SUPPLEMENTAL MATERIAL

Supplemental Methods

Supplemental References

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 301 patients in the CTSN moderate 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 cardiovascular disease (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:

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

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 readmissions. 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 a difference in effect of repair on early and late mortality.

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 (**Supplemental Figure II**), 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.

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 readmissions in which a reoperation took place were based on those observed in the CTSN moderate ischemic MR trial or, when missing, on the mean values observed for the index hospitalization in the CTSN severe ischemic MR trial where mitral-valve (MV) 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 301 moderate ischemic MR trial participants (**Supplemental Figure III**) and for the 251 severe ischemic MR trial participants (**Supplemental Figure IV**). Finally, the microsimulation model showed good performance when comparing predictions of cumulative costs and cumulative QALYs to those observed in the within-trial CEA for both CTSN trials (**Supplemental Figures V-VI**).

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, 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 MR grade (6, 12, and 24 months), time of quality-of-life measurement (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 300 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 151 individuals for CABG alone and 150 for CABG plus MV repair. 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 VII**). Exponential models for readmissions were considered inappropriate because of deviations from the non-parametric 2-year cumulative hazard function (**Supplemental Figure II**).

Subsequently, we performed a number of sensitivity analyses for long-term cost-effectiveness. First, we conducted analyses varying the hazard ratio of CABG plus MV repair regarding rates of mortality, heart failure and other cardiovascular readmissions within 95% UI limits. In addition, we examined the effect of increasing the risk of reoperation in the CABG only group, and ran simulations 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 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, 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. no repair 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

Variable	Heart failure	Other CVD	Non CVD	Competing 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)
Age x moderate ischemic MR trial	1.021 (0.990 to 1.056)	-	-	-
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)
CABG plus MV repair	1.092 (0.653 to 1.825)	0.788 (0.449 to 1.478)	1.052 (0.651 to 1.665)	-
< 3 months	-	-	-	0.533 (0.108 to 1.390)
≥ 3 months	-	-	-	1.457 (0.558 to 3.331)
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% UIs are based on the 2.5th and 97.5th percentiles obtained from the 1,000 bootstrap replicates. Abbreviations: MR = mitral regurgitation; MV = mitral-valve; UI = uncertainty interval.

Supplemental Table II. Study population baseline and operative characteristics

Variable	CABG alone N=151	CABG plus MV repair N=150
Age (y), mean ± SD	65.2 ± 11.3	64.3 ± 9.6
Male, n (%)	99 (66)	106 (71)
Diabetes, n (%)	66 (44)	76 (51)
Medical history, n (%)		
Myocardial infarction	97 (64)	103 (69)
Previous CABG	4/143 (3)	4/144 (3)
Previous PCI	24 (16)	26 (17)
Heart Failure	75 (50)	82 (55)
Atrial fibrillation	35/150 (23)	19/149 (13)
ICD	6 (4)	6 (4)
Stroke	9 (6)	15 (10)
Renal insufficiency	27/150 (18)	24 (16)
Oral anticoagulation use	68 (45)	63 (42)
Antiplatelets use	128 (85)	129 (86)
LVESVI (mL/m^2) , mean ± SD	54.8 ± 24.9	59.6 ± 25.7
SF-6D health index score, mean ± SD	0.66 ± 0.13	0.66 ± 0.12
Concomitant procedures, n (%)		
Management of left atrial appendage	8 (5)	12 (8)
Atrial maze	10 (7)	11 (7)
PFO closure	3 (2)	6 (4)
Atrial septal defect repair	0	2 (1)

Abbreviations: SD = standard deviation; SF-6D = Short-Form Six-Dimension; ICD = implantable cardioverter defibrillator; LVESVI = left ventricular end-systolic volume index; MV = mitral-valve; PCI = percutaneous coronary intervention; PFO, patent foramen ovale.

Supplemental Table III. Resource use associated with the index hospitalization

Variable	CABG alone N=151	CABG plus MV repair N=150
Number of grafts	3.3 ± 0.9	3.2 ± 0.9
Cardiopulmonary bypass time, min	107 ± 50	163 ± 55
Time to discharge, days (post-randomization)	8.4 ± 5.9	10.3 ± 8.2
ICU stay, days (post-randomization)	3.9 ± 5.7	4.8 ± 6.1
Discharge disposition, n (%)		
Home	112 (74)	106 (71)
Skilled nursing/inpatient rehabilitation facility	29 (19)	37 (24)
Hospice	1 (1)	0 (0)
Death	4 (3)	1 (1)
Other	5 (3)	6 (4)

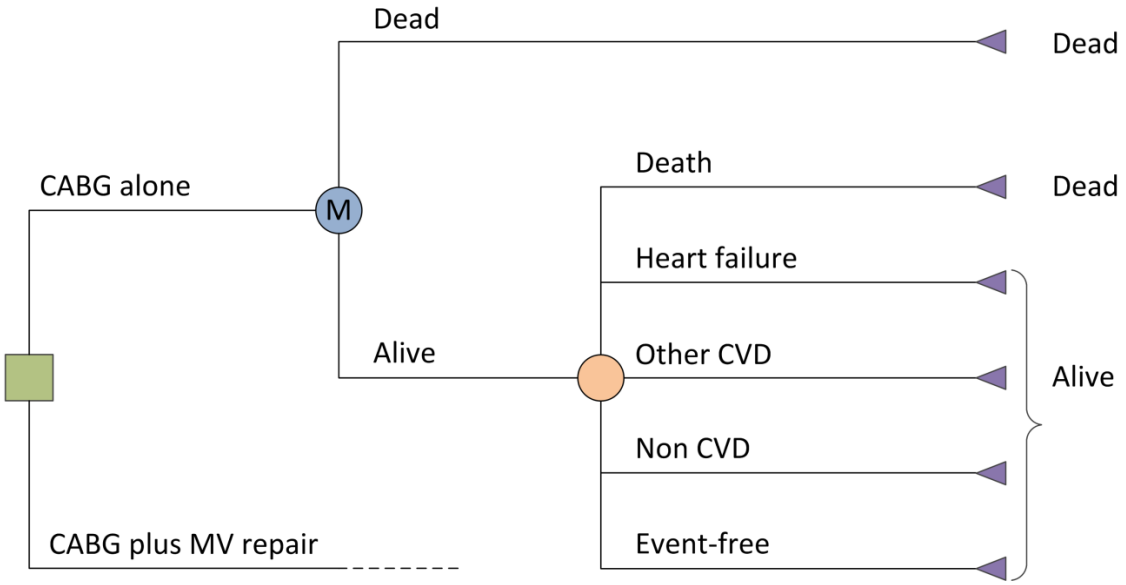
Abbreviations: ICU = intensive care unit; IQR = interquartile range; MV = mitral-valve; SD = standard deviation.

Supplemental Table IV: Resource use and costs associated with readmissions over 2-year follow-up

Readmission	CABG alone	CABG plus mitral-valve repair
Heart failure		
Count	35	36
Mean Length-of-stay ± SD, days	6.4 ± 5.5	10.8 ± 11.1
Mean ICU stay ± SD, days	0.9 ± 2.5	1.6 ± 3.8
Mean cost ± SD, $	14,043 ± 17,697	29,348 ± 45,807
MI/Angina		
Count	13	1
Mean Length-of-stay ± SD, days	2.3 ± 1.2	1.0
Mean ICU stay ± SD, days	0.2 ± 0.6	0
Mean cost ± SD, $	12,473 ± 10,563	12,187
Revascularization (CABG/PCI)		
Count	4	0
Mean Length-of-stay ± SD, days	2.5 ± 3.0	-
Mean ICU stay ± SD, days	0.3 ± 0.5	-
Mean cost ± SD, $	18,395 ± 15,113	-
Arrhythmia		
Count	4	9
Mean Length-of-stay ± SD, days	6.8 ± 2.6	5.2 ± 3.9
Mean ICU stay ± SD, days	0	0.6 ± 1.2
Mean cost ± SD, $	20,110 ± 23,422	22,410 ± 20,683
ICD/PM implant or revision		
Count	10	7
Mean Length-of-stay ± SD, days	2.4 ± 4.1	1.3 ± 0.8
Mean ICU stay ± SD, days	0.2 ± 0.4	0
Mean cost ± SD, $	25,905 ± 23,115	18,083 ± 11,993
Valve dysfunction/endocarditis		
Count	0	2
Mean Length-of-stay ± SD, days	-	20.5 ± 0.7
Mean ICU stay ± SD, days	-	11.5 ± 12.0
Mean cost ± SD, $	-	120,435 ± 57,297
Other cardiovascular readmissions		
Count	5	9
Mean Length-of-stay ± SD, days	7.8 ± 4.2	5.1 ± 4.7
Mean ICU stay ± SD, days	1.6 ± 2.6	0.7 ± 2.0
Mean cost ± SD, $	50,813 ± 73,142	25,718 ± 38,600
All cardiovascular readmissions		
Count	71	64
Mean Length-of-stay ± SD, days	5.0 ± 4.8	8.3 ± 9.6
Mean ICU stay ± SD, days	0.6 ± 1.9	1.5 ± 3.9
Mean cost ± SD, $	18,602 ± 26,080	29,208 ± 42,262
Non-cardiovascular readmissions		
Count	62	65
Mean Length-of-stay ± SD, days	7.0 ± 6.0	7.3 ± 5.9
Mean ICU stay ± SD, days	0.3 ± 1.1	0.4 ± 1.2
Mean cost ± SD, $	20,760 ± 28,480	21,902 ± 27,618
All readmissions		
Count	133	129
Mean Length-of-stay ± SD, days	5.9 ± 5.4	7.8 ± 7.9
Mean ICU stay ± SD, days	0.5 ± 1.6	0.9 ± 2.9
Mean cost ± SD, $	19,608 ± 27,142	25,527 ± 35,692

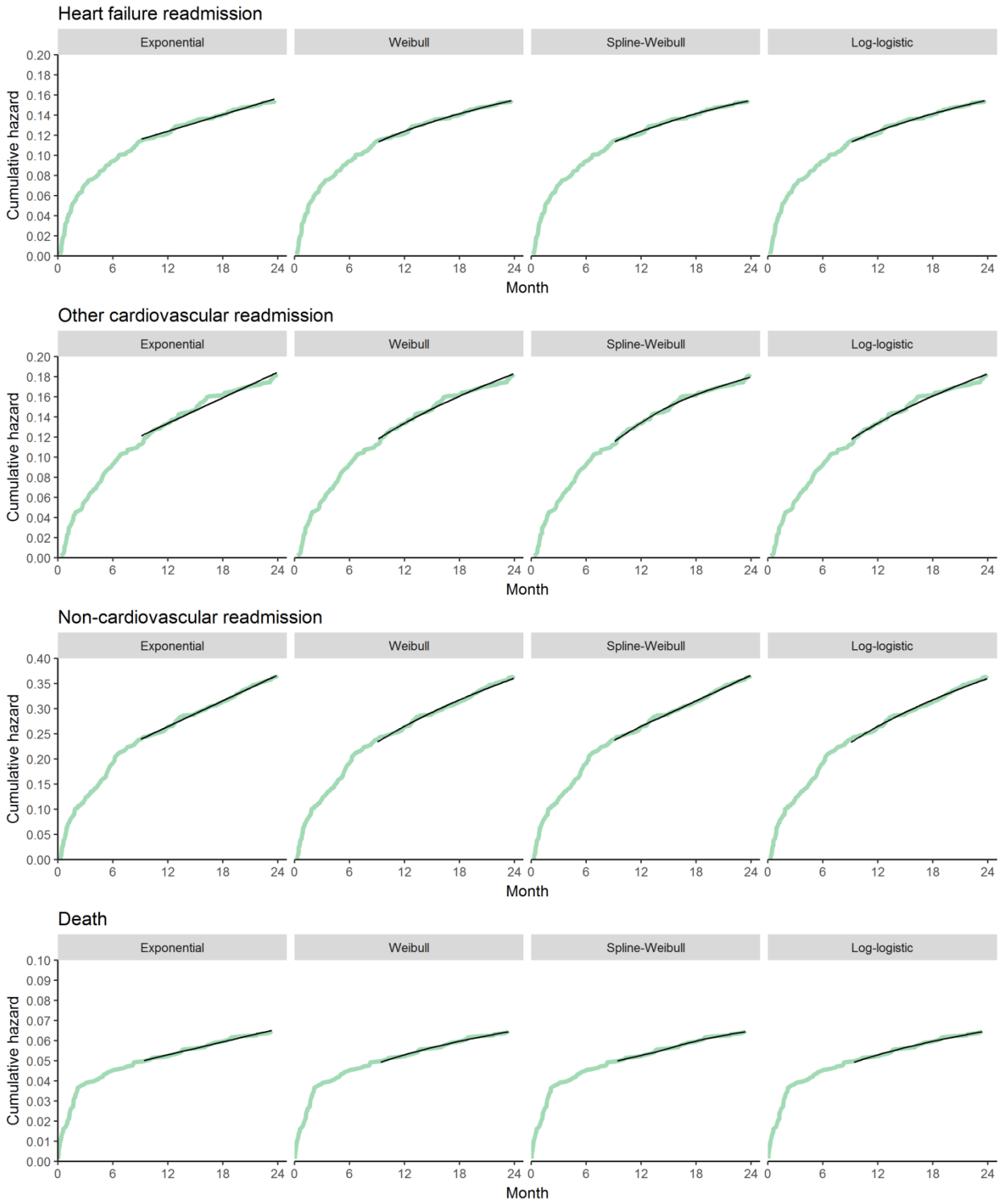
Descriptive statistics for undiscounted readmission costs were calculated using imputed datasets and applying Rubin's combination rules to obtain the mean and variance estimates.

Supplemental Figure I. Schematic representation of the microsimulation model



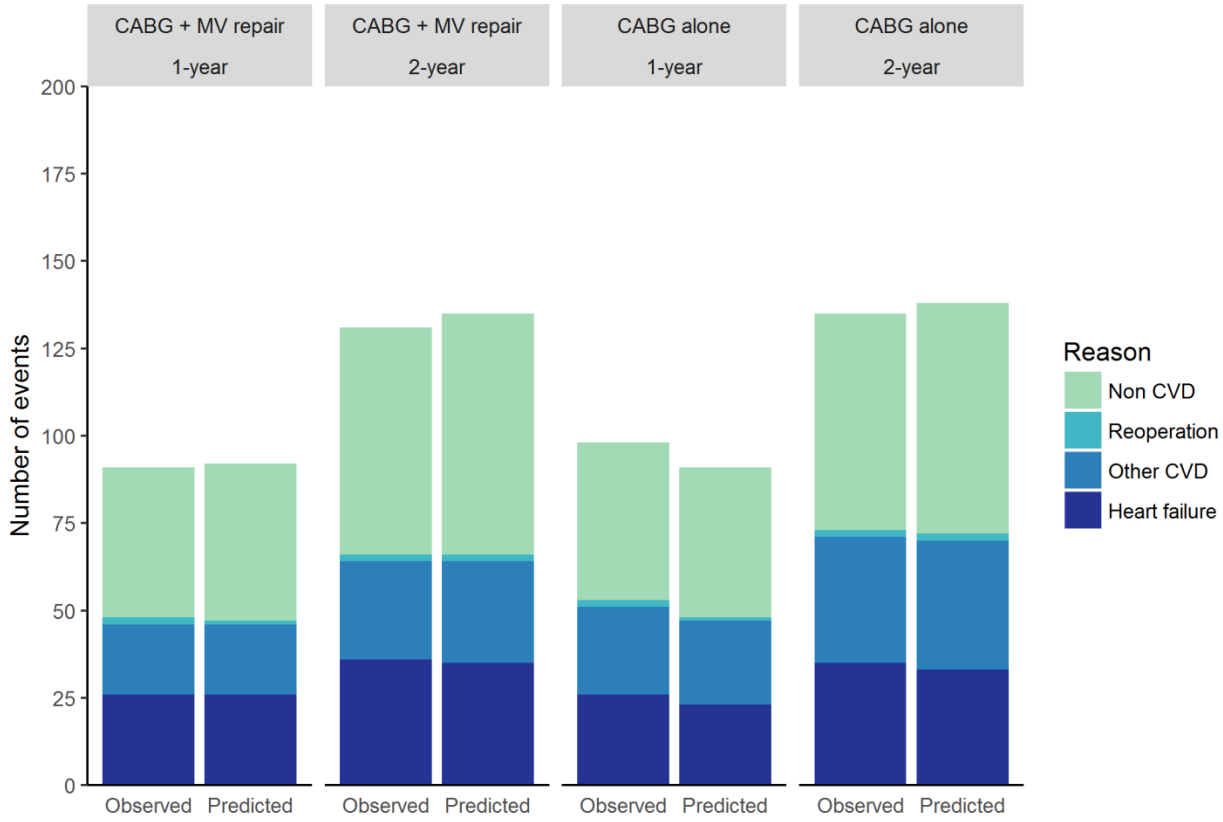
This figure schematically shows the individual-level (microsimulation) state-transition model denoted by *M*, consisting of two health states: alive and dead. For both CTSN moderate ischemic MR trial study arms the model structure is identical, and the dashed line indicates that a clone copy of the model is used for simulating CABG plus MV repair. Abbreviations: MV = mitral-valve.

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



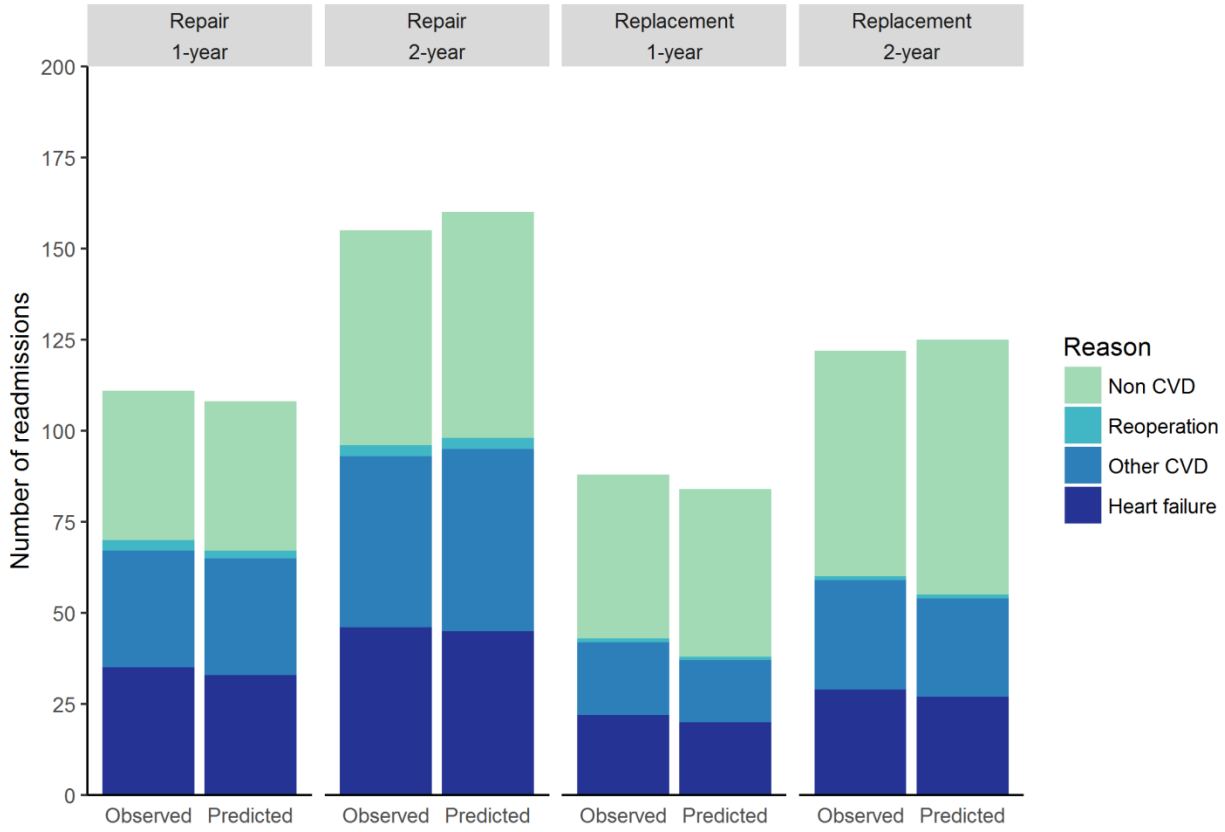
Shown are baseline cumulative hazard functions specified for a 60-year old female in the CABG alone group of the MMR trial with functions for extrapolation of baseline hazard rates.

Supplemental Figure III. Validity predictions of readmissions by type for moderate ischemic mitral regurgitation (CTSN moderate ischemic MR trial)



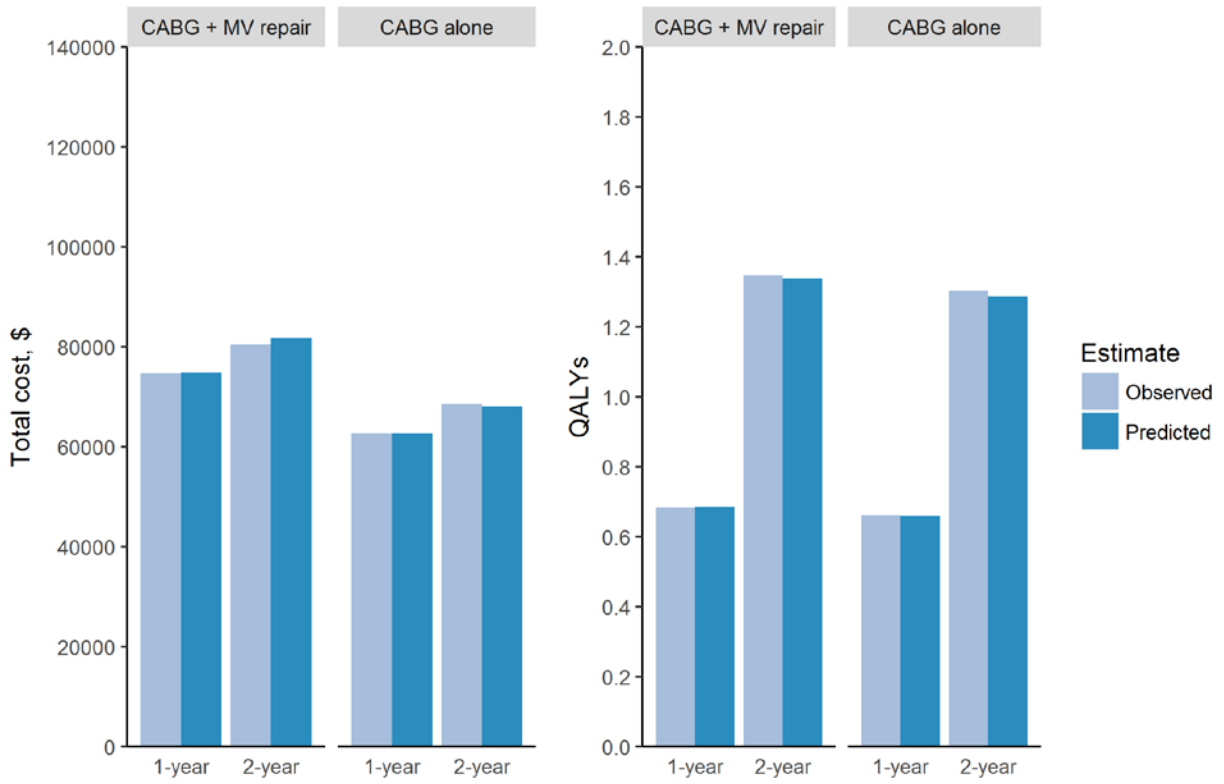
Shown are trial-based crude counts vs counts by readmission type predicted by the microsimulation model. Abbreviations: MV = mitral-valve.

Supplemental Figure IV. Validity predictions of readmissions by type for severe ischemic mitral regurgitation (CTSN severe ischemic MR trial)



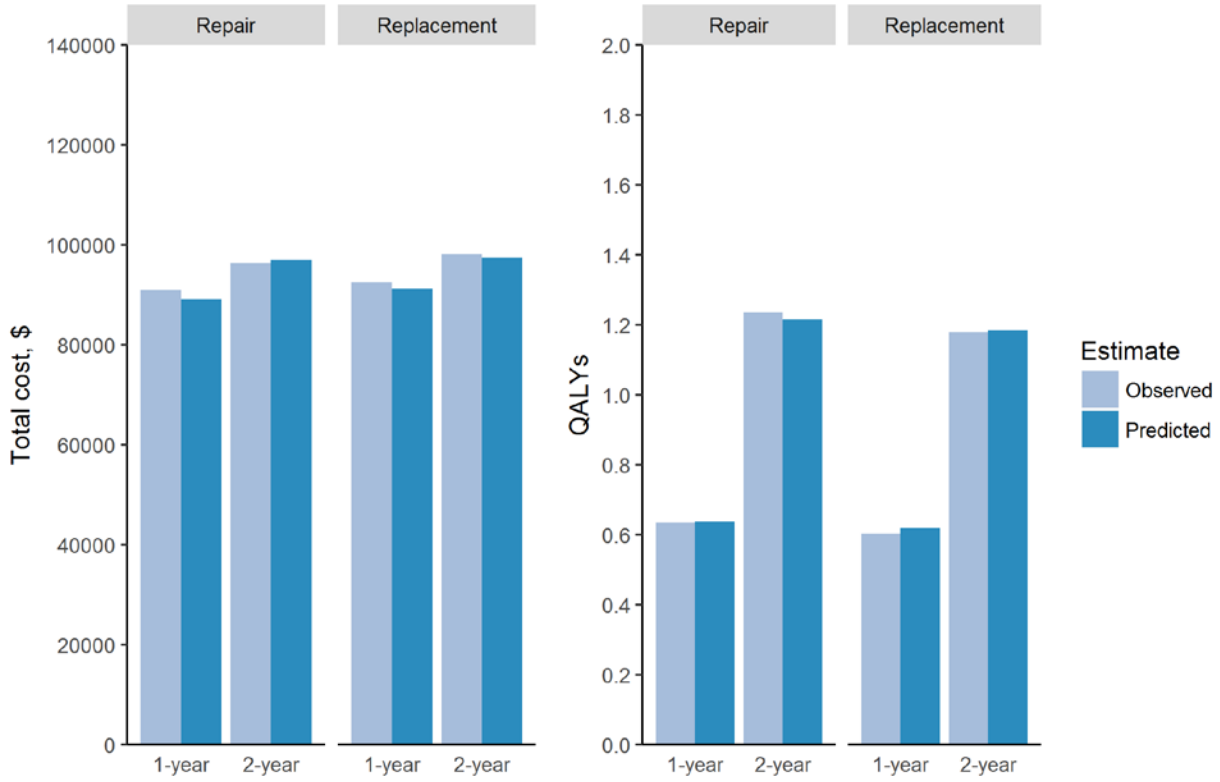
Shown are trial-based crude counts vs counts by readmission type predicted by the microsimulation model.

Supplemental Figure V. Validity of predictions of cost-effectiveness outcomes for moderate ischemic mitral regurgitation (CTSN moderate ischemic MR trial)



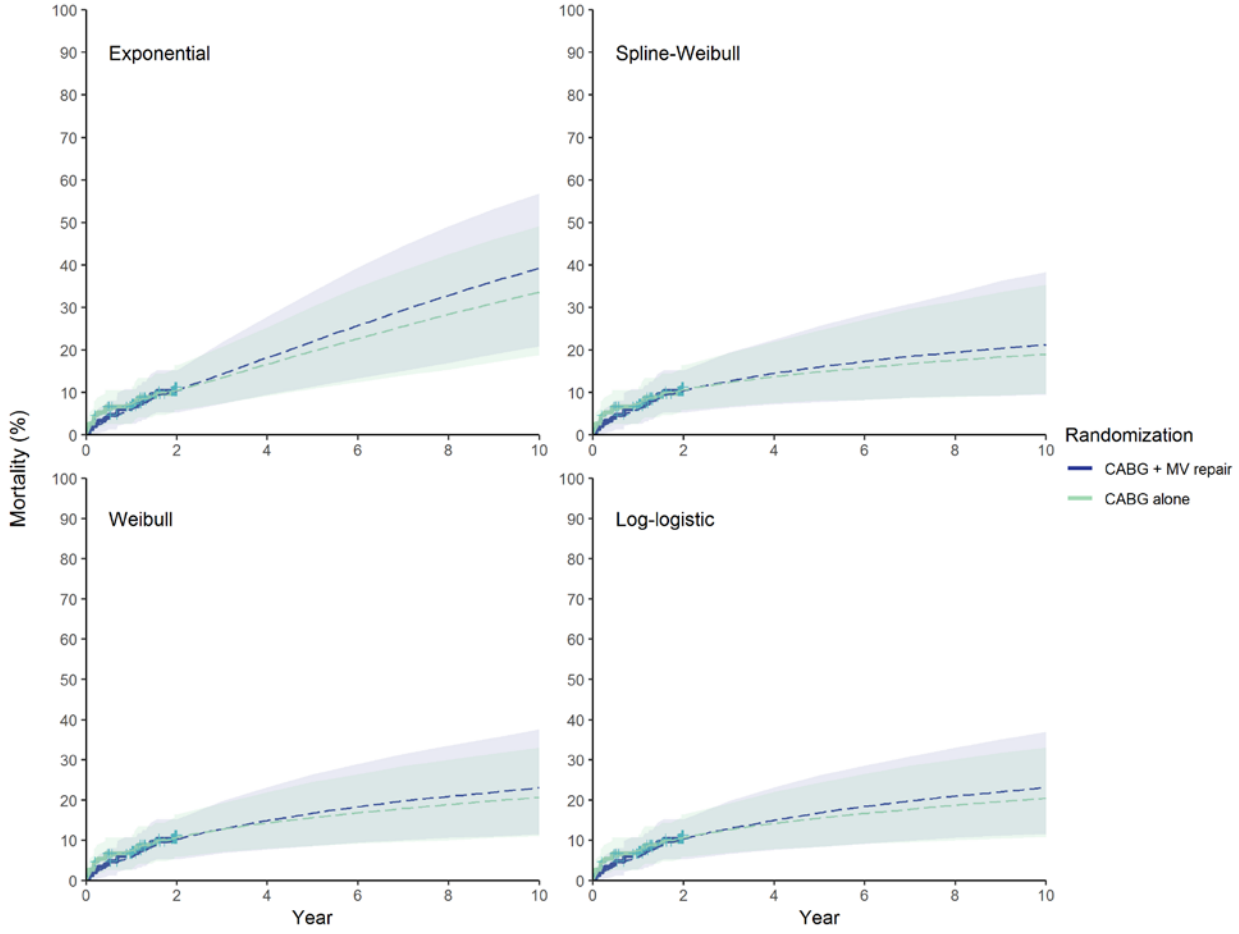
Shown are trial-based cumulative costs and QALYs vs those predicted by the microsimulation model.
Abbreviations: MV = mitral-valve; QALY = quality-adjusted life year.

Supplemental Figure VI. Validity of predictions of cost-effectiveness outcomes for severe ischemic mitral regurgitation (CTSN severe ischemic MR trial)



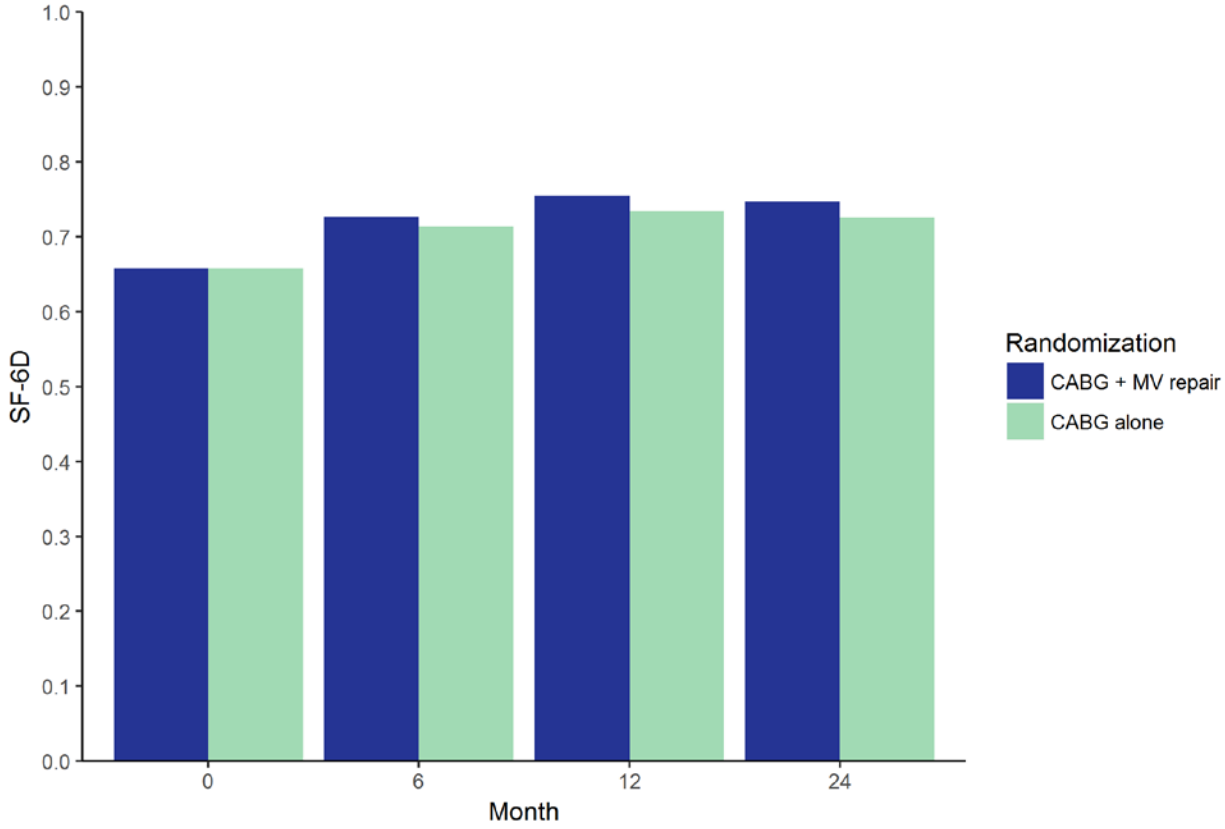
Shown are trial-based cumulative costs and QALYs vs those predicted by the microsimulation model.
Abbreviations: QALY = quality-adjusted life year.

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



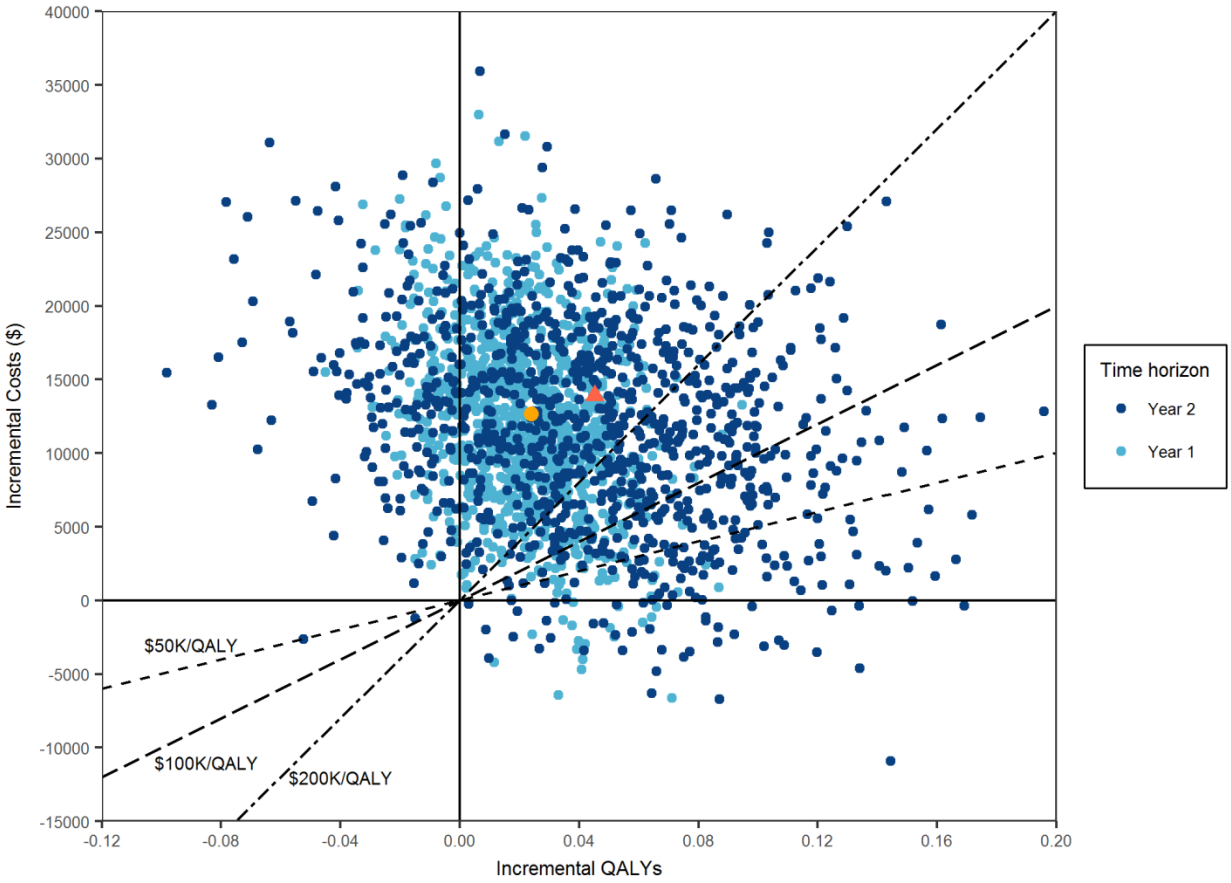
Shown are all-cause mortality estimates with 95% uncertainty intervals based on trial data (solid lines) and simulated mortality estimates (dashed lines) using exponential (reference case), spline-Weibull, Weibull and log-logistic models for extrapolation beyond two years. Numbers at risk with accounting for censored observations for the first 2 years of follow-up are shown in Figure 2. Abbreviations: MV = mitral-valve.

Supplemental Figure VIII. SF-6D utility index by study arm



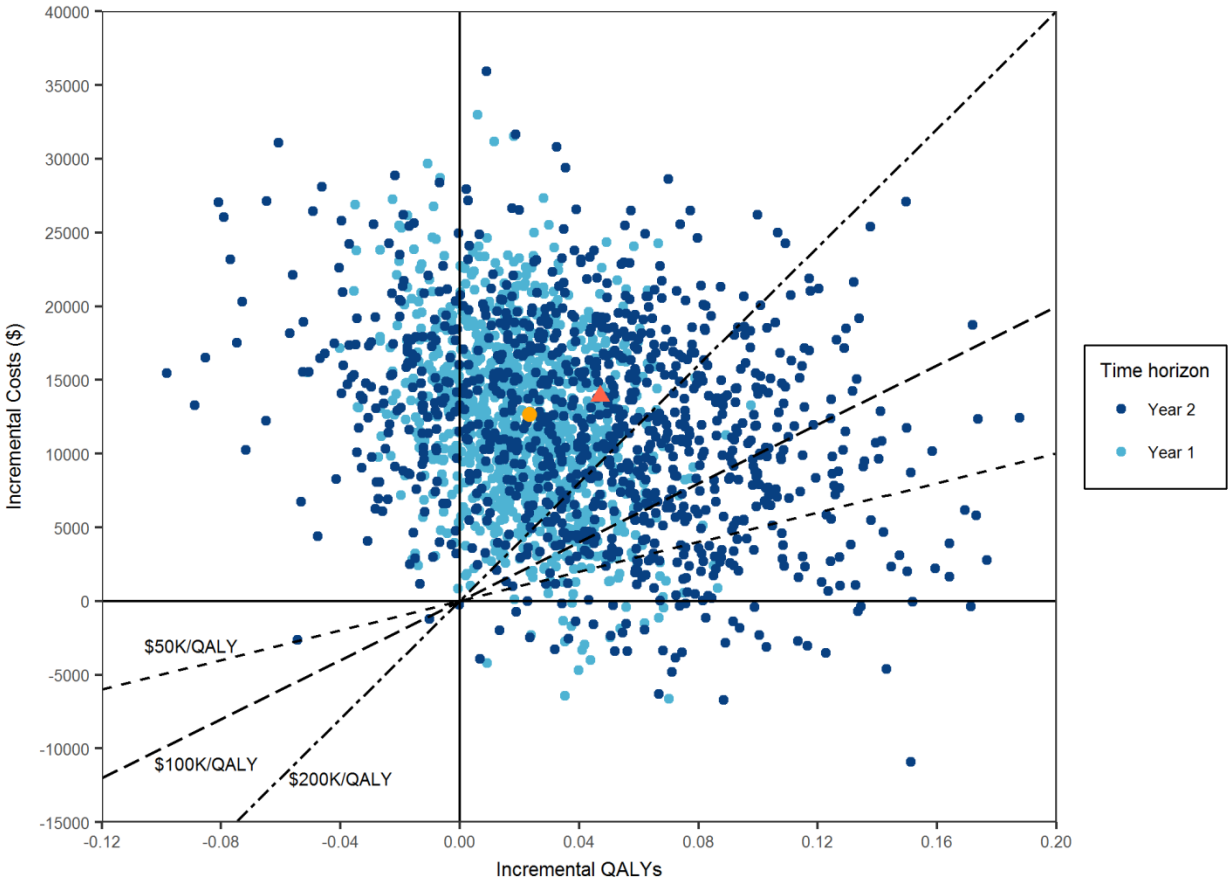
Shown are mean SF-6D utility index scores in N=150 for CABG plus mitral-valve repair and N=151 for CABG alone. Abbreviations: MV, mitral-valve.

Supplemental Figure IX. Within-trial cost-effectiveness analysis bootstrap results CABG plus mitral-valve repair vs CABG alone



Shown are differences in average costs and average QALYs as measured in each bootstrap replicate of the trial data with repair as the reference strategy. Trial data were bootstrapped 1,000 times. The yellow and red figures represent the point estimates (Δcosts, ΔQALYs) at 1-year ($12,656; 0.02) and 2-year ($13,922; 0.05) respectively. The three diagonals represent commonly used cost-effectiveness thresholds of $50K/QALY, $100K/QALY and $200K/QALY. The proportion of iterations below or to the right of the selected diagonal equals the likelihood of the CABG plus mitral-valve repair strategy being cost-effective as compared with CABG alone given the applicable cost-effectiveness threshold. Abbreviations: QALY, quality-adjusted life year.

Supplemental Figure X. Within-trial cost-effectiveness analysis bootstrap results CABG plus mitral-valve repair vs CABG alone assuming a gradual decline of quality-of-life prior to death



Shown are differences in average costs and average QALYs as measured in each bootstrap replicate of the trial data with repair as the reference strategy. Trial data were bootstrapped 1,000 times. The yellow and red figures represent the point estimates (Δcosts, ΔQALYs) at 1-year ($12,656; 0.02) and 2-year ($13,922; 0.05) respectively. The three diagonals represent commonly used cost-effectiveness thresholds of $50K/QALY, $100K/QALY and $200K/QALY. The proportion of iterations below or to the right of the selected diagonal equals the likelihood of the CABG plus mitral-valve repair strategy being cost-effective as compared with CABG alone given the applicable cost-effectiveness threshold. Abbreviations: QALY = quality-adjusted life year.

Supplemental References

1. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21:2175-2197.

2. Dunn A, Grosse SD, Zuvekas SH. Adjusting health expenditures for inflation: A review of measures for health services research in the united states. Health Serv Res. 2016.

3. White IR, Royston P. Imputing missing covariate values for the cox model. Stat Med. 2009;28:1982-1998.