–3.70</td>
</tr>
<tr>
<td>Renal failure</td>
<td>15.52</td>
<td>3.38–71.19</td>
<td>4.39</td>
<td>0.87–22.10</td>
</tr>
<tr>
<td>Chronic steroid use</td>
<td>4.14</td>
<td>1.44–11.89</td>
<td>4.16</td>
<td>1.76–9.82</td>
</tr>
<tr>
<td>Jehovah’s Witness</td>
<td>22.32</td>
<td>2.82–177.04</td>
<td>10.33</td>
<td>1.37–78.24</td>
</tr>
</tbody>
</table>

*Relative to rheumatic, endocarditis, ischemic, or other unknown etiology.
†Relative to elective procedure.

sity modeling but not shown in Table 1 included body weight and surface area, remote history of myocardial infarction, and active endocarditis and pre-operative warfarin and aspirin use. The following medical comorbidities, as coded by the surgeon performing the procedure, were used in the propensity score model: chronic renal failure, chronic steroid use, chronic pulmonary disease, dialysis or elevated creatinine, cerebrovascular and peripheral vascular disease, peptic ulcer or gastrointestinal bleeding, liver disease, and malignant disease. The beta coefficients for the covariables used in the propensity score model are provided in Table 4.

Survival

Median follow-up time for this study was 3.4 years. Hazard ratios and 95% confidence intervals of the covariables included in the various risk models are provided in Table 3. Nine patients (2.8%) in the MV repair group and 5 patients (1.6%) in the MV replacement group died within 30 days of surgery (NS, $P=0.28$). Patients undergoing MV repair experienced significantly improved probability of survival with a hazard ratio for death of 0.46 (95% CI 0.28 to 0.75), when compared with MV replacement patients. The proportional hazards assumption was tested but not violated in the models. Kaplan-Meier (KM) survival curves for the MV repair and replacement patients are presented in Figure 2.

The survival advantage for the mitral valve repairs was also seen when the patients who survived surgery were considered. The hazard ratio for death in MV repair patients who were alive 30 days after surgery was 0.34 (95% CI 0.20 to 0.60). Kaplan-Meier curves are presented in Figure 3.

Freedom from Reoperation

There was an increase in the risk of MV reoperation in the repair group, with the hazard ratio of 2.11 (95% CI 1.00 to 4.47). Kaplan-Meier curves for freedom from MV reoperation are presented in Figure 4.

MV Repair-Free Survival

The hazard ratio for death and MV reoperation in the MV repair group was 0.79 (95% CI 0.53 to 1.17). Kaplan-Meier curves for MV repair free survival are presented in Figure 5. The hazard ratio for death and MV reoperation in those alive 30 days after surgery was 0.65 (95% CI 0.42 to 1.00). The hazard ratios for MV repair compared with MV replacement for all endpoints and their associated confidence intervals are illustrated Figure 6.

Chordal Sparing

Significant baseline differences were observed between the MV repairs and the MV replacements with conservation of the subvalvular apparatus (chordal sparing surgery). A survival advantage for MV repair was again noted when the repair patients were compared with the MV replacement patients with chordal sparing surgery after multivariable analysis had been used to correct for baseline differences between the two study groups. The hazard ratios for death in the repair group were 0.52 (95% CI 0.32 to 0.85), and 0.34 (95% CI 0.19 to 0.61) in those patients who survived 30 days after surgery. Once again increased reoperation was noted in the repair group RR=2.93 (95% CI 1.31 to 6.54). The hazard ratio for death and MV reoperation was 0.92 (95% CI 0.63 to 1.36) in the repair patients, whereas the hazard ratio for death
and MV reoperation in those patients who survived 30 days after surgery was 0.65 (95% CI 0.40 to 1.04). The hazard ratios for mitral valve repair compared with the mitral replacement patients who had had chordal sparing surgery are illustrated in Figure 7.

**Discussion**

**Survival Advantage for Repair**

Our study has confirmed previous studies that have shown a survival advantage for MV repair over MV replacement. In the absence of randomized data, our study, using propensity matching, is reassuring and adds weight to the view that the survival advantage is a result of the salutary effect of MV repair rather than to selection biases inherent in the choice of therapy.

The survival advantage of repair over replacement was also seen when a cohort of MV replacement patients who had had conservation of their subvalvular apparatus were compared with MV repairs using multivariate analysis to correct for baseline differences. In other words, the incremental benefit to the repair group cannot be solely related to the effect of
chordal sparing surgery and its benefit of preserving postoperative left ventricular function.

Newer generation mechanical prostheses, with enhanced flow characteristics and reduced thromboembolic potential, may be associated with better outcomes compared with older prostheses. Although exact numbers are not available from the BC Cardiac Registries, practice at our institution for the last decade has been to use Carbomedics or St. Jude mechanical valves, a trend that we believe has been followed at the other cardiac surgery institutions in British Columbia. The survival benefit for repair is seen, however, even when newer-generation prostheses are used. Although most previous comparisons of repair and replacement have been single center, our study incorporates the experience of four cardiac surgery centers, and is therefore likely to reflect real world practice.

Limitations of the Study

Propensity scoring is a powerful tool that enables excellent matching of baseline characteristics, which may be superior to that obtained in a randomized trial. However, if important, unobserved covariables are not identified and not entered into the propensity model, significant baseline differences may still exist between the two groups. Propensity scoring is therefore not a substitute for randomization. We did not, for example, have information on heavy mitral calcification in our database and so could not correct for it in the propensity analysis. We believe, however, that in that particular instance, numbers of patients so affected are likely to be small and unlikely to present a significant bias in terms of exaggerating the mortality in the replacement group. In addition, propensity matching for age and renal function (risk factors for mitral annular calcification) is likely to further reduce the potential for imbalance between the two groups.

Given the richness of our database (in terms of the number of covariates recorded), it is unlikely then that important covariables were not identified, but the homogeneity of the two populations cannot be guaranteed outside the setting of a randomized trial.

It would have been useful to be able to determine with precision the cause of death in the MV repair and replacement patients (propensity-matched cohort).
groups, and in particular the proportion of these deaths that were valve related. Although data on the cause of death is available from vital statistics, such data, based as it is on death certificates, is unreliable and likely to over-represent cardiac deaths. We made a deliberate decision, therefore to focus on 'hard' and clearly verifiable end-points such as death and mitral reoperation. Similarly, specific valve-related complications are not captured in our study, although it might be reasonable to infer that the designation redo mitral surgery might encompass an important proportion of late valve related complications.

The best outcomes with mitral valve replacement probably occur when all chordal structures are optimally preserved.19 Because we were unable to identify the type of chordal sparing surgery carried out in the MV replacement patients, this cohort is likely to be a mixed group consisting of those with both complete and partial (posterior) chordal sparing.

Conclusions
In the absence of randomized controlled trials, propensity score methods were used to reduce the potential selection bias associated with the choice of repair versus replacement in a population-based cohort of patients undergoing MV surgery. MV repair was associated with better survival compared with MV replacement, although the benefit was relatively modest. A trend toward increased reoperation in the MV repair group was noted.

Appendix: Propensity Scoring: Description and Methodology
Although randomized controlled trials (RCT) are considered the gold standard for the assessment of efficacy, there are limitations to this approach. First, strict inclusion and exclusion criteria may limit the applicability of the study results to other, perhaps more typical, patient populations. Also, there may be ethical concerns or issues of feasibility that limit the use of RCTs. Under such circumstances, the non-randomized observational trial may be the most appropriate study design.

A key strength of the randomized study is random treatment assignment, which guarantees that, on average, there are no systematic differences in observed or unobserved covariates between the treatment groups. Balancing of observed and unobserved covariates results in an unbiased estimate of the treatment effect. Nonrandom treatment assignment, characteristic of observational studies, can lead to large differences in the observed and unobserved covariates between the treatment groups. Reasons for this imbalance include careful patient selection by the physician, as well as patient preferences. These differences can lead to biased estimates of treatment effects. Propensity scores are a useful technique to control for the imbalance in covariates that usually occurs in non-random treatment assignment.

Propensity scores do not replace traditional multivariable modeling.17 Rather, a propensity score analysis involves a two-step process—first a model to estimate the probability of treatment assignment and then a traditional multivariable model for estimating the treatment effect. The propensity score \( e(x) \) is defined as the estimated probability of assignment to one treatment (Z) over another, given the observed baseline covariables \( (X_i) \).

\[
e(x) = \Pr(Z = 1 \mid X_i)
\]

The propensity score is a balancing score that is a function of the observed covariables such that the distribution of the observed covariables is the same for the treated and untreated groups. The result is a quasi-randomized study in which two subjects, one treated and one control, with the same propensity score, were randomly assigned to each group, by virtue of their being equally likely to be selected for treatment or placebo. Thus, adjustment for propensity score tends to produce an unbiased estimate of treatment effect when the treatment assignment is strongly ignorable. Strongly ignorable means there are no systematic, unobserved, pretreatment differences between treatment and control that are related to the outcome of interest.

Another important characteristic of the propensity score is its ability to efficiently control for baseline differences between treatment groups using a single scalar value. This property is of particular value when prediction of treatment is complex and model development requires the introduction of large numbers of covariables, multiple interaction terms or higher order terms. Such complex models are difficult to interpret and estimation of main effects can be problematic. But, because the goal of the propensity score model is...
to obtain the best-estimated probability of treatment assignment, over-fitting this model is not a concern. Therefore, adjustment of bias in the background covariables can be made using the single, scalar propensity score, rather than all of the background covariables, individually in the model of treatment effect.

The propensity score can be estimated using logistic regression. The outcome is treatment assignment and the independent variables are the observed covariables plus plausible interactions and higher order terms. Details of the covariables used in the propensity score model are provided in Table 4. Missing covariables were treated as a different level in the propensity model.

Once the score is calculated the most common techniques for its use include: matching, stratification, and regression adjustment. Figure 8 illustrates the process of selecting a propensity-matched cohort from the overall population.

In this analysis of mitral valve replacement versus repair, we employed Greedy matching (SAS protocol) and regression adjustment. Greedy matching creates a study cohort of MV repair patients matched with the nearest available MV replacement patient based on the estimated propensity score. The propensity score ranges between 0 and 1. In our model, 5-digit propensity scores were estimated. Using Greedy matching, 18 patients were matched on all 5 digits, 98 were matched on 4 digits, 306 were matched on 3 digits, 184 were matched on 2 digits, and 38 were matched on 1 digit. The propensity score was then used as a covariate in Cox proportional hazards model to estimate the treatment effect, in addition to clinically important covariables.

Acknowledgments
The authors wish to acknowledge the surgeons and nurses who contributed information to the BC Cardiac Registries. Brad I. Munt received support from the Heart and Stroke Foundation of British Columbia and the Yukon.

References
Outcome of Mitral Valve Repair or Replacement: A Comparison by Propensity Score Analysis
Robert R. Moss, Karin H. Humphries, Min Gao, Christopher R. Thompson, James G. Abel, Guy Fradet and Brad I. Munt

Circulation. 2003;108:II-90-II-97
doi: 10.1161/01.cir.0000089182.44963.bb

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2003 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/108/10_suppl_1/II-90

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/