SUPPLEMENTAL MATERIAL

Supplemental Methods

Within-Trial Cost-effectiveness Analysis

In this within-trial cost-effectiveness analysis (CEA), costs associated with the index hospitalization and readmissions were summed up for 1- and 2-years of follow-up for each of the 251 patients in the severe ischemic mitral regurgitation (MR) trial. Generic quality-of-life weights (i.e., SF-6D utility scores) were linearly interpolated between the assessment visits. In case of an interval death prior to the 1 or 2-year time frame of analysis, we assumed that quality-of-life would drop immediately upon death (reference case analysis). We imputed future SF-6D utility scores assuming missing-at-random (MAR) for subjects who had died to facilitate linear interpolation. All imputations, including those for costs and SF-6D utility scores that were missing because of loss to follow-up, were done by the method of multiple imputation (see 'Imputation of Missing Cost and Quality-of-life Data'). We calculated the area-under-the-curve across longitudinal SF-6D measurements using the trapezoid rule in order to estimate QALYs. Subsequently, we estimated point estimates of 1-year and 2-year cumulative costs and QALYs for each study arm by averaging across imputed datasets. Costs and QALYs incurred during the second year of follow-up were discounted using a discount rate of 3%. Bootstrap uncertainty intervals around the point estimates were based on 1,000 bootstrap replications of the original data, which were generated by randomly sampling subjects with replacement stratified by study arm. Multiple imputation was performed for each bootstrap replicate. Ninety-five % uncertainty intervals (UIs) of (differences in) cumulative costs and QALYs were calculated by the bias-corrected and accelerated method to correct for skewness in the bootstrap distribution.

Development of the Microsimulation Model

For making long-term predictions of costs and QALYs, we developed an individual-level (microsimulation) state-transition model using TreeAge Pro 2017 (TreeAge Software. 2017. TreeAge Pro 2017. Williamstown, MA: TreeAge Software, Inc.) and R version 3.3.2 (2016, R Foundation for Statistical Computing) . The model consisted of an 'Alive' and 'Dead' state, with transitions for readmissions (for heart failure, other CVD or non-CVD reasons) and competing all-cause mortality (**Supplemental Figure I**). The probability of having a reoperation was estimated conditional on the reason for readmission (heart failure, other CVD or non-CVD).

The cycle-length in the model was fixed at 1-month and the following formula was used to calculate 'cause-specific'1-month risks while accounting for competing risks within each cycle, assuming constant hazard rates within each month:

$$P_i = \frac{\lambda_i}{\sum \lambda} \times \left(1 - \mathrm{e}^{-(\Sigma \lambda)}\right)$$

where $\lambda \in \{\lambda_1, \ldots, \lambda_i\}$ is defined as the set of 1-month transition rates (λ) for four causes: all-cause death, heart failure readmissions, other CVD readmissions and non-CVD reasons. P_i is defined as the 1-month risk calculated for cause *i*.

One-month transition rates were individualized using a multivariable Andersen-Gill model with the R coxph function. The Andersen-Gill model is an extension of the Cox proportional hazards model that allows for recurrent events. The following predictor variables were used: a trial variable (severe vs moderate ischemic MR), age, gender, a treatment variable (repair vs no repair), and time-varying covariates for readmissions. Trial, age, gender, treatment, and an interaction term for trial and treatment were included by default. Time-varying covariates for readmissions other than the modelled readmission outcome and interactions by trial were included when $p<0.20$. Time since randomization was the time scale for these models. After assessing the proportional hazards assumption by plotting scaled Schoenfeld residuals as a function of time using the R cox.zph function, we split follow-up time in an early and late phase with a breakpoint at 3 months to take into account the violation of the proportional hazards assumption for modelling competing mortality. This breakpoint allowed for a difference in the relative effect of repair vs replacement on early and late mortality rates.

Baseline hazard functions were estimated using a Turnbull method for left truncated survival data. We used follow-up data from 9 months through 2-years post-randomization to account for hazard rates that leveled off over time and for considering a sufficient amount of failure times. We explored various alternative distributions for extrapolation of the hazard functions. First, we fitted restricted cubic spline functions with four knots and log time as the

independent variable and the log cumulative hazard as the dependent variable.[1] Because restricted cubic spline functions are linear beyond the last knot, a Weibull distribution is assumed for estimated hazards beyond time for which we had trial data (referred to as a 'spline-Weibull' method). Knot locations were specified as the minimum, 33th and 67th percentile, and maximum log time. We also used a classical method for the Weibull distribution by fitting ordinary least squares regression equations on the log cumulative hazard with log time since randomization as dependent variable. Second, we assumed an exponential distribution, by fitting ordinary least squares regression equations on the cumulative hazard with time since randomization as dependent variable. Third, we assumed a log-logistic distribution, by fitting an ordinary least squares regression equation on the $\log\left((1 - survival_{time})/ survival_{time}\right)$ with log time since randomization as dependent variable. Based on assessment of plots of the cumulative hazard and time, for the reference case, we modelled readmission rates beyond the 24 months with spline-Weibull equations and competing mortality rates with an exponential distribution. Baseline rates across the first 24 months were modelled with the non-parametric 'Turnbull' hazard functions (**Supplemental Figure II**).

Finally, hazard ratios, baseline hazard rates, and probabilities of reoperation were re-estimated in 1,000 bootstrap replicates to account for parameter uncertainty. Quality-of-life weights of simulated subjects in the 'Alive' state were based on SF-6D utility scores, and a value of zero was assigned for the 'Dead' state. In the microsimulations, we used SF-6D utility scores from the baseline, 6-,12, and 24-months with interpolation to determine each individual's simulated score for a given model cycle. These SF-6D utility scores were based on the same bootstrap datasets used for modelling transition rates. For model cycles beyond 24-months we assumed that SF-6D utility scores would continue unchanged from the 24-month value. However, in case of re-operation we set the time clock back and used the individual's baseline SF-6D utility score again as start value. In addition, for each readmission occurring after month 24, we applied a one-off disutility toll when no reoperation occurred. This readmission disutility penalty (i.e., a decrement from the cumulative QALYs) was appropriately adjusted to fit the predicted length-of-stay measured in years.

Predictions of length-of-stay were done using a negative binomial model where the admission was the unit of analysis using data on readmissions excluding reoperations. We included age at admission, gender, randomization assignment, and reason for readmission (heart failure, other CVD or non-CVD) as covariates and fitted the prediction model using the same bootstrap datasets that were used for the models for transition rates. Hospitalization costs were modelled conditional on occurrence of readmission. To model the readmission costs, we developed a prediction model using the admission as unit of analysis and a gamma distribution and log-link function using the data on the non-operative readmissions and predictors as described above. Predictions of costs associated with simulated readmissions in which a reoperation took place were based on the mean values observed for the index hospitalization where mitral valve replacement took place and treatment cross-over did not occur.

Validity of the reference case microsimulation model was deemed to be satisfactory by comparing predicted to observed crude counts of readmissions and reoperations over 2-years of follow-up for the 251 severe ischemic MR trial participants (**Supplemental Figure III**). Finally, the microsimulation model showed good performance when comparing predictions of cumulative costs and cumulative QALYs to those observed in the within-trial CEA (**Supplemental Figure IV**).

Imputation of Missing Cost and Quality-of-life Data

Missing cost and quality of life data were imputed using multiple imputation. Costs were first inflated to 2015 U.S. dollars using the Personal Health Care (PHC) index for hospital care[2] before inclusion in the imputation models. Missing costs regarding the index hospitalization and longitudinal SF-6D utility scores (baseline, 30-day, 6-months, 12-months, 24-months) were imputed using the patient as the unit of the analysis and a multivariable flexible additive model (R function 'aregimpute' within the 'rms' package) using predictive mean matching and 50 imputations. In addition to index costs and longitudinal SF-6D utility scores, the imputation model included study arm, age, gender, race, diabetes at baseline, heart failure at baseline, renal disease at baseline, atrial fibrillation at baseline, concomitant CABG performed (yes/no), cross-over during the index hospitalization, disposition at discharge, number of adverse events during the index hospitalization, post-surgical time to discharge, length of stay, ICU duration, cardiopulmonary bypass time, OR duration, longitudinal ESVI (baseline, 6, 12, and 24 months), longitudinal mitral regurgitation grade (6, 12, and 24 months), time of quality-of-life measurement (30-day, 6, 12, and 24 months), an indicator variable for MACE and the Nelson-Aalen estimate of the cumulative hazard of MACE.[3] Missing readmission costs were imputed with hospital admission as the unit of the analysis with 100

imputations. In addition to readmission costs, this imputation model included age at admission, gender, race, primary reason of the readmission (heart failure, other CVD, non-CVD), ICU duration, length-of-stay, and disposition at discharge. Readmissions for which a reoperation occurred were not included in the readmission imputation model.

Long-Term Cost-Effectiveness Analysis

In the reference case, we first assigned the 2-year cumulative costs and QALYs, estimated in the within-trial CEA, as initial values for each subject. This was done for the 1,000 bootstrap replicates. We then used the microsimulation model to add cumulative costs and QALYs beyond 2 years while using a 3% discount rate and running the simulations over a 5 and 10 year time horizon. One-off quality-of-life penalties were additionally applied for readmissions occurring after two years as described above. For the reference case analysis we simulated N individuals x 200 random walks and for the probabilistic sensitivity analysis we used 1,000 bootstrap replicates x N individuals x 10 random walks for each treatment strategy. N was set to 126 individuals for repair and 125 for replacement. We calculated 95% UIs for cumulative costs and QALYs using the bias-corrected and accelerated method.

Uncertainty Analyses

First, for the within-trial CEA we assumed that patients who died would have a gradual decline in quality-of-life from the last value measured until death. In this case, we interpolated the SF-6D utility score using the score of the last exam visit before death and a value of zero at the time of death. Second, we repeated the long-term CEA using two different annual discount rates, 0 and 5%. Third, we used different distributions for extrapolating hazard rates. For modelling readmission rates, we used Weibull and log-logistic models. For modelling mortality we used spline-Weibull, Weibull and log-logistic distributions (**Supplemental Figure V**). Exponential (for readmissions) and Gompertz models were inappropriate because of large deviations from the non-parametric 2-year cumulative hazard function.

Subsequently, we performed a number of sensitivity analyses for long-term cost-effectiveness. First, we conducted analyses fixing the costs of the index hospitalization in the replacement arm and the repair arm at their lower and upper 95% UI limits. In addition, we ran simulations fixing hazard ratios of treatment assignment for heart failure and other cardiovascular readmission rates at their lower and upper 95% UI limits and fixing readmission costs at their lower and upper 95% UI limits. Finally, we assessed the effect of baseline age on cost-effectiveness outcomes by redefining the baseline age of each subject at a fixed age within a range from 50 to 85 with 5-year increments. We used an updated microsimulation model in these latter analyses, which also included prediction models for SF-6D utility scores and costs associated with the index hospitalization. For prediction of SF-6D we used generalized estimating equations (GEEs) with an independent correlation structure and SF-6D score at baseline, 30-day, 6-months, 12-months, and 24-months as dependent variable and baseline age, gender, and study arm as covariates. For index hospitalization costs we used a generalized linear model with a gamma distribution and a log-link function and baseline age, gender, and study arm as covariates. Again, these prediction models were re-fitted within the 1,000 bootstrap replicates. We did not change the age effect in the Anderson-Gill models, and thus did not include any interactions with baseline age, assuming no heterogeneity in effect of repair vs. replacement on readmissions and competing mortality. Within these fully model-based one-way sensitivity analyses on baseline age, we kept the values of all other model parameters at their original value.

Supplemental Table I. Hazard ratios (95% CI) obtained from Andersen-Gill models for heart failure, other cardiovascular and non-cardiovascular readmissions, and death

Predictor	Heart failure readmission	Other cardiovascular readmission	Non-cardiovascular readmission	Death
Age	0.998 (0.978 to 1.023)*	0.973 (0.953 to 0.997)	1.011 (0.996 to 1.027)	1.045 (1.018 to 1.077)
Male	0.854 (0.573 to 1.269)	1.040 (0.659 to 1.656)	0.725 (0.546 to 0.972)	0.688 (0.433 to 1.112)
Randomization to repair vs replacement	1.505 (0.899 to 2.610)	1.580 (0.790 to 3.227)	0.831 (0.569 to 1.220)	-
< 3 months post-randomization	-	-	-	0.438 (0.155 to 0.994)
≥ 3 months post-randomization	-	-	-	1.062 (0.494 to 2.087)
Previous readmission				3.928 (2.517 to 6.972)
Heart failure	4.266 (2.638 to 6.411)	3.085 (1.851 to 4.800)	1.764 (1.140 to 2.607)	-
Other CVD	-	3.019 (1.576 to 4.631)	-	-
Non CVD	2.511 (1.567 to 4.024)	1.455 (0.750 to 2.265)	3.255 (2.316 to 4.384)	-

95% CIs are based on the 2.5^{th} and 97.5^{th} percentiles obtained from the 1,000 bootstrap replicates.

*Accounts for interaction between age and SMR trial.

Supplemental Table II. Anticoagulation and antiplatelet medication use over 2-year follow-up

Variable	Repair N=126	Replacement N=125
Medication use among those alive at 12-month visit, N (%)		
Oral anticoagulation	41 (43)	46 (46)
Aspirin	73 (77)	85 (84)
Clopidogrel	16 (17)	5 (5)
Missing	1 (1)	0
Medication use among those alive at 24-month visit, N (%)		
Oral anticoagulation	33 (41)	44 (51)
Aspirin	67 (84)	64 (74)
Clopidogrel	10 (13)	10 (12)
Missing	0	0

Supplemental Table III. Study population baseline and operative characteristics

Variable	Repair N=126	Replacement N=125
Age (y), mean ± SD	68.9 ± 10.2	67.9 ± 9.0
Male, n (%)	77 (61)	78 (62)
Diabetes, n (%)	48 (38)	41 (33)
Medical history, n (%)		
Myocardial infarction	99 (79)	88 (70)
Prior PCI/CABG	59 (47)	55 (44)
Heart Failure	89 (71)	91 (73)
Atrial fibrillation	45 (36)	35 (28)
ICD	23 (18)	17 (14)
Stroke	14 (11)	11 (9)
Renal insufficiency	29 (23)	40 (32)
LVESVI (mL/m^2) , mean ± SD	61.1 ± 26.2	65.7 ± 27.3
Generic quality-of-life, mean ± SD		
SF-12 physical component summary	37.3 ± 8.1	37.2 ± 7.2
SF-12 mental component summary	47.9 ± 7.7	47.8 ± 9.1
SF-6D health index score	0.65 ± 0.12	0.68 ± 0.13
Concomitant procedures, n (%)		
CABG	93 (74)	94 (75)
Tricuspid valve repair	16 (13)	22 (18)
Atrial maze	15 (12)	16 (13)

Abbreviations: CABG, coronary artery bypass grafting; SF-6D, Short-Form Six-Dimension; SF-12, 12-Item Short Form Health Survey; ICD, implantable cardioverter defibrillator; LVESVI, left ventricular end-systolic volume index; PCI, percutaneous coronary intervention.

Supplemental Table IV. Resource use and costs associated with the index hospitalization

Variable	Repair N=126	Replacement N=125
Cardiopulmonary bypass time, min	139 ± 53	151 ± 50
Time to discharge, days	11.5 ± 9.0	11.9 ± 8.6
ICU stay, days (post-randomization)	5.7 ± 8.0	6.5 ± 8.3
Discharge disposition, n (%)		
Home	72 (57)	76 (61)
Skilled nursing/inpatient rehabilitation facility	43 (34)	36 (29)
Death	5 (4)	8 (6)
Other	6 (5)	5 (4)
Costs, $*		
Mean ± SD	75,505 ± 59,086	81,195 ± 64,703
Median (IQR)	57,829 (42,376 to 84,014)	65,595 (51,805 to 86,060)

*Based on complete cases (n=87 for repair and n=85 for replacement).

Abbreviations: ICU, intensive care unit; IQR, interquartile range.

Supplemental Table V. Resource use and costs associated with readmissions over 2-year follow-up

Readmission	Repair	Replacement
Heart failure		
Count	46	29
Mean Length-of-stay ± SD, days	7.6 ± 10.8	6.5 ± 5.7
Mean ICU stay ± SD, days	2.7 ± 7.6	0.5 ± 1.3
Mean cost ± SD, $	19,412 ± 31,677	12,137 ± 13,541
MI/Angina		
Count	18	3
Mean Length-of-stay ± SD, days	3.1 ± 1.9	3.7 ± 2.9
Mean ICU stay ± SD, days	0.2 ± 0.5	0.7 ± 1.2
Mean cost ± SD, $	8,654 ± 7,926	22,906 ± 25,824
Arrhythmia		
Count	10	13
Mean Length-of-stay ± SD, days	4.3 ± 3.3	5.3 ± 3.4
Mean ICU stay ± SD, days	1.3 ± 2.2	1.6 ± 4.4
Mean cost ± SD, $	21,664 ± 24,715	21,920 ± 27,227
ICD/PM implant or revision		
Count	13	9
Mean Length-of-stay ± SD, days	1.8 ± 1.5	2.7 ± 2.4
Mean ICU stay ± SD, days	0.3 ± 0.9	0
Mean cost ± SD, $	28,965 ± 31,475	22,314 ± 20,493
Valve dysfunction/endocarditis		
Count	2	1
Mean Length-of-stay ± SD, days	20.5 ± 20.5	20.0
Mean ICU stay ± SD, days	4.0	7.0
Mean cost ± SD, $	69,221 ± 43,799	72,329
Other cardiovascular readmissions		
Count	4	4
Mean Length-of-stay ± SD, days	4.0 ± 0.8	6.3 ± 4.8
Mean ICU stay ± SD, days	1.3 ± 2.5	0.5 ± 1.0
Mean cost ± SD, $	14,992 ± 24,960	25,969 ± 27,782
All cardiovascular readmissions		
Count	93	59
Mean Length-of-stay ± SD, days	5.7 ± 8.6	5.7 ± 5.1
Mean ICU stay ± SD, days	1.7 ± 5.5	0.8 ± 2.4

Mean cost ± SD, $	19,788 ± 28,711	18,351 ± 21,140
Non-cardiovascular admissions		
Count	59	62
Mean Length-of-stay ± SD, days	8.4 ± 8.4	10.2 ± 16.2
Mean ICU stay ± SD, days	1.7 ± 3.9	8.3 ± 47.2
Mean cost ± SD, $	19,337 ± 22,677	21,612 ± 35,602
All readmissions		
Count	152	121
Mean Length-of-stay ± SD, days	6.8 ± 8.6	8.0 ± 12.3
Mean ICU stay ± SD, days	1.7 ± 5.0	4.6 ± 33.5
Mean cost ± SD, $	19,613 ± 26,454	20,022 ± 29,377

Descriptive statistics for undiscounted readmission costs were calculated using imputed datasets and applying Rubin's combination rules to obtain the mean and variance estimates. One patient who underwent mitral-valve repair had a reoperation with replacement during the index hospitalization, and three had a reoperation with replacement later on: one during an admission for heart failure and two during an admission for valve dysfunction/endocarditis. One patient in the replacement group who initially received a bioprosthesis had a late reoperation and received a mechanical valve during an admission for valve dysfunction/endocarditis.

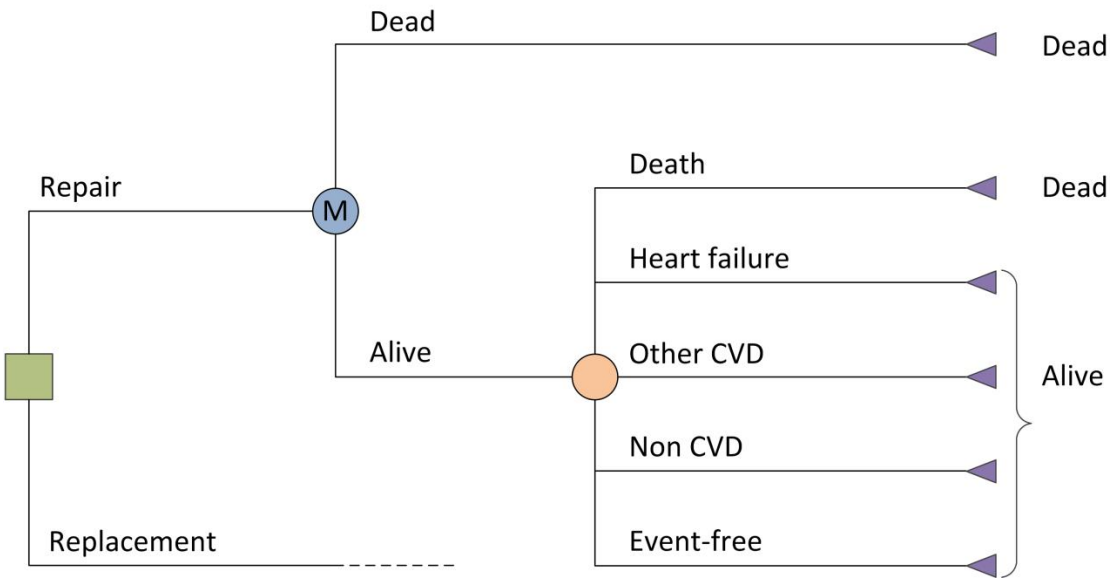
Supplemental Table VI. Sensitivity analyses of late reoperation risk in the mitral-valve replacement group using a 20-year time horizon

20-year reoperation risk in replacement	Costs repair, $	Costs replacement, $	Δ costs replacement vs repair, $	QALYs repair	QALYs replacement	Δ QALYs replacement vs repair	ICER, $/QALY replacement vs repair
Reference case model							
Baseline (2%)	127,504	125,313	-2,191	5.19	5.42	0.24	Dominant
5%	127,504	126,578	-925	5.19	5.42	0.24	Dominant
10%	127,504	128,874	1,371	5.19	5.42	0.23	5,884
15%	127,504	130,522	3,048	5.19	5.42	0.23	13,190
Clock set back to zero following reoperation, cross-over to replacement following reoperation in repair							
Baseline (2%)	128,836	125,644	-3,192	5.18	5.42	0.24	Dominant
5%	128,836	127,878	-958	5.18	5.40	0.21	Dominant
10%	128,836	132,665	3,829	5.18	5.37	0.19	20,025
15%	128,836	136,029	7,193	5.18	5.33	0.15	48,745

The 20-year cumulative incidence of reoperation was increased in the replacement group by increasing the monthly reoperation rate after 5 year of follow-up.

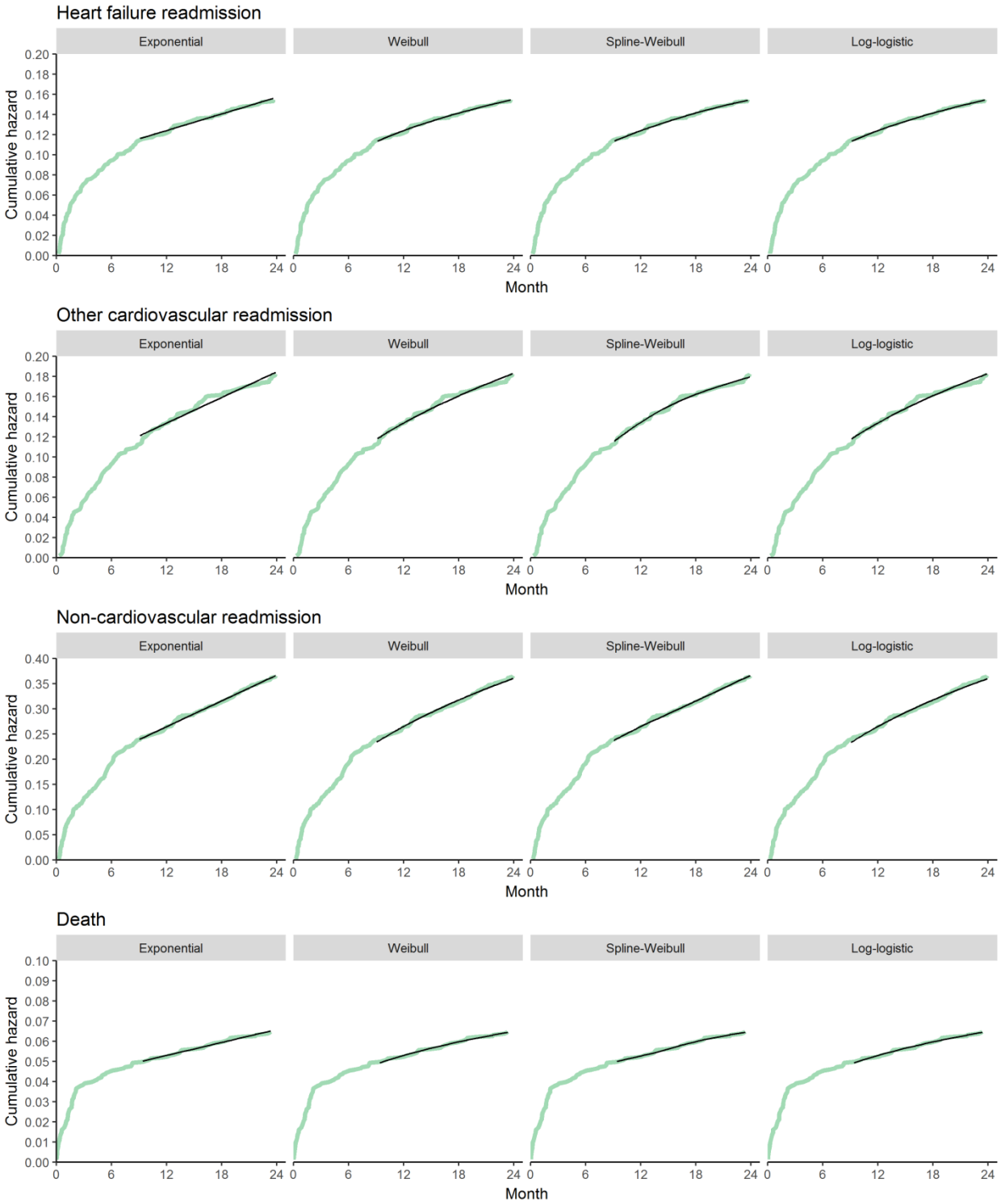
The 20-year cumulative incidence of reoperation in the repair group (5%) was not changed.

Supplemental Figure I. Schematic representation of the microsimulation model



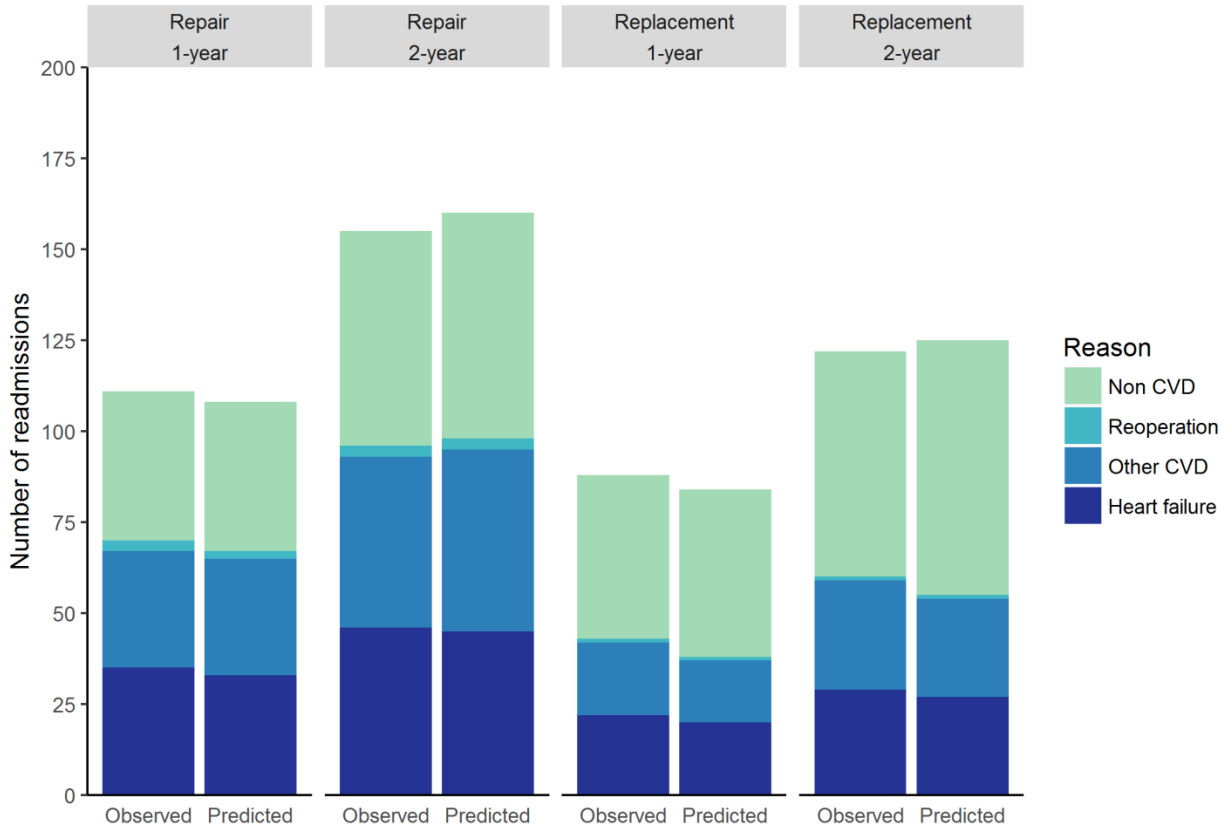
This figure schematically shows the individual-level (microsimulation) state-transition model denoted by *M*, consisting of two health states: alive and dead. For both study arms the model structure is identical, and the dashed line indicates that a clone copy of the model is used for simulating replacement.

Supplemental Figure II. Baseline cumulative hazards and hazard rate functions for extrapolation



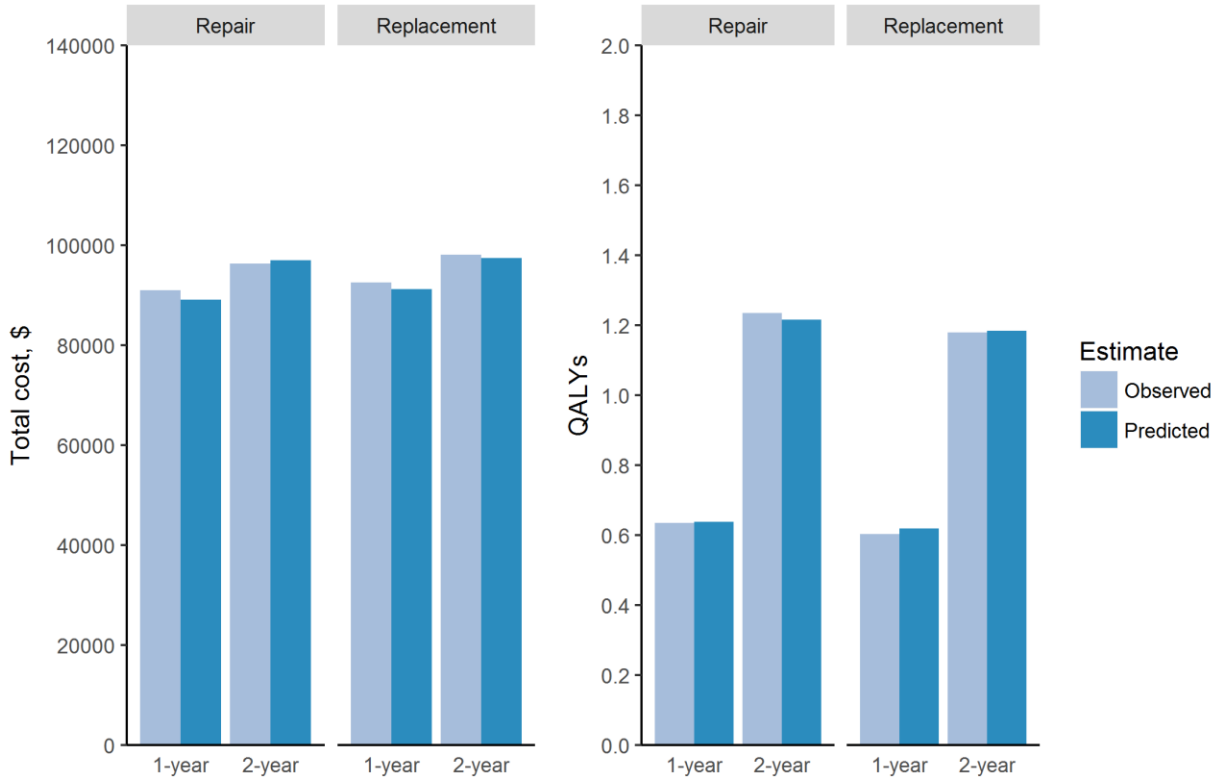
Shown are baseline cumulative hazard functions with functions for extrapolation of baseline hazard rates.

Supplemental Figure III. Validity predictions of readmissions by type



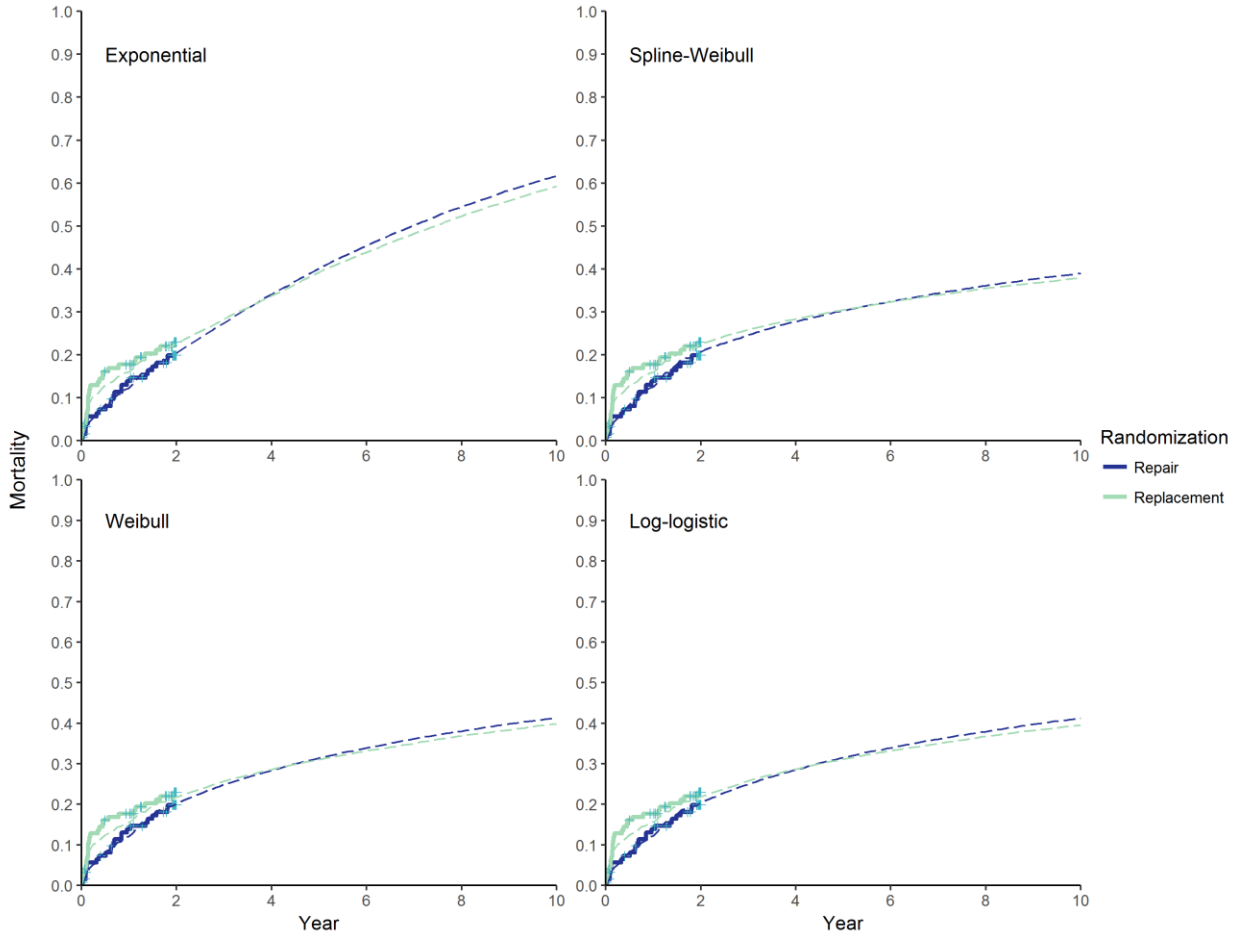
Shown are trial-based crude counts vs. counts by readmission type predicted by the microsimulation model.

Supplemental Figure IV. Validity of predictions of cost-effectiveness outcomes



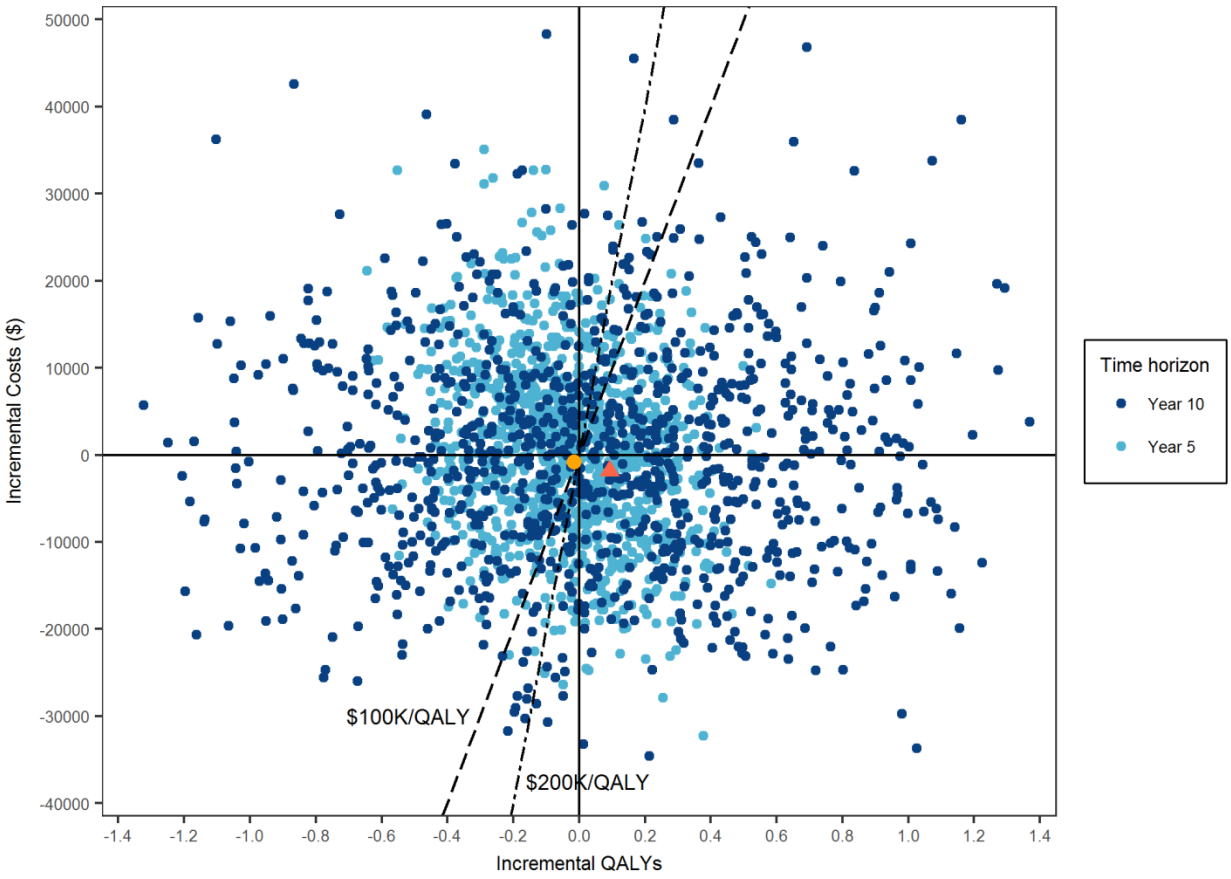
Shown are trial-based cumulative costs and QALYs vs. those predicted by the microsimulation model.

Supplemental Figure V. Observed and simulated all-cause mortality estimates assuming different survival distributions



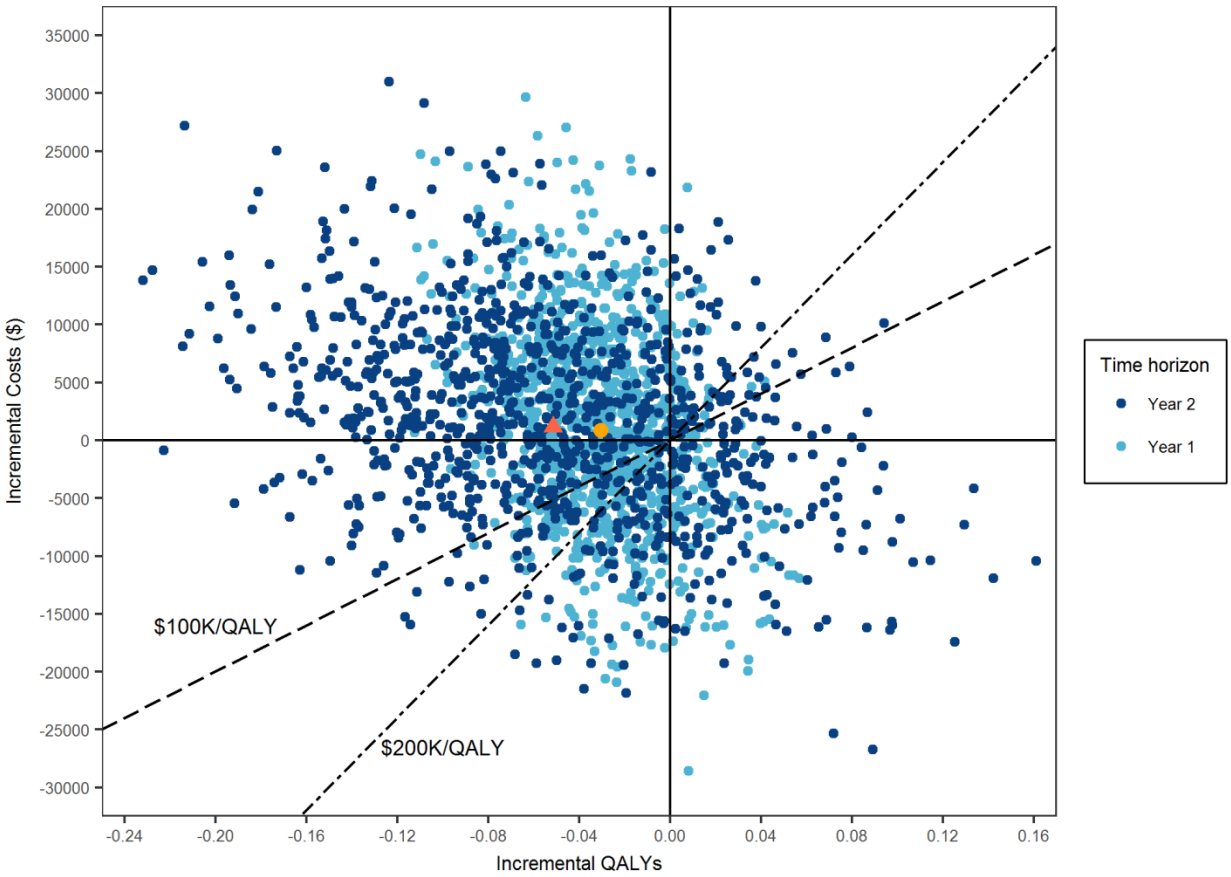
Shown are all-cause mortality estimates based trial data (solid lines) and simulated mortality estimates (dashed lines) using exponential (reference case), spline-Weibull, Weibull and log-logistic models.

Supplemental Figure VI. Long-term cost-effectiveness analysis bootstrap results comparing replacement vs repair



Shown are Δs in average costs and average QALYs as measured in each simulation with repair as the reference strategy. The yellow and red figures represent the point estimates (Δcosts, ΔQALYs) at 5-year ($-792; -0.05) and 10-year ($-1,814; +0.09) respectively. The two diagonals represent commonly used cost-effectiveness thresholds of $100K/QALY and $200K/QALY. The proportion of iterations below or to the right of the selected diagonal equals the likelihood of the replacement strategy being cost-effective as compared with repair given the applicable cost-effectiveness threshold.

Supplemental Figure VII. Within-RCT cost-effectiveness analysis bootstrap results comparing replacement vs repair assuming a gradual decline of quality-of-life prior to death



Shown are Δs in average costs and average QALYs as measured in each bootstrap replicate of the trial data with repair as the reference strategy. The yellow and red figures represent the point estimates (Δcosts, ΔQALYs) at 1-year ($848; -0.03) and 2-year ($1,166; -0.05) respectively. The two diagonals represent commonly used cost-effectiveness thresholds of $100K/QALY and $200K/QALY. The proportion of iterations below or to the right of the selected diagonal equals the likelihood of the replacement strategy being cost-effective as compared with repair given the applicable cost-effectiveness threshold.

Supplemental References

1. Royston P and Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*. 2002;21:2175-97.
2. Dunn A, Grosse SD and Zuvekas SH. Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States. *Health services research*. 2016.
3. White IR and Royston P. Imputing missing covariate values for the Cox model. *Statistics in medicine*. 2009;28:1982-98.