

## Appendix

### STS Individual Surgeon Composite Measure for Adult Cardiac Surgery

#### S.4. – S.11. Measure Specifications

Due to the complex methodology used to construct the composite measure, it is impractical to separately discuss the numerator and denominator. The following discussion describes in detail this multiprocedural, multidimensional composite measure.

The STS Individual Surgeon Composite Measure for Adult Cardiac Surgery includes five major procedures, i.e., isolated coronary artery bypass grafting (CABG), isolated aortic valve replacement (AVR), AVR+CABG, isolated mitral valve repair or replacement (MVRR), and MVRR+CABG, and comprises the following two domains:

##### Domain 1 – Risk-Adjusted Operative Mortality

Operative mortality is defined as death before hospital discharge or within 30 days of the operation.

##### Domain 2 – Risk-Adjusted Major Morbidity

Major morbidity is defined as the occurrence of any one or more of the following major complications:

1. Prolonged ventilation,
2. Deep sternal wound infection,
3. Permanent stroke,
4. Renal failure, and
5. Reoperations for bleeding, coronary graft occlusion, prosthetic or native valve dysfunction, and other cardiac reasons, but not for other non-cardiac reasons.

Individual surgeons with at least 100 eligible cases during the 3-year measurement window will receive a score for each domain and an overall composite score. In addition to calculating composite score point estimates with credible intervals, surgeons will be assigned rating categories designated by the following:

- 1 star – lower-than-expected performance
- 2 stars – as-expected performance
- 3 stars – higher-than-expected performance

**Patient Population:** The analysis population consists of patients aged 18 years or older who undergo isolated CABG, isolated AVR, AVR+CABG, isolated MVRR, and MVRR+CABG.

**Time Window:** 3 years

By including composite performance scores for a portfolio of five procedures that account for nearly 80% of a typical STS Adult Cardiac Surgery Database participant surgeon's clinical activity, this metric provides a more balanced and comprehensive perspective than focusing on just one procedure or one end point. Recognizing that surgeons' practices vary, each surgeon's composite performance is implicitly "weighted" by the proportion of each type of procedure he or she performs. For instance, the results of surgeons who primarily perform mitral procedures are affected most by their mitral surgery results. This approach is especially relevant for surgeons with highly specialized practices who may do relatively few isolated CABG procedures and whose performance would thus be difficult to assess using a CABG measure only. Finally, performance on each of these procedures is estimated using risk models specific to those procedures, in most cases the exact or slightly modified versions of previously published models (references provided below).

##### Final Composite Score:

The overall composite score was calculated as a weighted sum of (1 minus risk-adjusted mortality rate) and (1 minus risk-adjusted major morbidity rate). Mortality and morbidity rates were weighted inversely by their respective standard deviations across surgeons. This procedure is equivalent to first rescaling mortality and morbidity rates by their respective standard deviations across surgeons and then assigning equal weighting to the rescaled mortality rate and rescaled morbidity rate. Standard deviations derived from the data were used

to define the final composite measure as  $0.81 \times (1 \text{ minus risk-standardized mortality rate}) + 0.19 \times (1 \text{ minus risk-standardized complication rate})$ .

Details regarding the current STS adult cardiac surgery risk models can be found in the following manuscripts:

- Shahian DM, O'Brien SM, Filardo G, Ferraris VA, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1--coronary artery bypass grafting surgery. *Ann Thorac Surg*. 2009 Jul;88(1 Suppl):S2-22.
- O'Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac Surg* 2009;88(1 Suppl):S23–42.
- Shahian DM, O'Brien SM, Filardo G, Ferraris VA, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3--valve plus coronary artery bypass grafting surgery. *Ann Thorac Surg* 2009 Jul;88(1 Suppl):S43-62.

Additional details regarding the Individual Surgeon Composite Measure for Adult Cardiac Surgery are provided in the attached manuscript:

Shahian DM, He X, Jacobs JP, Kurlansky PA, Badhwar V, Cleveland JC Jr, Fazzalari FL, Filardo G, Normand SL, Furnary AP, Magee MJ, Rankin JS, Welke KF, Han J, O'Brien SM. The Society of Thoracic Surgeons Composite Measure of Individual Surgeon Performance for Adult Cardiac Surgery: A Report of The Society of Thoracic Surgeons Quality Measurement Task Force. *Ann Thorac Surg*. 2015;100:1315-25.

**S.14. Identify the statistical risk model method and variables** (Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)

Please see section 2b4.3 of the measure testing attachment.

The STS Adult Cardiac Surgery Database risk model manuscripts were published in 2009. Detailed information is provided in these manuscripts, and the definitions of all the variables in the final 2008 CABG, valve, and valve+CABG models are provided below. Note: not all were included in the final models.

**CABG**

Variable	Definition
Intercept	= 1 for all patients
Atrial fibrillation	= 1 if patient has history of preoperative atrial fibrillation, = 0 otherwise
Age	= Patient age in years
Age function 1	= max (age-50, 0)
Age function 2	= max (age-60, 0)
Age by reop function	= Age function 1 if surgery is a reoperation, = 0 otherwise
Age by status function	= Age function 1 if status is emergent or salvage, = 0 otherwise
BSA function 1	= max (1.4, min [2.6, BSA]) - 1.8
BSA function 2	= (BSA function 1) <sup>2</sup>
CHF but not NYHA IV	= 1 if patient has CHF and is not NYHA class IV, = 0 otherwise
CHF and NYHA IV	= 1 if patient has CHF and is NYHA class IV, = 0 otherwise
CLD mild	= 1 if patient has mild chronic lung disease, = 0 otherwise
CLD moderate	= 1 if patient has moderate chronic lung disease, = 0 otherwise
CLD severe	= 1 if patient has severe chronic lung disease, = 0 otherwise
Creatinine function 1	= max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise
Creatinine function 2	= max ([creatinine function 1] - 1.0, 0)
Creatinine function 3	= max ([creatinine function 1] - 1.5, 0)
CVD without prior CVA	= 1 if patient has history of CVD and no prior CVA, = 0 otherwise
CVD and prior CVA	= 1 if patient has history of CVD and a prior CVA, = 0 otherwise
Diabetes, noninsulin	= 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin	= 1 if patient has diabetes treated with insulin, = 0 otherwise
Dialysis	= 1 if patient requires dialysis preoperatively, = 0 otherwise
Ejection fraction function	= max (50 - ejection fraction, 0)
Female	= 1 if patient is female, = 0 otherwise
Female by BSA function 1	= BSA function 1 if female, = 0 otherwise
Female by BSA function 2	= BSA function 2 if female, = 0 otherwise
Hypertension	= 1 if patient has hypertension, = 0 otherwise
IABP or inotropes	= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
Immunosuppressive treatment	= 1 if patient given immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency, aortic	= 1 if patient has at least moderate aortic insufficiency, = 0 otherwise
Insufficiency, mitral	= 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency, tricuspid	= 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease	= 1 if patient has left main disease, = 0 otherwise
MI 1 to 21 days	= 1 if history of MI 1 to 21 days prior to surgery, = 0 otherwise
MI > 6 and < 24 hours	= 1 if history of MI >6 and <24 hours prior to surgery, = 0 otherwise
MI ≤ 6 hours	= 1 if history of MI ≤ 6 hours prior to surgery, = 0 otherwise
No. diseased vessel function	= 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise
PCI ≤ 6 hours	= 1 if patient had PCI ≤ 6 hours prior to surgery, = 0 otherwise
Peripheral vascular disease	= 1 if patient has peripheral vascular disease, = 0 otherwise
Race black	= 1 if patient is black, = 0 otherwise
Race Hispanic	= 1 if patient is nonblack Hispanic, = 0 otherwise
Race Asian	= 1 if patient is nonblack, non-Hispanic, and is Asian, = 0 otherwise
Reop, 1 previous operation	= 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, ≥ 2 previous operations	= 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Shock	= 1 if patient was in shock at time of procedure, = 0 otherwise
Status urgent	= 1 if status is urgent, = 0 otherwise
Status emergent	= 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage	= 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Stenosis aortic	= 1 if patient has aortic stenosis, = 0 otherwise
Unstable angina	= 1 if patient has unstable angina, no MI within 7 days of surgery, = 0 otherwise

## Valve

Variable	Definition
Intercept	= 1 for all patients
Atrial fibrillation	= 1 if patient has history of preop atrial fibrillation, = 0 otherwise
Age function 1	= max (age - 50, 0)
Age function 3	= max (age - 75, 0)
Age by reoperation function	= Age function 1 if surgery is a reoperation, = 0 otherwise
Age by status function	= Age function 1 if status is emergent or salvage, = 0 otherwise
Age by MVR function	= Age function 1 if operation is MVR, = 0 otherwise
Age by MVRRepair function	= Age function 1 if operation is MVRRepair, = 0 otherwise
BSA function 1	= max (1.4, min [2.6, BSA]) - 1.8
BSA function 2	= (BSA function 1) <sup>2</sup>
CHF but not NYHA IV	= 1 if patient has CHF and is not NYHA class IV, = 0 otherwise
CHF and NYHA IV	= 1 if patient has CHF and is NYHA class IV, = 0 otherwise
CHF by MVR function	= 1 if patient has CHF and operation is MVR, = 0 otherwise
CHF by MVRRepair function	= 1 if patient has CHF and operation is MVRRepair, = 0 otherwise
CLD function	= 0 if no CLD, = 1 if mild CLD, = 2 if moderate CLD, = 3 if severe CLD
CLD by MVR function	= CLD function if operation is MVR, = 0 otherwise
CLD by MVRRepair function	= CLD function if operation is MVRRepair, = 0 otherwise
Creatinine function 1	= max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise
CVD without prior CVA	= 1 if patient has history of CVD and no prior CVA, = 0 otherwise
CVD and prior CVA	= 1 if patient has history of CVD and a prior CVA, = 0 otherwise
Diabetes, noninsulin	= 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin	= 1 if patient has diabetes treated with insulin, = 0 otherwise
Diabetes by MVR function	= 1 if patient has diabetes and operation is MVR, = 0 otherwise
Diabetes by MVRRepair function	= 1 if patient has diabetes and operation is MVRRepair, = 0 otherwise
Dialysis	= 1 if patient requires dialysis preoperatively, = 0 otherwise
Dialysis by MVR function	= 1 if patient has history of dialysis and operation is MVR, = 0 otherwise
Dialysis by MVRRepair function	= 1 if patient has history of dialysis and operation is MVRRepair, = 0 otherwise
Ejection fraction function	= max (50-ejection fraction, 0)
Endocarditis, active	= 1 if patient has active endocarditis, = 0 otherwise
Female	= 1 if patient is female, = 0 otherwise
Female by MVR function	= 1 if female and operation is MVR, = 0 otherwise
Female by MVRRepair function	= 1 if female and operation is MVRRepair, = 0 otherwise
Female by BSA function 1	= BSA function 1 if female, = 0 otherwise
Female by BSA function 2	= BSA function 2 if female, = 0 otherwise
Hypertension	= 1 if patient has hypertension, = 0 otherwise
IABP or inotropes	= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
IABP by MVR function	= 1 if patient requires preop IABP/inotropes and operation is MVR, = 0 otherwise
IABP by MVRRepair function	= 1 if patient requires preop IABP/inotropes and operation is MVRRepair, = 0 otherwise
Immunosuppressive treatment	= 1 if patient received immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency mitral	= 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency tricuspid	= 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease	= 1 if patient has left main disease, = 0 otherwise
MI ≤ 21 days	= 1 if patient has history of MI within 21 days of surgery, = 0 otherwise
MVR	= 1 if valve operation is mitral valve replacement, = 0 otherwise
MVRRepair	= 1 if valve operation is mitral valve repair, = 0 otherwise
No. diseased vessel function	= 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise
Peripheral vascular disease	= 1 if patient has peripheral vascular disease, = 0 otherwise
Race black	= 1 if patient is black, = 0 otherwise
Race Hispanic	= 1 if patient is nonblack Hispanic, = 0 otherwise
Reop, 1 prior operation	= 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, ≥ 2 prior operations	= 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Reop by MVR function	= 1 if surgery is a reoperation and operation is MVR, = 0 otherwise
Reop by MVRRepair function	= 1 if surgery is a reoperation and operation is MVRRepair, = 0 otherwise
Shock	= 1 if patient was in shock at time of procedure, = 0 otherwise
Shock by MVR function	= 1 if shock and operation is MVR, = 0 otherwise
Shock by MVRRepair function	= 1 if shock and operation is MVRRepair, = 0 otherwise
Status urgent	= 1 if status is urgent, = 0 otherwise
Status emergent	= 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage	= 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Status by MVR function	= 1 if status is emergent or salvage and operation is MVR, = 0 otherwise
Status by MVRRepair function	= 1 if status is emergent or salvage and operation is MVRRepair, = 0 otherwise
Stenosis aortic	= 1 if patient has aortic stenosis, = 0 otherwise
Stenosis mitral	= 1 if patient has mitral stenosis, = 0 otherwise
Unstable angina	= 1 if patient has unstable angina, no MI within 7 days of surgery, = 0 otherwise



## VALVE+CABG

Variable	Definition
Intercept	= 1 for all patients
Preoperative AFib	= 1 if patient has history of preoperative atrial fibrillation, = 0 otherwise
Age function 1	= max (age - 50, 0)
Age function 3	= max (age - 75, 0)
Age by reop function	= Age function 1 if surgery is a reoperation, = 0 otherwise
Age by status function	= Age function 1 if status is emergent or salvage, = 0 otherwise
Age by MVR function	= Age function 1 if operation is MVR, = 0 otherwise
Age by MVRepair function	= Age function 1 if operation is MVRepair, = 0 otherwise
BSA function 1	= max (1.4, min [2.6, BSA]) - 1.8
BSA function 2	= (BSA function 1) <sup>2</sup>
CHF but not NYHA IV	= 1 if patient has CHF and is not NYHA class IV, = 0 otherwise
CHF and NYHA IV	= 1 if patient has CHF and is NYHA class IV, = 0 otherwise
CHF by MVR function	= 1 if patient has CHF and operation is MVR, = 0 otherwise
CHF by MVRepair function	= 1 if patient has CHF and operation is MVRepair, = 0 otherwise
CLD function	= 0 if no CLD, = 1 if mild CLD, = 2 if moderate CLD, = 3 if severe CLD
CLD by MVR function	= CLD function if operation is MVR, = 0 otherwise
CLD by MVRepair function	= CLD function if operation is MVRepair, = 0 otherwise
Creatinine function 1	= max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise
Creatinine by MVR function	= Creatinine function 1 if valve operation is MVR, = 0 otherwise
Creatinine by MVRepair function	= Creatinine function 1 if valve operation is MVRepair, = 0 otherwise
CVD without prior CVA	= 1 if patient has history of CVD and no prior CVA, = 0 otherwise
CVD and prior CVA	= 1 if patient has history of CVD and a prior CVA, = 0 otherwise
Diabetes, noninsulin	= 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin	= 1 if patient has diabetes treated with insulin, = 0 otherwise
Dialysis	= 1 if patient requires dialysis preoperatively, = 0 otherwise
Dialysis by MVR function	= 1 if patient has history of dialysis and operation is MVR, = 0 otherwise
Dialysis by MVRepair function	= 1 if patient has history of dialysis and operation is MVRepair, = 0 otherwise
Ejection fraction function	= max (50 - ejection fraction, 0)
EF by MVR function	= Ejection fraction function if valve operation is MVR, = 0 otherwise
EF by MVRepair function	= Ejection fraction function if valve operation is MVRepair, = 0 otherwise
Endocarditis, active	= 1 if patient has active endocarditis, = 0 otherwise
Endocarditis by MVR function	= 1 if patient has active endocarditis and valve operation is MVR, = 0 otherwise
Endocarditis by MVRepair function	= 1 if patient has active endocarditis and valve operation is MVRepair, = 0 otherwise
Female	= 1 if patient is female, = 0 otherwise
Female by MVR function	= 1 if female and operation is MVR, = 0 otherwise
Female by MVRepair function	= 1 if female and operation is MVRepair, = 0 otherwise
Female by BSA function 1	= BSA function 1 if female, = 0 otherwise
Female by BSA function 2	= BSA function 2 if female, = 0 otherwise
Hypertension	= 1 if patient has hypertension, = 0 otherwise
IABP or inotropes	= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
IABP by MVR function	= 1 if patient requires preop IABP or inotropes and operation is MVR, = 0 otherwise
IABP by MVRepair function	= 1 if patient requires preop IABP or inotropes and operation is MVRepair, = 0 otherwise
Immunosuppressive treatment	= 1 if patient has received immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency, mitral	= 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency, tricuspid	= 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease	= 1 if patient has left main disease, = 0 otherwise
MI 1-21 days	= 1 if history of MI 1 to 21 days prior to surgery, = 0 otherwise
MI ≤ 21 days <sup>a</sup>	= 1 if patient has history of MI within 21 days prior to surgery, = 0 otherwise (for CVA and PLOS; coded as < 24 hours and 1-21 days for others)
MI < 24 hours	= 1 if history of MI < 24 hours prior to surgery, = 0 otherwise
MVR	= 1 if valve operation is mitral valve replacement, = 0 otherwise
MVRepair	= 1 if valve operation is mitral valve repair, = 0 otherwise
No. diseased coronary vessel function	= 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise

Peripheral vascular disease	= 1 if patient has peripheral vascular disease, = 0 otherwise
Race black	= 1 if patient is black, = 0 otherwise
Race Hispanic	= 1 if patient is nonblack Hispanic, = 0 otherwise
Reop, 1 previous operation	= 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, $\geq 2$ previous operations	= 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Reop by MVR function	= 1 if surgery is a reoperation and operation is MVR, = 0 otherwise
Reop by MVRepair function	= 1 if surgery is a reoperation and operation is MVRepair, = 0 otherwise
Shock	= 1 if patient was in shock at time of procedure, = 0 otherwise
Shock by MVR function	= 1 if shock and operation is MVR, = 0 otherwise
Shock by MVRepair function	= 1 if shock and operation is MVRepair, = 0 otherwise
Status urgent	= 1 if status is urgent, = 0 otherwise
Status emergent	= 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage	= 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Status by MVR function	= 1 if status is emergent or salvage and operation is MVR, = 0 otherwise
Status by MVRepair function	= 1 if status is emergent or salvage and operation is MVRepair, = 0 otherwise
Stenosis, mitral	= 1 if patient has mitral stenosis, = 0 otherwise
Unstable angina	= 1 if patient has unstable angina and no MI within 7 days of surgery, = 0 otherwise

---

## S.15. Detailed risk model specifications

### STS Risk Model Variables – STS Adult Cardiac Surgery Database Version 2.73

CABG	Operative Mortality	Stroke	Renal Failure	Prolonged Ventilation	Deep Stern Infx	Reop	Mortality/ Morbidity	Length of Stay>14	Length of Stay<6
<b>B. Demographics</b>									
Age (140)	X	X	X	X	X	X	X	X	X
Gender (150)	X	X	X	X	X	X	X	X	X
RaceBlack (300)		X	X	X	X	X	X	X	X
RaceAsian (310)		X	X	X	X	X	X	X	X
Ethnicity (350)		X	X	X	X	X	X	X	X
<b>D. Risk Factors</b>									
WeightKg (630)	X	X	X	X	X	X	X	X	X
HeightCm (640)	X	X	X	X	X	X	X	X	X
Diabetes (780)	X	X	X	X	X	X	X	X	X
DiabCtrl (790)	X	X	X	X	X	X	X	X	X
CreatLst (750)	X	X	X	X	X	X	X	X	X
Dialysis (810)	X	X	NA	X	X	X	X	X	X
Hypertn (820)		X	X	X			X	X	X
InfEndTy (840)									
ChrLungD (860)	X		X	X	X	X	X	X	X
ImmSupp (970)	X		X	X	X	X	X	X	X
PVD (980)	X	X	X	X	X	X	X	X	X
CVD (1010)	X	X	X	X		X	X	X	X
CVA (1020)	X	X	X	X		X	X	X	X
<b>E. Previous Interventions</b>									
PrCAB (1215)	X	X	X	X	X	X	X	X	X
PrValve (1216)	X	X	X	X	X	X	X	X	X
POCPCIn (1520)	X		X	X		X	X	X	X
<b>F. Preoperative Cardiac Status</b>									
MIWhen (1550)	X	X	X	X			X	X	X
CHF (1580)	X		X	X	X	X	X	X	X
ClassNYH (1585)	X		X	X	X	X	X	X	X
CardPres (1610)	X		X	X					
CarShock (1620)	X	X	X	X		X	X	X	X
Resusc (1630)	X	X	X	X	X	X	X	X	X
ArrhyAfib (1700)	X	X	X	X		X	X	X	X
<b>G. Preoperative Medications</b>									
MedInotr (1790)	X		X	X		X	X	X	X
<b>H. Hemodynamics and Cath</b>									
NumDisV (1930)	X	X	X	X	X	X	X	X	X
LMainDis (1940)				X			X		
HDEF (1960)	X	X	X	X	X	X	X	X	X
VDStenA (2152)				X			X	X	X
VDStenM (2240)									
VDInsufA (2155)									X
VDInsufM (2270)	X			X		X	X	X	X
VDInsufT (2320)			X	X			X		X
<b>I. Operative</b>									
Incidence (2380)	X	X	X	X	X	X	X	X	X
Status (2390)	X	X	X	X	X	X	X	X	X
<b>L. Mechanical Cardiac Assist Devices</b>									
IABP-Timing (4620)	X		X	X		X	X	X	X

Valve (AVRepl, MV Repl, MVRepr)	Operative Mortality	Stroke	Renal Failure	Prolonged Ventilation	Deep Stern Infx	Reop	Mortality/ Morbidity	Length of Stay>14	Length of Stay<6
<b>B. Demographics</b>									
Age (140)	X	X	X	X	X	X	X	X	X
Gender (150)	X	X	X	X	X	X	X	X	X
RaceBlack (300)		X	X	X		X	X	X	X
RaceAsian (310)									
Ethnicity (350)		X	X	X		X	X	X	X
<b>D. Risk Factors</b>									
WeightKg (630)	X	X	X	X	X	X	X	X	X
HeightCm (640)	X	X	X	X	X	X	X	X	X
Diabetes (780)	X		X	X	X	X	X	X	X
DiabCtrl (790)	X		X	X	X	X	X	X	X
CreatLst (750)	X	X	X	X		X	X	X	X
Dialysis (810)	X	X	NA	X	X	X	X	X	X
Hypertn (820)	X	X	X	X			X		X
InfEndTy (840)	X	X	X	X		X	X	X	X
ChrLungD (860)	X		X	X	X	X	X	X	X
ImmSupp (970)	X		X				X	X	
PVD (980)	X	X				X	X	X	X
CVD (1010)		X	X	X		X	X	X	X
CVA (1020)		X	X	X		X	X	X	X
<b>E. Previous Interventions</b>									
PrCAB (1215)	X	X	X	X	X	X	X	X	X
PrValve (1216)	X	X	X	X	X	X	X	X	X
POCPClin (1520)									
<b>F. Preoperative Cardiac Status</b>									
MIWhen (1550)	X			X		X	X	X	X
CHF (1580)	X		X	X		X	X	X	X
ClassNYH (1585)	X		X	X		X	X	X	X
CardPres (1610)	X								
CarShock (1620)	X	X		X		X	X	X	
Resusc (1630)	X	X	X	X		X	X	X	X
ArrhyAfib (1700)	X	X		X		X	X	X	X
<b>G. Preoperative Medications</b>									
MedInotr (1790)	X		X	X	X	X	X	X	X
<b>H. Hemodynamics and Cath</b>									
NumDisV (1930)		X		X			X	X	X
LMainDis (1940)	X		X		X				
HDEF (1960)	X		X	X	X	X	X	X	X
VDStenA (2152)				X		X	X	X	X
VDStenM (2240)	X								
VDInsufA (2155)									
VDInsufM (2270)		X							
VDInsufT (2320)			X	X		X	X	X	X
<b>I. Operative</b>									
Incidenc (2380)	X	X	X	X	X	X	X	X	X
Status (2390)	X	X	X	X	X	X	X	X	X
<b>L. Mechanical Cardiac Assist Devices</b>									
IABP-Timing (4620)	X		X	X	X	X	X	X	X



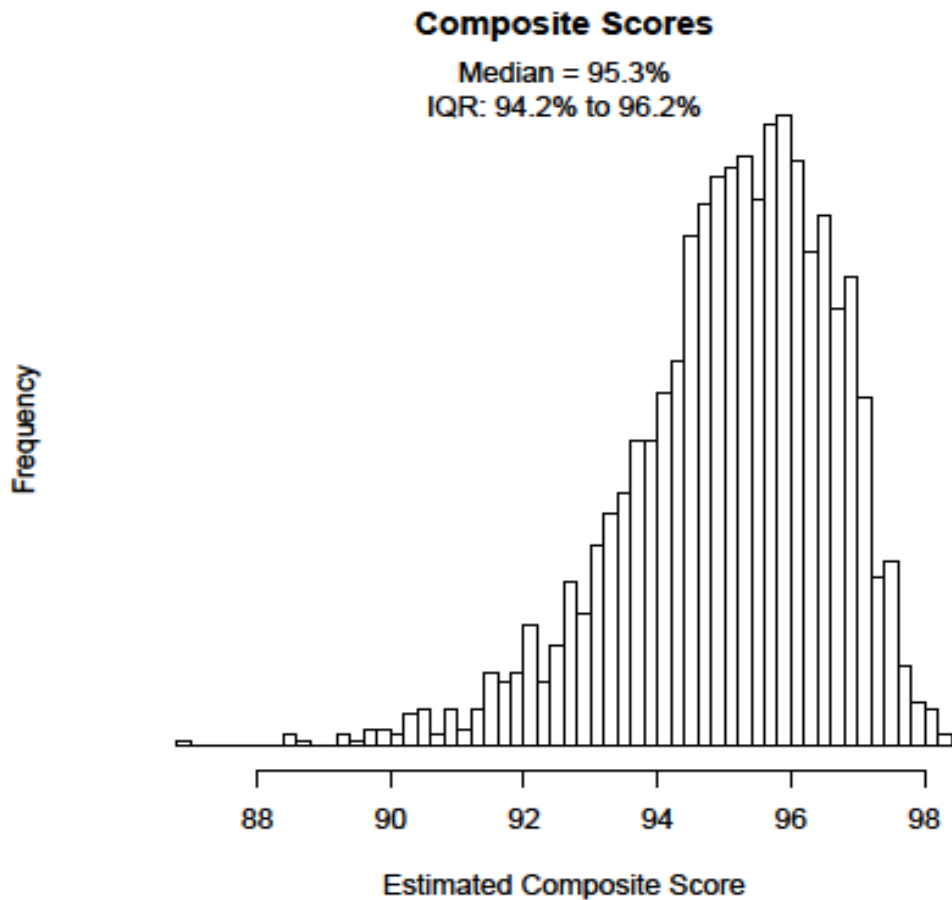
Valve+CABG (AVRepl+CABG, MVRepl+CABG, MVRepl+CABG)	Operative Mortality	Stroke	Renal Failure	Prolonged Ventilation	Deep Stern Infx	Reop	Mortality/Morbidity	Length of Stay>14	Length of Stay<6
<b>B. Demographics</b>									
Age (140)	X	X	X	X	X	X	X	X	X
Gender (150)	X	X	X	X	X	X	X	X	X
RaceBlack (300)			X	X		X	X	X	X
RaceAsian (310)									
Ethnicity (350)			X	X		X	X	X	X
<b>D. Risk Factors</b>									
WeightKg (630)	X	X	X	X	X	X	X	X	X
HeightCm (640)	X	X	X	X	X	X	X	X	X
Diabetes (780)	X	X	X	X	X		X	X	X
DiabCtrl (790)	X	X	X	X	X		X	X	X
CreatLst (750)	X	X	X	X		X	X	X	X
Dialysis (810)	X	X	NA	X	X	X	X	X	X
Hypertn (820)		X	X	X	X		X	X	X
InfEndTy (840)	X	X	X	X		X	X	X	X
ChrLungD (860)	X		X	X	X	X	X	X	X
ImmSupp (970)	X		X	X		X	X	X	X
PVD (980)	X	X	X	X		X	X	X	
CVD (1010)	X	X	X	X	X	X	X	X	X
CVA (1020)	X	X	X	X	X	X	X	X	X
<b>E. Previous Interventions</b>									
PrCAB (1215)	X	X	X	X	X	X	X	X	X
PrValve (1216)	X	X	X	X	X	X	X	X	X
POCPClin (1520)									
<b>F. Preoperative Cardiac Status</b>									
MIWhen (1550)	X	X	X	X		X	X	X	
CHF (1580)	X	X	X	X		X	X	X	X
ClassNYH (1585)	X	X	X	X		X	X	X	X
CardPres (1610)	X	X	X	X					
CarShock (1620)	X	X	X	X		X	X	X	
Resusc (1630)	X	X	X	X		X	X	X	X
ArrhyAfib (1700)	X	X	X	X		X	X	X	X
<b>G. Preoperative Medications</b>									
MedInotr (1790)	X		X	X		X	X	X	X
<b>H. Hemodynamics and Cath</b>									
NumDisV (1930)	X	X	X	X	X	X	X	X	X
LMainDis (1940)	X			X					
HDEF (1960)	X		X	X		X	X	X	X
VDStenA (2152)									
VDStenM (2240)	X							X	
VDInsufA (2155)									
VDInsufM (2270)							X		
VDInsufT (2320)	X		X	X			X		X
<b>I. Operative</b>									
Incident (2380)	X	X	X	X	X	X	X	X	X
Status (2390)	X	X	X	X	X	X	X	X	X
<b>L. Mechanical Cardiac Assist Devices</b>									
IABP-Timing (4620)	X		X	X		X	X	X	X

Please see section 2b4.4 of the measure testing attachment for additional information.

**1b.2.** Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. (This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.

The measure was calculated using STS data for patients undergoing cardiac surgery during January 2017 - December 2019. Five major procedures were included: isolated CABG, isolated AVR, AVR + CABG, isolated mitral valve repair or replacement procedures, and mitral valve repair or replacement + CABG procedures. Initially, 2919 surgeons were identified using their National Provider Identifiers. Surgeons without a National Provider Identifier (e.g., from a foreign country) or with invalid National Provider Identifiers were excluded (49). Surgeons were also required to have reported at least one of any type of cardiac procedure during each of the three 12-month periods (i.e., January-December 2017, 2018 and 2019). This was to ensure that the included surgeons had not just finished training or, conversely, had retired, and that they had actively participated in the most recent STS harvest. This requirement excluded 495 surgeons. From the remaining 2,375 surgeons, we included 2098 surgeons who met the annual completeness threshold of 98% of the operative mortality fields to assure accuracy of the operative mortality endpoint and had performed at least 10 major procedures during the 3-year period, both to facilitate statistical computations and because results would not be calculated or reported for surgeons with lower volumes than this. In the table below, we provide the number of measured entities (# surgeons), the number of eligible patient records (# operations), and the distribution of composite score estimates by percentiles. Surgeons with at least 10 eligible records during the study period were included in the hierarchical model for estimating composite scores. While surgeons with 10 eligible cases are included in the hierarchical model procedure, composite scores will typically only be reported by the STS for surgeons with at least 100 cases during a 3-year time period. Thus, we tabulate results for all eligible surgeons and the subset with at least 100 eligible cases. The distribution of scores across surgeons is also summarized in a histogram below.

<b>Stat</b>	<b>Surgeons with ≥10 Eligible Cases</b>	<b>Surgeons with ≥100 Eligible Cases</b>
# Participant	2098	1841
# Operations	600207	584571
Mean	0.951	0.952
STD	0.01547	0.01508
IQR	0.0198	0.0193
0%	0.869	0.886
10%	0.931	0.932
20%	0.939	0.940
30%	0.945	0.946
40%	0.949	0.950
50%	0.953	0.954
60%	0.957	0.958
70%	0.960	0.961
80%	0.964	0.965
90%	0.969	0.969
100%	0.984	0.984



**1b.4.** Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. (This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.

This composite measure gauges the performance of STS surgeons and is not a patient or operation level measure. We do not have a simple way to generate data stratified by patient characteristics at the composite level.

### 3030 - STS Individual Surgeon Composite Measure for Adult Cardiac Surgery Improvement - 4b1.

Performance data for the Individual Surgeon Composite (3030) were first distributed to consenting surgeons in January 2020; overall performance trends for this measure are therefore not yet available. As noted elsewhere in these submission materials, measure 3030 aggregates individual surgeon performance on five surgical procedures (isolated CABG, isolated AVR, AVR+CABG, MVRR, MVRR+CABG) and provides each surgeon with mortality and morbidity domain scores and an overall composite score and star rating, based on their own case mix. Therefore, in the absence of multi-year performance trends for measure 3030, we are providing (below) the star rating trends for the five procedures aggregated within it.

The data demonstrate that the general trend since the introduction of each measure has been a decrease in the percentage of surgical programs with 1-star and 3-star ratings and a corresponding increase in 2-star programs. This trend is consistent with the performance improvement goals of the STS star rating program, which seek to reduce variation in performance and to drive all participants in the STS Adult Cardiac Surgery Database toward the 2-star (or "as expected") category.

	Stars	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
CABG	*	3.77	4.37	4.55	5.29	5.82	4.59	9.19	9	9.6	11
	**	88.57	88.27	89.21	84.65	84.4	86.64	75.86	76	76.5	75.5
	***	7.66	7.36	6.24	10	9.74	8.77	14.95	15	14	13.5
AVR	*	1.67	1.96	2.62	2.17	3.11	4.22	3.35	3	3.5	N/A
	**	92.26	92.84	92.70	90.3	88.75	87.89	88.98	91	90.6	N/A
	***	6.07	5.20	4.68	7.53	8.15	7.89	7.67	6	5.9	N/A
AVR + CABG	*	1.84	2.16	2.73	2.06	2.49	2.51	3.14	N/A	N/A	N/A
	**	93.5	93.03	92.76	92.26	90.72	90.42	90.7	N/A	N/A	N/A
	***	4.66	4.81	4.51	5.68	6.79	7.07	6.17	N/A	N/A	N/A
MVRR	*	1.85	2.41	3.64	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	**	91.81	87.06	85.65	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	***	6.34	10.53	10.71	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MVRR + CABG	*	2.55	2.08	2.74	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	**	88.0	89.97	91.78	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	***	9.45	7.96	5.48	N/A	N/A	N/A	N/A	N/A	N/A	N/A



# The Society of Thoracic Surgeons Composite Measure of Individual Surgeon Performance for Adult Cardiac Surgery: A Report of The Society of Thoracic Surgeons Quality Measurement Task Force

David M. Shahian, MD, Xia He, MS, Jeffrey P. Jacobs, MD, Paul A. Kurlansky, MD, Vinay Badhwar, MD, Joseph C. Cleveland, Jr, MD, Frank L. Fazzalari, MD, MBA, Giovanni Filardo, PhD, MPH, Sharon-Lise T. Normand, PhD, Anthony P. Furnary, MD, Mitchell J. Magee, MD, J. Scott Rankin, MD, Karl F. Welke, MD, Jane Han, MS, and Sean M. O'Brien, PhD

Department of Surgery and Center for Quality and Safety, Massachusetts General Hospital, and Harvard Medical School, Boston, Massachusetts (DMS); Duke Clinical Research Institute, Durham, North Carolina (XH, SMO); Johns Hopkins All Children's Heart Institute, St. Petersburg, Florida (JPJ); Columbia HeartSource, Columbia University College of Physicians and Surgeons, New York, New York (PAK); Department of Cardiothoracic Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania (VB); Division of Cardiothoracic Surgery, University of Colorado School of Medicine, Aurora, Colorado (JCC); Cardiac Surgery Department, University of Michigan Health System, Ann Arbor, Michigan (FLF); Institute for Health Care Research and Improvement, Baylor Health Care System, Dallas, Texas (GF); Department of Health Care Policy, Harvard Medical School and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (S-LTN); Starr-Wood Cardiac Group, Portland, Oregon (APF); HCA North Texas Division, Dallas, Texas (MJM); Cardiothoracic Surgery Associates, Nashville, Tennessee (JSR); Children's Hospital of Illinois and the University of Illinois College of Medicine, Peoria, Illinois (KFW); The Society of Thoracic Surgeons, Chicago, Illinois (JH)

**Background.** Previous composite performance measures of The Society of Thoracic Surgeons (STS) were estimated at the STS participant level, typically a hospital or group practice. The STS Quality Measurement Task Force has now developed a multiprocedural, multidimensional composite measure suitable for estimating the performance of individual surgeons.

**Methods.** The development sample from the STS National Database included 621,489 isolated coronary artery bypass grafting procedures, isolated aortic valve replacement, aortic valve replacement plus coronary artery bypass grafting, mitral, or mitral plus coronary artery bypass grafting procedures performed by 2,286 surgeons between July 1, 2011, and June 30, 2014. Each surgeon's composite score combined their aggregate risk-adjusted mortality and major morbidity rates (each weighted inversely by their standard deviations) and reflected the proportion of case types they performed. Model parameters were estimated in a Bayesian framework. Composite star ratings were examined using 90%, 95%, or 98%

Bayesian credible intervals. Measure reliability was estimated using various 3-year case thresholds.

**Results.** The final composite measure was defined as  $0.81 \times (1 \text{ minus risk-standardized mortality rate}) + 0.19 \times (1 \text{ minus risk-standardized complication rate})$ . Risk-adjusted mortality (median, 2.3%; interquartile range, 1.7% to 3.0%), morbidity (median, 13.7%; interquartile range, 10.8% to 17.1%), and composite scores (median, 95.4%; interquartile range, 94.4% to 96.3%) varied substantially across surgeons. Using 98% Bayesian credible intervals, there were 207 1-star (lower performance) surgeons (9.1%), 1,701 2-star (as-expected performance) surgeons (74.4%), and 378 3-star (higher performance) surgeons (16.5%). With an eligibility threshold of 100 cases over 3 years, measure reliability was 0.81.

**Conclusions.** The STS has developed a multiprocedural composite measure suitable for evaluating performance at the individual surgeon level.

(Ann Thorac Surg 2015;100:1315–25)

© 2015 by The Society of Thoracic Surgeons

---

For related article, see page 1141

---

Accepted for publication June 26, 2015.

Presented at the Fifty-first Annual Meeting of The Society of Thoracic Surgeons, San Diego, CA, Jan 24–28, 2015.

This document has been approved by the STS Executive Committee.

Address correspondence to Dr Shahian, Massachusetts General Hospital, 55 Fruit St, Bulfinch 284, Boston, MA 02114; e-mail: dshahian@partners.org.

The past quarter century has witnessed many changes in the measurement of cardiac surgical quality. As the relative proportion of coronary artery bypass grafting (CABG) operations has declined, detailed data regarding other types of procedures are now being collected, and corresponding risk models have been developed [1–3].

The [Appendix](http://dx.doi.org/10.1016/j.athoracsur.2015.06.122) can be viewed in the online version of this article [<http://dx.doi.org/10.1016/j.athoracsur.2015.06.122>] on <http://www.annalsthoracicsurgery.org>.

Because of progressively decreasing mortality rates and because survival is only one measure of the quality of care, The Society of Thoracic Surgeons (STS) Quality Measurement Task Force (QMTF) first developed a multidimensional composite performance measure for CABG in 2007 [4, 5] and subsequently for aortic valve replacement (AVR) [6] and AVR + CABG [7]. Additional composite measures for mitral repair or replacement, with or without concomitant CABG, will be released in 2015. The composite score for each of these five commonly performed adult cardiac surgical procedures is estimated at the STS participant (usually hospital or practice group) level. These composite measures effectively provide more end points while also affording a much more comprehensive perspective on quality of care, because they include both risk-adjusted mortality and morbidity (and for isolated CABG, two additional process measures).

In 2010 STS began voluntary public reporting of cardiac surgical outcomes for hospitals and surgical groups [8–10]. As of early 2015, approximately 44% of these database participants agreed to publish their STS composite measure results for CABG, AVR, or AVR + CABG on the *Consumer Reports* or STS Web sites.

This report describes the next major step in the 25-year evolution of cardiac surgery quality measurement. The STS QMTF has now developed a composite performance measure for individual adult cardiac surgeons, based on aggregate risk-adjusted morbidity and mortality for five common procedures. With the development of this composite measure, QMTF addresses a number of major concerns that have previously been raised regarding surgeon-level metrics. It combines results from five of the most frequently performed cardiac surgical procedures, encompassing most of a typical surgeon's practice, as opposed to basing performance on just one or a few separate procedures. Furthermore, it includes not just risk-adjusted mortality but also the risk-adjusted occurrence of any of five major complications, thus providing more comprehensive quality assessment and additional end points. This measure will be useful to surgeons in identifying potential areas for improvement, and it has numerous advantages compared with existing surgeon metrics if used for accountability purposes.

## Patients and Methods

### Study Sample

Data were obtained from the STS Adult Cardiac Surgery Database, version 2.73, for the 3-year period July 1, 2011, to June 30, 2014. Five major procedures were included: isolated CABG, isolated AVR, AVR + CABG, isolated mitral valve repair or replacement procedures, and mitral valve repair or replacement + CABG procedures.

A total of 2,757 surgeons in the STS Adult Cardiac Surgery Database were identified using their National Provider Identifiers. Surgeons without a National Provider Identifier (eg, from a foreign country) or with invalid National Provider Identifiers were excluded.

Surgeons were also required to have reported at least one of any type of cardiac procedure during each of the three 12-month periods (ie, July 2011 to June 2012, July 2012 to June 2013, and July 2013 to June 2014). This was to ensure that the included surgeons had not just finished training or, conversely, had retired, and that they had actively participated in the most recent STS harvest. This requirement excluded 428 surgeons. From the remaining 2,329 surgeons, we included only the 2,286 who had performed at least 10 major procedures during the 3-year period, both to facilitate statistical computations and because results would not be calculated or reported for surgeons with lower volumes than this.

### End Points

The STS surgeon composite is based on two outcomes domains, risk-adjusted operative mortality and risk-adjusted morbidity. Operative mortality was defined as death before hospital discharge or within 30 days of the operation. Consistent with previous STS composite measures, morbidity was defined as the occurrence of any one or more of the following major complications: prolonged ventilation, deep sternal infection, permanent stroke, renal failure, and reoperations for bleeding, coronary graft occlusion, prosthetic or native valve dysfunction, and other cardiac reasons, but not for other noncardiac reasons.

This composite measure does not include the additional two process domains (internal mammary artery use and perioperative medications) from the original STS CABG composite measure for hospitals and practice groups. This approach was necessary for computational reasons to efficiently combine the results from five procedures, most of which did not have comparable process measures available. It is also consistent with an evolving trend favoring the use of outcomes measures in assessing health care quality.

### Risk Adjustment

Surgeon-specific risk-adjusted operative mortality and major complication rates were estimated using a bivariate random-effects logistic regression model [11]. To adjust for case mix, each patient's risk score for operative mortality and his or her corresponding risk score for major complications were first calculated, using existing and modified STS risk models as described below. The goal of calculating a risk score was to reduce the number of covariates in the hierarchical model by summarizing the predictive information from a large number of baseline covariates into a single number. Adjustment for each covariate individually in the hierarchical model would be theoretically preferable but is impractical due to the large number of records and covariates and the computationally intensive nature of Bayesian hierarchical model estimation.

To study the consequences of using an overall risk score for each patient instead of individual covariates in our models, sensitivity analyses were performed in which each surgeon's risk-adjusted mortality and complication rates were estimated in models that adjusted for 41 and

47 individual patient covariates, respectively. These estimates were compared with those derived from models adjusting for a single composite risk score. To make this analysis computationally manageable, model variables were estimated by maximum likelihood (ie, empirical Bayes) instead of performing a fully Bayesian analysis. To further simplify this sensitivity analysis, mortality and complication rates were estimated in separate models, not simultaneously in a single model, and the cohort was restricted to isolated CABG. For each end point (operative mortality and major complications) we calculated each surgeon's risk-standardized rate of the end point using each model and compared the results.

After sensitivity analyses demonstrated the validity of this risk score approach (see Results), the operative mortality risk score (predicted risk of death) was then used as a covariate in the hierarchical model for operative mortality, and the major complication risk score (predicted risk of major morbidity) was used as a covariate in the hierarchical model for major complications. To reduce potential bias, the hierarchical model included both individual patient-level risk scores and the average value of these patient-level risk scores calculated separately for each surgeon (see the [Appendix \[11\]](#)) [12–14].

For patients undergoing isolated CABG, isolated AVR, or AVR + CABG, risk scores were calculated according to the published STS 2008 mortality and major complications models for isolated CABG, isolated valve, or valve + CABG [1–3]. To ensure high calibration for the current study cohort, coefficients of each model were reestimated using the current 3-year study sample and current end point definitions. Risk scores for patients undergoing a mitral operation without CABG were calculated using a modified version of the published STS 2008 mortality and major complications models for isolated valve procedures [2]. These modifications allowed inclusion of patients undergoing tricuspid repair, an increasingly common adjunct, urgent and emergency procedures, all arrhythmia ablation procedures for atrial fibrillation, atrial septal defect and patent foramen ovale closures, and active and treated endocarditis. Also included are more granular classifications and adjustment for the degree of tricuspid regurgitation (less than moderate, moderate, and severe).

Coefficients of the modified models were estimated using the current 3-year study cohort and end point definitions. Risk scores for patients undergoing a mitral operation with concomitant CABG were calculated using a similarly modified version of the published STS 2008 mortality and major complications models for valve + CABG operations [3], also with reestimated coefficients.

### Final Composite Score

The overall composite score was calculated as a weighted sum of (1 minus risk-adjusted mortality rate) and (1 minus risk-adjusted major morbidity rate). Mortality and morbidity rates were weighted inversely by their respective standard deviations across surgeons. This procedure is equivalent to first rescaling mortality and morbidity rates by their respective standard deviations

across surgeons and then assigning equal weighting to the rescaled mortality rate and rescaled morbidity rate. Standard deviations derived from the data were used to define the final composite measure as  $0.81 \times (1 \text{ minus risk-standardized mortality rate}) + 0.19 \times (1 \text{ minus risk-standardized complication rate})$ .

### Model Estimation

Model parameters were estimated in a Bayesian framework ([Appendix \[11\]](#)). Posterior means and credible intervals (CrIs) were calculated using Markov chain Monte Carlo (MCMC) simulations, as implemented in OpenBUGS 3.2.2 software. Posterior summaries were calculated by generating 20,000 sets of simulated parameter values after a burn-in period to ensure convergence.

### Bayesian CrIs

A variety of Bayesian CrIs were determined from the simulated values. For example, 98% Bayesian CrIs were obtained by calculating the 200th lowest and 200th highest values of a surgeon's composite score across the 20,000 simulated values. These CrIs were examined to select the one best suited to accurately identify a reasonable percentage of above-average and below-average performing surgeons.

### Reliability

Model reliability was estimated as described in the [Appendix \[11\]](#).

### Star Ratings

In addition to calculating composite score point estimates with CrIs, star ratings were also assigned to all surgeons. These star ratings mirror the approach used in STS hospital-level and program-level composite scores. They facilitate better understanding of surgeon performance by patients and families, who are often unable to accurately interpret numeric scores and confidence intervals [15].

## Results

In this study cohort, the five procedures included in the STS surgeon composite measure encompass, on average, 78% of a surgeon's total cardiac surgical procedure volume (median, 83.2%; range, 6.4% to 100%). For approximately 5% of these surgeons, these index procedures represented less than 50% of their total procedural volume.

In sensitivity analyses, the Pearson correlation coefficients between surgeon performance estimates adjusted for individual patient covariates vs patient risk scores were 0.9999 for risk-adjusted operative mortality and 0.9988 for risk-adjusted major complications. This provides assurance that the use of an overall patient risk score in the hierarchical models does not substantially affect estimates of the surgeon composite score.

Overall results for all procedures ( $n = 621,489$ ) and all surgeons in the study sample are provided in [Table 1](#), and the summary results by procedure type are reported in [Table 2](#). Substantial variation exists in the numbers of

Table 1. Overall Results for All Procedures and Surgeons in the Study Cohort—July 1, 2011, to June 30, 2014

Variable	No. (%)
Surgeons	2,286
Operations	621,489
Mortality	14,463 (2.3)
Any major morbidity	85,217 (13.7)
Prolonged ventilation	63,658 (10.2)
Deep sternal infection	1,706 (0.3)
Permanent stroke	8,900 (1.4)
Renal failure	15,405 (2.5)
Reoperations	18,598 (3.0)

specific procedures and mortality and morbidity rates by procedure; including all of them in a methodologically appropriate manner should provide a more complete and balanced perspective than focusing on single procedures.

Figure 1 depicts the Bayesian-estimated distributions of risk-adjusted mortality (median, 2.3%; interquartile range, 1.7% to 3.0%) and morbidity (median, 13.7%; interquartile range, 10.8% to 17.1%). The distributions show substantial variation across surgeons, confirming that these end points are useful for discriminating performance. Figure 2 shows the distribution of calculated composite scores (point estimates) based on 3 years of data.

Table 3 reports the star rating for the overall composite and its mortality and morbidity components using three different Bayesian CrIs. As these intervals increase from 90% to 98%, the total number of surgeons classified as above or below average decreases, but the Bayesian probability that the surgeon is truly a high or low performer increases (ie, the likelihood of being misclassified decreases). A 98% CrI was selected, which produces approximately the same total proportion (~25%) of low and high performers as the original STS isolated CABG composite estimated at the hospital or practice group level.

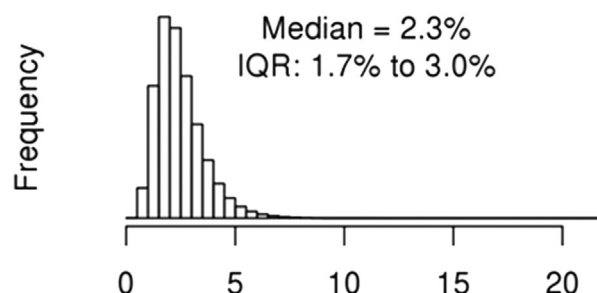
The 98% Bayesian CrI shows the range in which the surgeon's true risk-adjusted score is likely to lie. If the

Table 2. Characteristics by Procedure Group

Procedure Group	Total	Mortality		Morbidity	
	No. (%)	No.	Rate, %	No.	Rate, %
Isolated CABG	417,261 (67.1)	8,295	2.0	51,281	12.3
Isolated AVR	84,751 (13.6)	2,059	2.4	11,458	13.5
Isolated MVR	14,948 (2.4)	539	3.6	2,905	19.4
AVR + CABG	53,081 (8.5)	2,124	4.0	10,801	20.3
MVR + CABG	6,547 (1.1)	474	7.2	2,125	32.5
Isolated MV repair	30,347 (4.9)	339	1.1	2,953	9.7
MV repair + CABG	14,554 (2.3)	635	4.4	3,694	25.4
Total	621,489 (100.0)	14,465	2.3	85,217	13.7

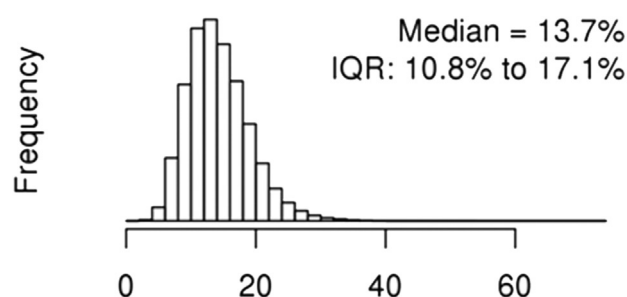
AVR = aortic valve replacement; CABG = coronary artery bypass grafting; MV = mitral valve; MVR = mitral valve replacement.

## Mortality



Participant-Specific Adjusted Rate (%)

## Morbidity



Participant-Specific Adjusted Rate (%)

Fig 1. Bayesian-estimated distribution of risk-adjusted (top) mortality and (bottom) morbidity rates. (IQR = interquartile range.)

lower limit of the 98% Bayesian CrI exceeds the STS average value, then it is at least 99% probable (98% CrI plus the 1% upper tail) that the surgeon's true performance is better than the STS average value and that surgeon is assigned a 3-star rating. Conversely, if the upper limit of the 98% Bayesian CrI is less than the STS average value, then it is at least 99% probable (98% CrI plus 1% lower tail) that the surgeon's true performance is worse than the STS average value; that surgeon is assigned a 1-star rating. Surgeons not meeting either of these criteria were assigned 2-star ratings (average, or as expected).

Table 4 provides the observed and risk-adjusted morbidity and mortality rates for each star rating category and Bayesian CrI studied. In each case, a monotonic decrease occurs in the observed or risk-adjusted mortality and morbidity rates as the star ratings increase, which provides internal validation that the overall composite scores and star ratings accurately reflect the constructs they were designed to measure.



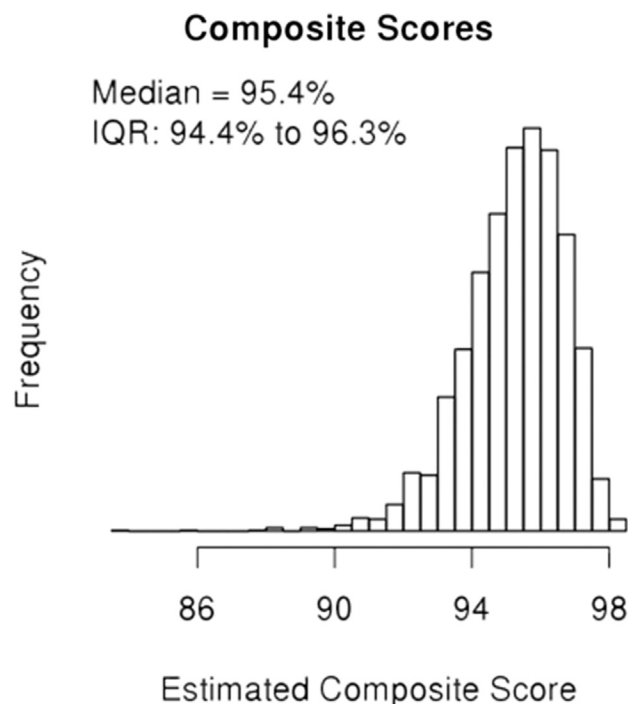


Fig 2. Distribution of calculated composite scores (point estimates) based on 3 years of data. (IQR = interquartile range.)

Table 5 summarizes the estimated reliability (signal-to-noise ratio) of the STS surgeon composite measure at various thresholds of total index procedure volume over 3 years. These data address the question of whether there is a lower limit of total index procedure volume, and associated composite measure reliability, that should make a surgeon ineligible to receive a composite score and star rating. Overall reliability of the surgeon composite is quite high, approximately 80%, at all volume thresholds, although the reliability for a specific surgeon with very low volumes may still be problematic. For context, the reliability of the most recent isolated CABG

Table 3. Star Ratings for Various Bayesian Credible Intervals

Variable	1 Star No. (%)	2 Stars No. (%)	3 Stars No. (%)
Composite			
90% CrI	362 (15.8)	1,351 (59.1)	573 (25.1)
95% CrI	293 (12.8)	1,521 (66.5)	472 (20.6)
98% CrI	207 (9.1)	1,701 (74.4)	378 (16.5)
Mortality			
90% CrI	113 (4.9)	1,942 (84.9)	231 (10.1)
95% CrI	70 (3.1)	2,079 (90.9)	137 (6.0)
98% CrI	37 (1.6)	2,179 (95.3)	70 (3.1)
Morbidity			
90% CrI	411 (18.0)	1,264 (55.3)	611 (26.7)
95% CrI	343 (15)	1,451 (63.5)	492 (21.5)
98% CrI	262 (11.5)	1,623 (71)	401 (17.5)

CrI = credible interval.

Table 4. Observed and Adjusted Event Rates by Star Rating (All 2,286 Surgeons Included)

Star Rating	Mortality, %			Morbidity, %		
	1 Star	2 Stars	3 Stars	1 Star	2 Stars	3 Stars
Observed						
90% CrI	4.1	2.4	1.4	21.6	14.3	9.5
95% CrI	4.2	2.4	1.4	22.2	14.2	9.2
98% CrI	4.2	2.4	1.3	23.1	14.2	9.0
Risk-adjusted						
90% CrI	4.1	2.5	1.3	21.1	14.3	9.3
95% CrI	4.1	2.5	1.3	21.7	14.3	9.0
98% CrI	4.2	2.5	1.2	22.6	14.2	8.8

CrI = credible interval.

composite measure for hospitals, based on 1 year of data, was 0.77 (0.74, 0.80), and reliability of the most recent isolated AVR composite measure for hospitals, based on 3 years of data, was 0.52 (0.47, 0.57).

Reliability increases slightly at higher volume thresholds, but this comes at the cost of progressively fewer surgeons who are eligible to receive a score. There is no one correct answer for the optimal threshold, but 100 total index cases over 3 years seems a reasonable compromise that retains a high proportion of surgeons and patients but excludes very low-volume surgeons from receiving a score.

Finally, Figure 3 shows composite scores and star ratings plotted against total aggregate volume of the five index procedures over 3 years. Median volumes were highest for 3-star surgeons, although some low-volume surgeons achieved a 3-star rating. There were no 1-star high-volume surgeons.

## Comment

### What Is the Appropriate Level of Attribution for Cardiac Surgical Performance Measurement?

Although several states (eg, New York [16–19], Pennsylvania, Massachusetts) have implemented individual surgeon report cards for cardiac operations, a number of considerations have historically led the STS to focus on hospitals or practice groups. Methodologic concerns have included small sample size and low measure reliability. In addition, from a health policy perspective, many argue that attributing the health care outcome of a patient to 1 individual, the responsible physician or surgeon, is inconsistent with the increasingly team-based nature of modern health care, especially for more complex diagnoses and procedures such as those in cardiac surgery. Cardiac surgical outcomes are influenced by many factors, including the appropriate referral and optimal preoperative stabilization of patients by cardiologists, cardiac anesthesia management, cardiopulmonary perfusion, operating room scrub technicians and nurses, resident or staff assistant surgeons, physician assistants and nurse practitioners, postoperative intensive care unit management,



Table 5. Composite Measure Reliability at Various Eligibility Thresholds of Surgeon Volume for Index Procedures (3-Years)<sup>a</sup>

Threshold Number of Index Cases Over 3 Years	Surgeons Included (No.)	Patients Included (No.)	Reliability $\hat{\rho}^2$ (95% PrI)
10	2,286	621,489	0.77 (0.75, 0.79)
25	2,234	620,586	0.78 (0.76, 0.80)
36	2,205	619,691	0.79 (0.77, 0.80)
50	2,165	617,976	0.79 (0.78, 0.81)
100	1,976	603,594	0.81 (0.79, 0.82)
150	1,737	573,491	0.81 (0.80, 0.83)
200	1,432	520,724	0.82 (0.81, 0.84)

<sup>a</sup> Number of surgeons and patients included for each threshold.

$\hat{\rho}^2$  is the estimated squared correlation between the set of surgeon-specific estimates of composite performance measure values and their corresponding unknown true values (used as the measure of reliability in this study).

PrI = probability interval.

hospital support services, postdischarge extended care facilities, and rehabilitation. Having cardiac surgeons and their teams working together consistently in the operating

room seems to be particularly important to achieving optimal results [20].

Notwithstanding the complex, team-based nature of cardiac surgical care, surgeons still have primary overall responsibility for their patients, and they directly affect patient outcomes from many perspectives, including diagnosis, operative planning, technical conduct of the procedure, supervision of perioperative management, and follow-up. If other components of the team or hospital are contributing to adverse patient outcomes (eg, anesthetic technique leading to prolonged ventilation), surgeons should be leaders in addressing such problems. Accurate characterization of a surgeon's performance based on a broad sample of common procedures is the first step in identifying opportunities for improvement.

In addition, the federal government, some states, and commercial payers are increasingly requesting data on individual physician and surgeon performance for various accountability and reimbursement programs, and this trend will continue to accelerate. In the absence of well-designed gold standard metrics, examples abound of health care performance measures based on suboptimal data sources (eg, claims or billing data) and opaque, proprietary analytic methodologies, sometimes leading

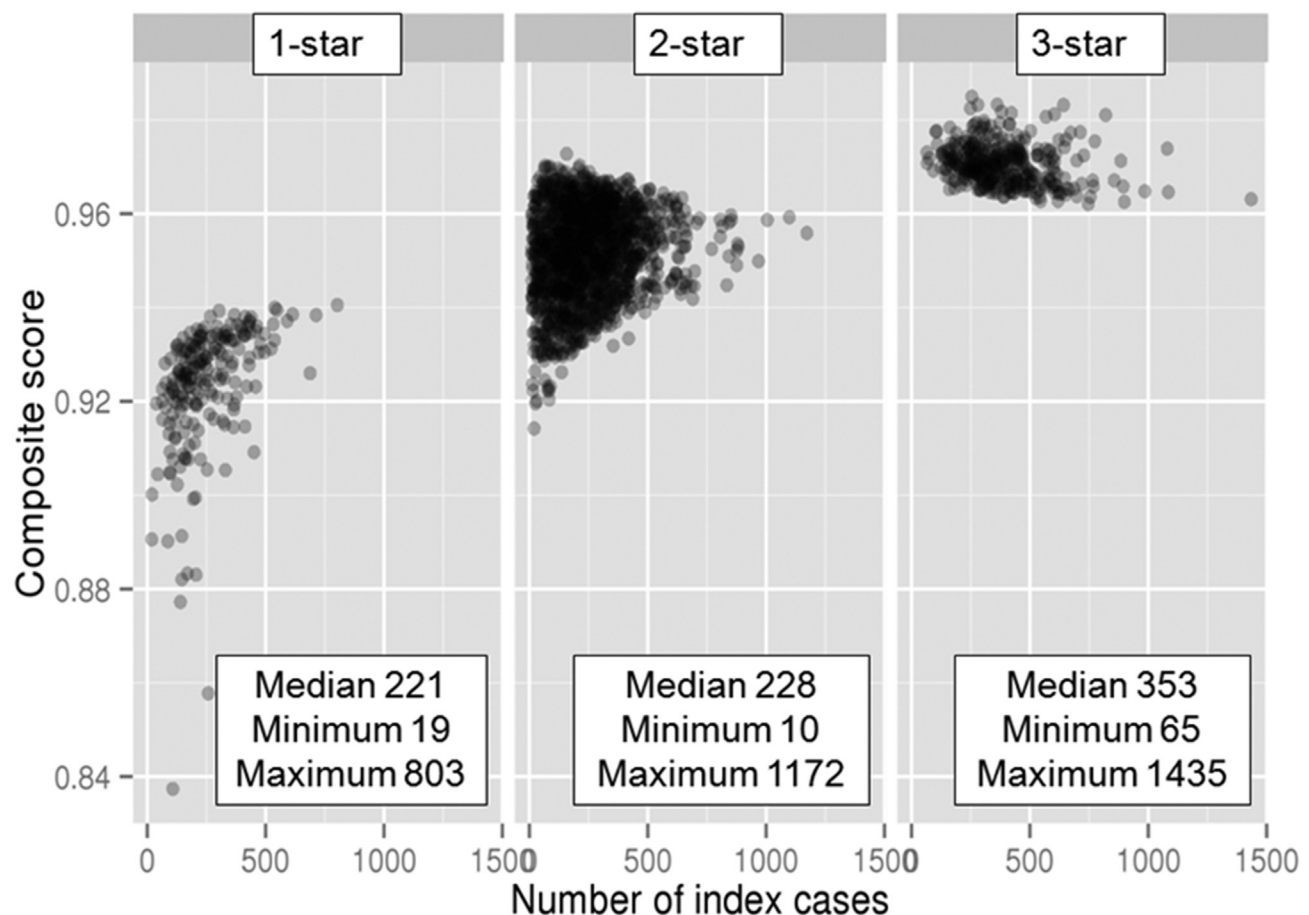


Fig 3. Composite scores and (left) 1-star star, (middle) 2-star, and (right) 3-star ratings plotted against total aggregate volume of the five index procedures over 3 years.

to completely discrepant results for the same provider [21–23]. Because physician-level performance data are increasingly requested by various governmental, commercial, and regulatory entities, it is essential that they be derived from the best data and analytics and that they are nationally benchmarked.

Based on 7 years of experience producing nationally endorsed performance measures at the hospital or group level, and nearly 5 years of experience with public reporting, the STS QMTF believes it is now better positioned to address the methodologic issues that had previously discouraged the development of surgeon-level measures.

### *Multiprocedural, Multidimensional, Performance Measurement*

Assessing the quality of cardiac surgeons based on their isolated CABG mortalities provides insight into only one outcome, albeit the most important one, and one procedure. Regarding the latter, the number and proportion of isolated CABG procedures has been steadily declining for the past decade due to improved medical therapy and advances in interventional cardiology (eg, drug-eluting stents). Consequently, isolated CABG is a much less dominant component of most cardiac surgeon's practices than it was a decade ago. Also, compared with hospital cardiac surgical units or practice groups, greater variability exists among individual surgeons in the types and relative frequencies of procedures they perform. Some may do few isolated CABG procedures but rather focus on mitral or aortic valve procedures, and their quality ratings should reflect this. Several state report cards (eg, New York and Pennsylvania) now include separate ratings for isolated CABG, isolated valve, and valve plus CABG. However, in contrast to the current STS surgeon composite, many of these measures do not use separate risk models for the various valve procedures (eg, aortic vs mitral), and the sample sizes for these various procedures remain small because they are analyzed separately.

Our goal in the development of the STS surgeon composite was to combine, in a scientifically rigorous manner, risk-adjusted mortality and morbidity domains across a broad portfolio of procedures. The final composite score reflects the proportion of each type of case performed by each specific surgeon. Because results are combined across five procedures, this approach also provides larger sample sizes and more end points to analyze, which helps to address reliability concerns (see below). Finally, procedure-specific STS risk models are used, which are generally preferred compared with generic models.

### *Sample Size and Reliability*

Measurement of health care performance requires attention to many details, including data source and quality, treatment of missing data, selection of appropriate measures and target populations, time frame for collection of patient and provider data, level of attribution, statistical methodology, and the criteria for classifying performance categories. Arguably, one of the most

vexing and nearly universal problems is sample size, and this is particularly problematic at the individual physician or surgeon level.

For most publicly reported condition-specific and procedure-specific outcomes, the number of observations available in a typical reporting period is small, rarely more than several hundred, and usually less for individual providers. The resulting performance estimates have an associated uncertainty, represented by frequentist confidence intervals or Bayesian CrIs. The smaller the event rate and sample sizes, the wider the corresponding confidence intervals around the point estimate and the less certainty there is regarding the true underlying value [24]. Because of substantial random variation or "noise," the reliability of performance measures is often low, ratings and rankings vary substantially over time [25], and differentiating provider performance and identifying high and low performers is difficult. In one study of major surgical procedures, Dimick and colleagues [26] demonstrated that only CABG was typically performed with sufficient volume to allow reliable detection of a doubling of the mortality rate, and these data were taken from an era when isolated CABG was performed much more frequently. Krell and colleagues [27] studied general and vascular surgical outcomes measures derived from data from the American College of Surgeons National Surgical Quality Improvement Program registry and found that most measures had low reliability to detect performance differences for common procedures. Even more challenging issues have been observed with common medical diagnoses [28, 29].

Compared with other surgical specialties, adult cardiac surgeons perform a relatively larger number of a smaller ensemble of major cases. Despite this, the numbers of procedures of a given type performed by most cardiac surgeons in a typical data analysis period remains small, and confidence intervals are correspondingly wide.

### *Addressing Sample Size Concerns*

A variety of approaches are available to deal with small sample sizes and low reliability at the individual cardiac surgeon level, and several of these were incorporated into the current STS measure. The first involves the time period over which surgeon data are collected and analyzed. Historically, this has been 1 year, but individual physician metrics are now increasingly collected over rolling 3-year intervals, as in the current measure. Longer time frames (eg, 5 years) would provide even greater sample sizes and reliability but at the cost of potentially lower relevance for some of the data. Other aspects of the structures or processes of care may have changed in the interim at an institution (eg, a new cardiac anesthesia or intensive care unit team), and older data may not reflect current practice.

Another strategy to increase surgeon sample sizes is to expand the number of end points being analyzed and to focus on end points having sufficient frequency. The mortality rate has historically been used as the primary or sole end point to measure performance for complex surgical procedures, but mortality rates in cardiac surgery

have declined substantially, making this approach less useful. Furthermore, despite its obvious importance, survival is only one aspect of quality. Equally important to many patients is the avoidance of serious complications, such as stroke, renal failure, reoperation, sternal infection, and prolonged ventilation, some of which have life-altering consequences (eg, neurologic disability, permanent dialysis). Assessing multiple postoperative outcomes, as in this surgeon composite measure, provides a more comprehensive perspective on quality as well as additional end points. Using a similar strategy, the original STS isolated CABG composite identified 23% of providers as outliers compared with 1% when using risk-adjusted mortality alone [4, 5].

Certain statistical techniques improve the accuracy of estimates from small samples and are incorporated into Bayesian approaches, as in this surgeon composite. These methods “shrink” the results of providers toward the population mean, with the degree of shrinkage being inversely proportional to the sample size. Less shrinkage occurs when large samples are available because the sample estimate is more likely a valid estimate of the underlying true performance. Conversely, with small sample sizes, there is more shrinkage of the observed results towards the population mean. Most statisticians [30–33] believe this approach provides more accurate estimates of true performance and avoids falsely labeling providers as high or low outliers based on small samples with large standard errors.

#### *The STS Surgeon Composite Measure*

By incorporating the features described above, the STS surgeon composite measure has been designed to address most methodologic concerns regarding physician-level metrics. Sample size issues are mitigated by measuring composite end points (death and any major morbidity) for five different procedures over a 3-year rolling time interval and by using Bayesian estimation. By including composite performance scores for a portfolio of five procedures that account for nearly 80% of a typical STS database participant surgeon's clinical activity, this metric provides a more balanced and comprehensive perspective than focusing on just one procedure, such as CABG, or one end point such as death.

Recognizing that surgeons' practices vary, each surgeon's composite performance is implicitly “weighted” by the proportion of each type of procedure he or she performs. For instance, the results of surgeons who primarily perform mitral procedures are affected most by their mitral surgery results. This approach is especially relevant for surgeons with highly specialized practices who may do relatively few isolated CABG procedures and whose performance would thus be difficult to assess using a CABG measure only.

Finally, performance on each of these procedures is estimated using risk models specific to those procedures, in most cases the exact or slightly modified versions of previously published models [11]. By incorporating these methodologic approaches, the STS surgeon composite measure achieves excellent statistical

reliability, higher than that of previously reported individual physician or surgeon measures and even many hospital-level measures.

For the nonstatistician, this surgeon composite measure can be thought of as simply an extension and modification of the methodology used in all previous STS composite measures, which were designed for hospital-level or program-level use. One difference in methodology is that this is an outcomes-only composite; that is, the two process domains from the original STS CABG composite are not included.

A second difference is that this is a multiprocedural composite measure. It aggregates the observed and expected mortality rates and the observed and expected morbidity rates across five common adult cardiac surgical procedures. Expected values are derived from existing or slightly modified STS risk models. Given this methodology, higher composite scores are achieved when mortality and morbidity rates across all these procedures are low, and lower composite scores result when morbidity and mortality rates are high.

#### *Appropriate Interpretation*

This surgeon composite, like most other risk-adjusted performance measures in health care, uses indirect rather than direct standardization. Indirectly standardized results can be used to compare an individual provider with the benchmark population of providers, asking the question, “how do his or her results compare with what would have been expected if their patients had been cared for by an average provider?” As opposed to direct standardization, indirectly standardized results should not be used to directly compare 2 surgeons [34, 35]. Each surgeon's performance is based only on the patients he or she actually treated. A surgeon who is able to achieve better-than-average results with a predominately low-risk population may not necessarily achieve similarly excellent results if confronted with a high-risk group of patients because there are few if any data regarding his or her experience with such patients.

Such case-mix differences may lead to seemingly paradoxical results. Even with perfect patient-level risk adjustment, if surgeons A and B have a substantially different case mix (e.g., surgeon A operates on a substantially higher or lower proportion of high-risk patients than surgeon B), surgeon A could have risk-standardized mortality ratios that were equivalent to those of surgeon B for each type of patient and operation (high-risk vs low-risk) yet have a higher or lower overall risk-standardized mortality ratio [36, 37].

#### *Limitations*

The STS surgeon composite measure is based on observational data. These are the only data commonly available for measuring health care performance because randomizing patients for routine, noninvestigational care is not possible. The STS data used in these analyses come from one of the largest and longest standing clinical data registries in the world, whose granular data elements are regularly updated. External audit of 10% of sites each



year has demonstrated very high agreement between submitted data and medical record audits (96% overall in 2014). However, despite rigorous risk adjustment using granular clinical data, the potential for unmeasured confounders still exists.

The statistical models used in this composite measure are quite complex because they involve not only multiple end points but also multiple procedures, and the relative mix of these procedures may vary from surgeon to surgeon. There is rarely only one way such a model could be constructed. Considering the practical constraints of the computational burden for these highly complex models, QMTF clinicians and statisticians have developed an excellent model that reliably estimates and discriminates overall surgeon performance. However, other approaches to model development are possible, and QMTF continues to explore new methodologies.

As with any composite metric, there is potential information loss when only a single summary measure is reported. For example, it is possible that a surgeon may have excellent outcomes for CABG but less desirable outcomes for valve operations, or vice versa. For both quality assessment and improvement activities, this composite score is a high-level summary that should be supplemented by drill down to specific procedures. Furthermore, the surgeon composite should be used in combination with the relevant hospital-level or group-level performance metric for the procedure of interest.

Finally, although the five procedures included in this metric represent the bulk of a typical surgeon's work, some surgeons with highly specialized practices (eg, thoracic aortic surgery) may not be able to receive a score or the score may not reflect the particular area in which they specialize. Development of separate composite metrics for such specialized practices and less commonly performed procedures may be a better approach than trying to incorporate them within the overall surgeon metric.

## Conclusions

The STS QMTF has developed a surgeon-level composite measure based on risk-adjusted morbidity and mortality for five common cardiac surgical procedures, using 3 years of data. This measure has high statistical reliability and provides a comprehensive overall assessment of surgeon performance.

## References

- Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. *Ann Thorac Surg* 2009;88:S2–22.
- O'Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac Surg* 2009;88:S23–42.
- Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3—valve plus coronary artery bypass grafting surgery. *Ann Thorac Surg* 2009;88:S43–62.
- Shahian DM, Edwards FH, Ferraris VA, et al. Quality measurement in adult cardiac surgery: part 1—conceptual framework and measure selection. *Ann Thorac Surg* 2007;83:S3–12.
- O'Brien SM, Shahian DM, Delong ER, et al. Quality measurement in adult cardiac surgery: part 2—statistical considerations in composite measure scoring and provider rating. *Ann Thorac Surg* 2007;83:S13–26.
- Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons isolated aortic valve replacement (AVR) composite score: a report of the STS Quality Measurement Task Force. *Ann Thorac Surg* 2012;94:2166–71.
- Shahian DM, He X, Jacobs JP, et al. The STS AVR + CABG composite score: a report of the STS Quality Measurement Task Force. *Ann Thorac Surg* 2014;97:1604–9.
- Shahian DM, Edwards FH, Jacobs JP, et al. Public reporting of cardiac surgery performance: part 1—history, rationale, consequences. *Ann Thorac Surg* 2011;92:S2–11.
- Shahian DM, Edwards FH, Jacobs JP, et al. Public reporting of cardiac surgery performance: part 2—implementation. *Ann Thorac Surg* 2011;92:S12–23.
- Ferris TG, Torchiana DF. Public release of clinical outcomes data—online CABG report cards. *N Engl J Med* 2010;363:1593–5.
- The Society of Thoracic Surgeons. Statistical methodology for STS risk models and performance measures. The STS surgeon composite – technical appendix (updated February 2015). Available at: <http://www.sts.org/quality-research-patient-safety/statistical-methodology-risk-models-and-measures>. Accessed August 1, 2015.
- Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998;54:638–45.
- Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *J Roy Stat Soc B* 2006;68:859–72.
- Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001;135:112–23.
- Hibbard JH, Peters E, Slovic P, Finucane ML, Tusler M. Making health care quality reports easier to use. *Jt Comm J Qual Improv* 2001;27:591–604.
- Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's Cardiac Surgery Reporting System: four years later. *Ann Thorac Surg* 1994;58:1852–7.
- Hannan EL, Wu C, Ryan TJ, Bennett E, Culliford AT, Gold JP, et al. Do hospitals and surgeons with higher coronary artery bypass graft surgery volumes still have lower risk-adjusted mortality rates? *Circulation* 2003;108:795–801.
- Hannan EL, Cozzens K, King SB 3rd, Walford G, Shah NR. The New York State cardiac registries: history, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes. *J Am Coll Cardiol* 2012;59:2309–16.
- Hannan EL, Siu AL, Kumar D, Kilburn H Jr, Chassin MR. The decline in coronary artery bypass graft surgery mortality in New York State. The role of surgeon volume. *JAMA* 1995;273:209–13.
- Carey JS, Parker JP, Brandeau C, Li Z. The "occasional open heart surgeon" revisited. *J Thorac Cardiovasc Surg* 2008;135:1254–60.
- Leonardi MJ, McGory ML, Ko CY. Publicly available hospital comparison web sites: determination of useful, valid, and appropriate information for comparing surgical quality. *Arch Surg* 2007;142:863–8.
- Rothberg MB, Morsi E, Benjamin EM, Pekow PS, Lindenauer PK. Choosing the best hospital: the limitations of public quality reporting. *Health Aff (Millwood)* 2008;27:1680–7.
- Healthcare Association of New York State. HANYS' Report on Report Cards. Understanding publicly reported hospital quality measures. Available at: [http://www.hanys.org/quality/data/report\\_cards/2013/docs/2013\\_hanys\\_report\\_card\\_book.pdf](http://www.hanys.org/quality/data/report_cards/2013/docs/2013_hanys_report_card_book.pdf). Accessed August 1, 2015.

24. Austin PC, Reeves MJ. Effect of provider volume on the accuracy of hospital report cards: a Monte Carlo study. *Circ Cardiovasc Qual Outcomes* 2014;7:299–305.
25. Siregar S, Groenwold RH, Jansen EK, Bots ML, van der Graaf Y, van Herwerden LA. Limitations of ranking lists based on cardiac surgery mortality rates. *Circ Cardiovasc Qual Outcomes* 2012;5:403–9.
26. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA* 2004;292:847–51.
27. Krell RW, Hozain A, Kao LS, Dimick JB. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg* 2014;149:467–74.
28. Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care* 1996;34:737–53.
29. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician “report cards” for assessing the costs and quality of care of a chronic disease. *JAMA* 1999;281:2098–105.
30. The COPSS-CMS White Paper Committee: Ash AS, Fienberg SE, Louis TA, Normand ST, Stuke TA, Utts J. Statistical issues in assessing hospital performance. Commissioned by the Committee of Presidents of Statistical Societies. Available at: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf>. Accessed September 18, 2013.
31. Normand SL, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc* 1997;92:803–14.
32. Normand SL, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci* 2007;22:206–26.
33. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Stat Soc A* 1996;159:385–443.
34. Shahian DM, Normand SL. Comparison of “risk-adjusted” hospital outcomes. *Circulation* 2008;117:1955–63.
35. Shahian DM, Normand SL. What is a performance outlier? *BMJ Qual Saf* 2015;24:95–9.
36. Manktelow BN, Evans TA, Draper ES. Differences in case-mix can influence the comparison of standardised mortality ratios even with optimal risk adjustment: an analysis of data from paediatric intensive care. *BMJ Qual Saf* 2014;23:782–8.
37. Pouw ME, Peelen LM, Lingsma HF, et al. Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization. *PLoS One* 2013;8:e59160.

## DISCUSSION

**DR JOSEPH E. BAVARIA** (Philadelphia, PA): I congratulate Dr Shahian and his colleagues on tackling a very complex and a very contentious topic, namely, individual surgeon performance metrics. As I practice in Pennsylvania, I have operated in this environment for about 20 years. There are many questions this report may raise, but I will simply address a few, and maybe let others talk about the data methods that may relate to surgeons at all levels within the STS.

So the first question is, I am a little worried about the morbidity index and its effect on star ratings and surgeons in those categories, especially—and we talked about this—prolonged ventilation. In my view, this is the least serious of these morbidities, but it is equally weighted and also represents the greatest percentage of morbidity by at least three times in your data. Moreover, it may actually be the one morbidity metric that is the “least controlled” by the surgeon as it is usually an intensive care unit (ICU) institutional and cultural issue. Is it possible for a surgeon to have good mortality statistics, good morbidity statistics, but a poor prolonged ventilation metric, and that puts him or her into a 2-star or lesser star rating? That is my first question.

The second question is simply related to the tension between composite scores and individual procedural scores. As you know, in Pennsylvania, there is no composite score; everybody is reported by individual procedures. It seems that patients having a procedure would like to know what that surgeon’s performance is for that particular procedure. I translate it to my office. They do not really care about anything else. If I am doing aortic valve replacement (AVR), they want to know what my AVR outcomes are, and they can get that from the Pennsylvania state database. They could not get that right now under your system. Do you have any insight on the great decision before us on whether we report as composite only or do we report on individual therapeutic procedures? Thanks.

**DR SHAHIAN:** Thank you very much, Dr Bavaria, for these excellent questions. Regarding the inclusion of prolonged ventilation as part of the any-or-none morbidity bundle, this is something that we have discussed numerous times since the development of our first composite measure. The question, as you have alluded to, is whether prolonged ventilation rises to the same level of importance as the other complications with which it is grouped, such as reoperation or renal failure. We recently had the opportunity to analyze the impact of various individual complications on mortality, and the results were very interesting.

If you study isolated coronary artery bypass grafting (CABG) patients from a recent cohort and focus on patients who had no complications whatsoever, the raw mortality rate was remarkably low at 0.7%, a tribute to how far we have come in cardiac surgery. If you add just the one complication of prolonged ventilation, the mortality rate increases by a factor of almost nine, to 6%. If you then analyze this in a multivariate environment, adjusting for patients that have also had other complications, such as stroke, renal failure, or reoperation, then the adjusted odds of mortality associated with prolonged ventilation is 5.

Prolonged ventilation seems to be capturing an important cause of death that is not simply the result of these other complications, although they can also increase ventilation time. Prolonged ventilation may reflect longer cardiopulmonary bypass times, intraoperative fluid uptake, “wet lungs,” or the type of anesthetic management. It clearly is an independent, very strong predictor of mortality, and we think it is quite justifiable to include it in our morbidity metric.

As to who controls ventilation protocols, we agree that this is often managed primarily by anesthesia, but given the implications of prolonged ventilation, surgeons have a responsibility to be involved in this aspect of care and to work collaboratively with their anesthesia colleagues.



With regard to the tension between composite and individual procedure measures, we agree with you completely. You need an overall estimate of quality, and then you would like the ability to drill down to the actual procedure a patient is having. Both are important. Because many states are looking for a single metric to characterize quality, we wanted to develop something would be a good overall measure of a typical surgeon's practice, together with the ability to drill down to the individual procedure level. However, the latter is challenging due to relatively small sample sizes for individual surgeons.

**DR ANDREW SEELY** (Ottawa, ON, Canada): I am always interested in how to provide this information back to surgeons so that they improve their own care, improve their own performance, and you elected to provide the star rating, the 3-, 2-, and

1-star rating, rather than just focus on the composite score itself. Could you tell us why you wanted to add that star rating as well? Do you think that that is going to have more impact, if you will, on surgeons improving their own care?

**DR SHAHIAN:** For surgeons who are more sophisticated in terms of analyzing data, I think that the actual composite score will probably be more useful. But we followed the lead of folks like Judith Hibbard at Oregon, who have looked at ways that consumers can better understand data. Consumers do not understand numbers and confidence intervals. They need something simple and straightforward, and these star ratings are meant to be an aid to consumer interpretation, much more for the patients than for the docs, who I think are much more used to dealing with the actual data.

# The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 1—Coronary Artery Bypass Grafting Surgery

David M. Shahian, MD,<sup>a</sup> Sean M. O'Brien, PhD,<sup>b</sup> Giovanni Filardo, PhD, MPH,<sup>c</sup> Victor A. Ferraris, MD,<sup>d</sup> Constance K. Haan, MD,<sup>e</sup> Jeffrey B. Rich, MD,<sup>f</sup> Sharon-Lise T. Normand, PhD,<sup>g</sup> Elizabeth R. DeLong, PhD,<sup>b</sup> Cynthia M. Shewan, PhD,<sup>h</sup> Rachel S. Dokholyan, MPH,<sup>b</sup> Eric D. Peterson, MD, MPH,<sup>b</sup> Fred H. Edwards, MD,<sup>e</sup> and Richard P. Anderson, MD<sup>i†</sup>

<sup>a</sup>Massachusetts General Hospital, Boston, Massachusetts; <sup>b</sup>Duke Clinical Research Institute, Durham, North Carolina; <sup>c</sup>Institute for Health Care Research and Improvement, Baylor Health Care System, Dallas, Texas; <sup>d</sup>University of Kentucky Chandler Medical Center, Division of Cardiovascular and Thoracic Surgery, Lexington, Kentucky; <sup>e</sup>University of Florida, Division of Cardiothoracic Surgery, Jacksonville, Florida; <sup>f</sup>Sentara Cardiovascular Research Institute, Norfolk, Virginia; <sup>g</sup>Department of Health Care Policy, Harvard Medical School, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; <sup>h</sup>The Society of Thoracic Surgeons, Chicago, Illinois; and <sup>i</sup>Seattle, Washington

**Background.** The first version of The Society of Thoracic Surgeons National Adult Cardiac Surgery Database (STS NCD) was developed nearly 2 decades ago. Since its inception, the number of participants has grown dramatically, patient acuity has increased, and overall outcomes have consistently improved. To adjust for these and other changes, all STS risk models have undergone periodic revisions. This report provides a detailed description of the 2008 STS risk model for coronary artery bypass grafting surgery (CABG).

**Methods.** The study population consisted of 774,881 isolated CABG procedures performed on adult patients aged 20 to 100 years between January 1, 2002, and December 31, 2006, at 819 STS NCD participating centers. This cohort was randomly divided into a 60% training (development) sample and a 40% test (validation) sample. The development sample was used to identify predictor variables and estimate model coefficients. The validation sample was used to assess model calibration and discrimination. Model outcomes included operative mortality, renal failure, stroke, reoperation for any cause, prolonged ventilation, deep sternal wound infection, composite major morbidity or mortality, prolonged length of stay (> 14 days), and short length of stay (< 6 days and alive). Candidate predictor variables were selected based on their availability in versions 2.35, 2.41, and 2.52.1 of the STS NCD and their presence in (or ability to be mapped to) version 2.61. Potential predictor

variables were screened for overall prevalence in the study population, missing data frequency, coding concerns, bivariate relationships with outcomes, and their presence in previous STS or other CABG risk models. Supervised backwards selection was then performed with input from an expert panel of cardiac surgeons and biostatisticians. After successfully validating the fit of the models, the development and validation samples were subsequently combined, and the final regression coefficients were estimated using the overall combined (development plus validation) sample.

**Results.** The c-index for the mortality model was 0.812, and the c-indices for other endpoints ranged from 0.653 for reoperation to 0.793 for renal failure in the validation sample. Plots of observed versus predicted event rates revealed acceptable calibration in the overall population and in numerous subgroups. When patients were grouped into categories of predicted risk, the absolute difference between the observed and expected event rates was less than 1.5% for each endpoint. The final model intercept and coefficients are provided.

**Conclusions.** New STS risk models have been developed for CABG mortality and eight other endpoints. Detailed descriptions of model development and testing are provided, together with the final algorithm. Overall model performance is excellent.

(Ann Thorac Surg 2009;88:S2–22)

© 2009 by The Society of Thoracic Surgeons

In 1986, The Society of Thoracic Surgeons (STS) convened an Ad Hoc Committee on Risk Factors for Coronary Artery Bypass Graft Surgery (CABG) [1] and an

Ad Hoc Committee to Develop a National Database for Cardiothoracic Surgery [2]. This was prompted by the

†This author is deceased. Former Chair, Quality, Research and Patient Safety Council, The Society of Thoracic Surgeons, Chicago, IL.

Address correspondence to Dr Shahian, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114; e-mail: dshahian@partners.org.

Drs Shahian, O'Brien, Filardo, Ferraris, Haan, Rich, Normand, DeLong, Shewan, Peterson, Edwards, Anderson, and Ms Dokholyan have no conflicts of interest to declare regarding this work.

#### Abbreviations and Acronyms

BSA	= body surface area
CABG	= coronary artery bypass graft surgery
CHF	= congestive heart failure
EF	= ejection fraction
GFR	= glomerular filtration rate
HCFA	= Health Care Financing Administration
IABP	= intra-aortic balloon pump
NYHA	= New York Heart Association
NCD	= National Adult Cardiac Surgery Database
O/E	= observed to expected ratio
QMTF	= Quality Measurement Task Force
STS	= The Society of Thoracic Surgeons

release earlier that year of inadequately risk-adjusted hospital mortality data by the Health Care Financing Administration (HCFA), now the Centers for Medicare and Medicaid Services. Although the HCFA analytical methodology was widely criticized, STS leadership recognized that the underlying principle of collecting and analyzing data to improve patient outcomes was valid, particularly for complex and costly procedures such as coronary artery bypass grafting surgery (CABG). They believed that it was the responsibility of professional organizations to develop credible clinical data registries for their own specialties, and that risk models derived from such registries would circumvent many of the concerns resulting from the use of unadjusted administrative data. Such clinical registries would be used as credible data sources for quality assessment and improvement activities as well as for research.

These early activities ultimately led to the development of the STS National Adult Cardiac Surgery Database (NCD) [3, 4]. Since its release to members in 1990, the STS NCD has evolved to become one of the largest specialty-specific clinical data registries in the world. It currently has more than 950 participants enrolled, representing just under 90% of the cardiac surgery providers in the United States, with data on more than 3.6 million procedures. Similar STS data registries have now been developed for congenital heart surgery and general thoracic surgery, and future plans include the development of specialty modules (eg, quality metrics, atrial fibrillation surgery, thoracic aortic surgery). Recent enhancements, including the addition of unique physician and patient identifiers, will facilitate linkages with other registries and greatly expand the potential of the STS NCD for longitudinal follow-up, comparative effectiveness, and cost efficiency studies.

In addition to the development of the STS NCD as a comprehensive, nationally representative data registry, the second major goal of the STS was to assure that analyses derived from this registry would be appropriately adjusted for preoperative patient severity, a major deficiency of the HCFA reports that were initially published in 1986. This was accomplished by first identifying

risk factors for specific procedures and outcomes, beginning with isolated CABG, then using these predictor variables to develop risk models. With statistical risk models, which are most often based on logistic regression, the expected outcome for a patient with a given set of risk factors can be determined, and that can be compared with the observed outcome. The observed (O) and expected (E) outcomes are summed over all patients of a particular surgeon or hospital to yield the risk-standardized mortality ratio (O/E), which can then be multiplied by the average rate in the reference population to calculate risk-standardized mortality rates [5–7].

STS CABG risk models have undergone periodic updates and revisions, the most recent of which was based upon 2000 to 2002 STS NCD data. In 2007, the STS Database Modernization Task Force completed a major specification upgrade of the STS NCD data collection instrument from version 2.52.1 to version 2.61. This included refinement, modification, consolidation, or elimination of some data elements, as well as an attempt to harmonize definitions with those of the American College of Cardiology National Cardiovascular Data Registry whenever possible. Given these changes, as well as the number of years since the last risk model update, the STS Quality Measurement Task Force (QMTF) was asked to develop new risk models for isolated CABG, isolated valve repair or replacement, and combined CABG plus valve procedures. The authors of this report include the QMTF members who participated in this initiative.

Implementation of these new models in January 2008 coincided with the release of STS NCD version 2.61. This report, Part 1 of 3, describes the development of the new mortality and morbidity models for isolated CABG surgery.

#### Study Purpose

The primary goal of this study was to develop risk-prediction algorithms for patients undergoing isolated CABG surgery. As the major intended use of these algorithms was to compare participant outcomes to the overall STS national experience, risk factors were generally restricted to patient and clinical characteristics present preoperatively.

#### Risk Model Development and Transparency

The availability of user-friendly statistical software programs and the exponential increase in computing speed have greatly facilitated statistical analyses such as logistic regression, the basis for many risk models. However, despite these technological advances, clinical judgment, experience, intuition, and practicality still play a critical role in risk model development. There are many points in model development at which legitimate differences in approach may lead to substantial differences in the resulting statistical models and the inferences derived from them [8].

We believe the degree of transparency provided in this report regarding the development of the STS CABG risk

models is essential in today's health care environment. In an era when society demands full transparency regarding health care performance, the methodologies used to evaluate that performance should be just as transparent [9, 10]. This fundamental principle is among the standards established by the American Heart Association and American College of Cardiology for statistical models used for public reporting [11].

## Study Population and Endpoints

All isolated CABG procedures performed on adult patients aged 20 to 100 years between January 1, 2000, and December 31, 2006, were initially considered for inclusion, although the final development and validation samples were derived from 2002 to 2006 data. Patients missing data on sex ( $n = 195$ ) were excluded, as these patients are not included in STS performance feedback reports to database participants. That left a study population of 774,881 surgical procedures from 819 database participants. Patients on dialysis preoperatively ( $n = 12,415$ ) were excluded when developing the risk model for postoperative renal failure.

### *Training and Validation Samples*

The study population was randomly divided into a 60% training (development) sample and a 40% test (validation) sample. The development sample was used to identify predictor variables and estimate model coefficients. Data from the validation sample were used to assess model fit, discrimination, and calibration. After choosing variables and assessing model fit, the development and validation samples were subsequently combined, and the final model coefficients were estimated using the combined (development plus validation) data.

### *Endpoints*

Risk models were developed for the nine endpoints listed below. Only mortality was recorded beyond the index hospitalization. Morbidity data included only in-hospital complications, although beginning in STS NCD version 2.61, sternal infections will be recorded for up to 30 days postoperatively. The nine endpoints are as follows: (1) operative mortality: death during the same hospitalization as surgery, regardless of timing, or within 30 days of surgery regardless of venue; (2) permanent stroke (cerebrovascular accident): a central neurologic deficit persisting longer than 72 hours; (3) renal failure: a new requirement for dialysis or an increase of the serum creatinine to more than 2.0 mg/dL and double the most recent preoperative creatinine level; (4) prolonged ventilation (longer than 24 hours); (5) deep sternal wound infection; (6) reoperation for any reason; (7) major morbidity or mortality: a composite defined as the occurrence of any of the above endpoints; (8) prolonged postoperative length of stay (PLOS): length of stay (LOS) more than 14 days (alive or dead); and (9) short postoperative LOS (SLOS): LOS less than 6 days and patient alive at discharge (this SLOS definition differs from the previous STS risk models, which did not exclude patients who died in-hospital; patients who died within 5 days of surgery are

included in the new models but are treated as not having a short stay).

Table 1 summarizes the frequencies of these endpoints in the study population for each predictor variable category (ie, the bivariate relationships).

## Selection of Candidate Predictor Variables

### *Initial Data Screening of Candidate Predictor Variables*

We began by considering all possible candidate variables from the development set (Table 2). Because the primary goal of the STS risk models is to adjust surgical outcomes, in general only preoperative patient variables are included. However, because these models are also used for other purposes such as individual patient prediction and counseling, there were a few modifications (which are discussed in the relevant sections) in the application of this general principle.

As there were a large number of procedures and endpoints available, we were not statistically constrained to highly parsimonious models, nor is such an approach generally favored in regression modeling [12–14]. Discarding valid data elements can waste valuable information that has been collected at substantial effort and cost. Furthermore, although much of the discrimination of a predictive model may be contained in a relatively small number of variables [15, 16], some predictor variables that add only modestly to discrimination may still be important predictors of outcomes at the patient level [17, 18].

### *Expert Panel Review for Clinical Relevance and Face Validity*

All candidate variables available in version 2.61 were individually discussed by a panel of cardiac surgeons and health policy experts to assure that clinical relevance as well as multiple aspects of validity (face, construct, and content) had been considered.

### *Data Version for Model Development*

Although these new risk models were to be introduced in conjunction with the release of STS NCD version 2.61, they were developed with data collected under the three previous data versions (2.35, 2.41, and 2.52.1) because no 2.61 data were yet available. The QMTF began its predictor selection process with two caveats. First, any candidate variable had to be collected consistently across the three previous data versions. Second, it had to also be available in version 2.61 or have the ability to be mapped to this new version. For example, history of smoking and renal failure were not candidate variables as they were either not included in, or were unable to be mapped to, version 2.61. Renal function is now assessed by the last preoperative serum creatinine value, which is collected in all data versions. Because the definition of hypercholesterolemia has changed substantially over successive STS data versions, and because counterintuitive results have been observed in some previous analyses of hyper-

Table 1. Distribution of Risk Factors and Frequency of Adverse Outcomes in Overall Study Population, Isolated Coronary Artery Bypass Graft Surgery (2002–2006)

Variable	Number of Patients		Percent of Patients Experiencing Endpoint								
	N	%	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Overall											
Total	774,881	100.0	2.3	1.4	3.6	9.7	0.4	5.2	14.4	5.6	51.2
Age, years											
< 55	137,318	17.72	1.0	0.5	1.7	7.1	0.3	3.7	10.0	2.7	67.1
55–64	221,697	28.61	1.3	0.9	2.4	7.8	0.4	4.2	11.4	3.8	59.4
65–74	245,132	31.63	2.4	1.6	3.9	10.0	0.5	5.5	14.9	5.9	47.7
≥ 75	170,734	22.03	4.7	2.3	6.4	13.9	0.5	7.5	20.9	9.6	33.0
Sex											
Male	560,006	72.27	2.0	1.2	3.4	8.7	0.4	5.1	13.4	4.9	55.0
Female	214,875	27.73	3.4	1.9	4.1	12.2	0.5	5.6	17.0	7.2	41.5
Race											
Caucasian	665,941	85.94	2.3	1.3	3.5	9.3	0.4	5.1	13.9	5.3	52.2
Black	44,405	5.73	2.7	2.0	5.2	13.5	0.7	6.3	19.0	8.2	41.3
Hispanic	25,103	3.24	2.6	1.5	4.3	11.3	0.5	5.6	16.1	6.1	48.4
Asian	12,509	1.61	2.7	1.9	3.8	12.6	0.3	7.2	17.4	7.0	45.2
Other	21,222	2.74	2.3	1.3	3.6	10.4	0.5	5.5	14.8	6.0	48.7
Missing	5,701	0.74	2.3	1.4	4.1	9.4	0.4	5.2	14.5	6.1	48.9
Body surface area (m <sup>2</sup> )											
< 1.50	14,339	1.85	6.2	2.4	4.6	16.2	0.3	8.3	22.1	9.8	36.5
1.50–1.74	111,458	14.38	3.8	2.0	4.0	12.6	0.3	6.5	17.7	7.4	42.5
1.75–1.99	280,677	36.22	2.4	1.5	3.5	9.6	0.4	5.4	14.4	5.6	50.7
≥ 2.00	363,817	46.95	1.7	1.0	3.6	8.6	0.5	4.6	13.1	4.8	55.0
Missing	4,590	0.59	3.7	1.4	4.0	7.6	0.3	4.7	13.9	6.8	46.0
Body mass index (kg/m <sup>2</sup> )											
< 25	169,091	21.82	3.3	1.7	3.5	11.0	0.3	6.7	16.3	6.7	47.6
25–29	303,371	39.15	2.1	1.3	3.1	8.6	0.3	4.9	13.1	4.8	54.2
30–34	186,148	24.02	1.8	1.2	3.6	9.0	0.5	4.5	13.4	5.0	53.1
≥ 35	110,213	14.22	2.3	1.2	5.2	12.0	0.8	4.9	16.8	6.8	45.7
Missing	6,058	0.78	3.7	1.4	4.2	8.6	0.3	4.8	14.5	6.7	47.2
Diabetes mellitus											
No diabetes	492,800	63.60	2.1	1.2	2.8	8.8	0.3	5.0	13.0	4.7	54.8
Diabetes–noninsulin	195,421	25.22	2.3	1.6	4.3	10.1	0.5	5.2	15.2	6.0	48.2
Diabetes–insulin	84,406	10.89	3.6	1.8	7.1	13.9	1.0	6.5	20.6	9.7	37.5
Diabetes–missing treatment	1,439	0.19	3.1	2.2	4.6	11.1	0.7	4.6	15.7	8.8	41.9
Missing	815	0.11	3.8	0.7	2.6	8.8	0.5	4.5	12.3	6.9	43.7
Hypertension											
No	167,260	21.59	1.9	0.9	2.2	8.1	0.3	4.6	11.7	4.2	58.2
Yes	606,813	78.31	2.5	1.5	4.0	10.1	0.5	5.4	15.1	5.9	49.3
Missing	808	0.10	3.8	0.7	2.4	9.3	0.5	5.2	12.7	6.7	43.9
Hypercholesterolemia											
No	199,894	25.80	3.0	1.6	3.9	11.0	0.5	5.8	16.1	6.5	48.7
Yes	573,257	73.98	2.1	1.3	3.5	9.2	0.4	5.0	13.8	5.2	52.1
Missing	1,730	0.22	4.1	1.6	3.5	10.3	0.3	4.7	13.9	7.3	47.5
Past or present smoker											
No	295,999	38.20	2.4	1.4	3.7	9.0	0.4	5.1	13.9	5.3	50.1
Yes	477,911	61.68	2.3	1.3	3.6	10.1	0.5	5.3	14.7	5.7	51.9
Missing	971	0.13	3.4	0.7	3.1	9.9	0.4	5.3	13.5	9.1	41.0
Chronic lung disease											
None	612,211	79.01	2.0	1.3	3.3	8.4	0.3	4.9	13.0	4.7	53.7
Mild	85,005	10.97	2.8	1.5	4.2	12.0	0.6	5.8	16.9	7.0	45.7
Moderate	47,745	6.16	3.8	1.6	5.3	15.8	0.8	6.8	20.8	9.6	39.5
Severe	22,302	2.88	7.0	2.0	7.7	22.8	1.1	9.5	29.0	15.3	29.2
Missing	7,618	0.98	2.6	1.4	3.2	8.2	0.2	3.9	12.6	5.7	53.4



Table 1. Continued

Variable	Number of Patients		Percent of Patients Experiencing Endpoint								
	N	%	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Peripheral vascular disease											
No	653,260	84.30	2.0	1.2	3.2	8.8	0.4	4.8	13.1	4.8	53.5
Yes	120,480	15.55	4.4	2.3	6.1	14.4	0.7	7.5	21.2	9.6	38.7
Missing	1,141	0.15	3.9	1.1	3.1	11.8	0.3	5.5	14.4	7.4	43.9
Cerebrovascular disease											
No	668,073	86.22	2.1	1.1	3.3	9.0	0.4	4.9	13.4	5.0	53.4
Yes	105,792	13.65	4.0	2.9	5.8	14.0	0.6	7.2	20.7	9.3	37.7
Missing	1,016	0.13	3.2	0.6	2.4	8.9	0.3	4.2	11.4	6.8	43.2
CVA											
No CVA	717,721	92.62	2.2	1.2	3.4	9.3	0.4	5.1	13.8	5.2	52.5
Remote CVA (> 2 weeks)	53,341	6.88	4.2	3.1	6.1	15.3	0.7	7.4	22.0	10.3	35.5
Recent CVA (≤ 2 weeks)	1,763	0.23	5.0	4.9	6.5	18.8	0.9	8.7	25.4	13.0	32.5
CVA—missing timing	745	0.10	3.8	3.5	6.9	14.0	0.5	5.9	21.7	10.7	34.2
Missing	1,311	0.17	3.3	1.0	2.6	7.4	0.2	4.3	11.4	5.9	47.8
Endocarditis											
No endocarditis	773,002	99.76	2.3	1.4	3.6	9.7	0.4	5.2	14.4	5.5	51.2
Treated endocarditis	472	0.06	4.4	0.8	5.3	15.3	0.6	8.5	19.9	8.9	33.7
Active endocarditis	110	0.01	2.7	1.8	6.3	20.0	1.8	11.8	24.5	20.0	41.8
Endocarditis—missing type	90	0.01	4.4	4.4	5.8	11.1	1.1	3.3	15.6	2.2	55.6
Missing	1,207	0.16	4.1	1.0	3.5	9.0	0.2	4.6	12.8	7.0	46.2
Renal failure											
No	731,626	94.42	2.1	1.3	3.2	9.0	0.4	5.0	13.4	5.0	52.8
Yes	42,153	5.44	7.2	2.7	14.7	22.5	1.0	9.9	31.9	15.8	23.4
Missing	1,102	0.14	3.1	0.8	2.8	7.4	0.3	3.4	10.8	6.4	46.6
Renal function											
Creatinine < 1.00 mg/dL	274,197	35.39	1.6	1.1	1.5	8.0	0.3	4.4	11.2	4.0	55.6
Creatinine 1–1.49 mg/dL	398,833	51.47	2.0	1.3	3.4	8.9	0.4	5.0	13.5	4.9	53.1
Creatinine 1.5–1.99 mg/dL	57,779	7.46	4.5	2.3	10.8	16.1	0.7	7.8	25.2	10.6	34.5
Creatinine 2.0–2.49 mg/dL	12,463	1.61	6.9	2.9	14.3	21.3	0.9	9.4	31.5	15.3	24.7
Creatinine ≥ 2.5 mg/dL	7,906	1.02	8.2	3.2	20.4	23.4	0.9	11.1	37.9	18.6	20.4
Dialysis	12,415	1.60	8.4	2.7	*NA	25.3	1.2	10.5	31.5	16.4	19.6
Missing	11,288	1.46	3.3	1.2	3.1	7.6	0.3	4.3	12.9	5.9	50.1
Immunosuppressive treatment											
No	758,368	97.87	2.3	1.4	3.6	9.6	0.4	5.2	14.2	5.4	51.5
Yes	14,976	1.93	5.4	1.8	6.3	15.6	0.8	8.7	22.5	10.8	37.0
Missing	1,537	0.20	3.3	0.8	2.8	6.5	0.4	4.6	11.4	6.1	46.8
Prior CABG Surgery											
No	735,033	94.86	2.2	1.4	3.5	9.4	0.4	5.1	14.1	5.4	51.7
Yes	36,693	4.74	5.3	1.6	5.8	14.7	0.5	7.5	20.9	7.8	42.6
Missing	3,155	0.41	2.7	1.1	3.3	8.9	0.5	4.8	12.9	6.8	48.9
Prior valve surgery											
No	769,434	99.30	2.3	1.4	3.6	9.7	0.4	5.2	14.4	5.5	51.3
Yes	2,280	0.29	5.9	1.9	6.8	15.3	0.7	8.6	22.5	11.1	32.0
Missing	3,167	0.41	2.9	1.2	3.5	8.6	0.5	4.4	12.7	6.3	50.0
Prior other cardiac surgery											
No	755,653	97.52	2.3	1.4	3.6	9.6	0.4	5.2	14.3	5.5	51.3
Yes	15,218	1.96	3.9	1.5	4.9	13.1	0.6	6.6	18.6	7.6	45.5
Missing	4,010	0.52	2.8	1.0	2.9	8.5	0.4	4.4	12.2	5.9	50.5
Number of previous CV surgeries											
No previous CV surgery	723,623	93.39	2.2	1.4	3.5	9.4	0.4	5.1	14.0	5.4	51.7
One prior CV surgery	40,474	5.22	4.7	1.6	5.4	13.8	0.5	7.3	19.9	7.7	44.1
Two or more prior CV Surgeries	4,840	0.62	6.2	1.4	5.6	14.7	0.6	8.0	22.0	8.4	41.5
Missing	5,944	0.77	2.9	1.1	3.2	9.8	0.6	5.5	14.3	6.1	51.2

Table 1. Continued

Variable	Number of Patients		Percent of Patients Experiencing Endpoint								
	N	%	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Prior PCI											
No PCI	606,824	78.31	2.3	1.4	3.6	9.5	0.4	5.1	14.2	5.6	51.1
PCI ≤ 6 hours	7,373	0.95	8.9	2.1	7.4	25.8	0.6	10.5	32.6	11.3	35.5
PCI > 6 hours	155,161	20.02	2.3	1.2	3.5	9.6	0.4	5.4	14.3	5.1	52.4
PCI-missing timing	2,456	0.32	3.0	0.6	3.3	8.0	0.8	5.1	13.4	6.8	48.9
Missing	3,067	0.40	3.3	1.0	3.0	9.8	0.5	4.7	13.7	6.2	47.3
Acuity status											
Elective	381,116	49.18	1.5	1.1	2.9	6.6	0.4	4.3	11.1	4.1	55.6
Urgent	356,287	45.98	2.4	1.5	3.9	10.8	0.5	5.6	15.7	6.2	48.5
Emergent	34,513	4.45	8.1	2.6	8.3	29.6	0.7	10.4	34.1	13.3	33.2
Emergent salvage	1,967	0.25	38.6	4.9	17.4	52.7	0.7	18.4	70.0	23.5	12.6
Missing	998	0.13	3.2	1.3	3.6	9.2	0.5	4.8	13.7	6.5	43.6
MI											
No prior MI	424,599	54.80	1.5	1.1	2.8	6.9	0.3	4.5	11.3	4.1	55.4
MI > 21 days	137,522	17.75	2.1	1.3	3.5	8.9	0.5	5.2	13.9	5.4	50.4
MI 8–21 days	26,205	3.38	4.0	1.8	6.0	14.4	0.8	7.5	20.7	10.2	38.1
MI 1–7 days	148,659	19.18	3.4	1.8	4.8	14.0	0.5	6.1	18.9	7.5	45.7
MI > 6 and < 24 hours	21,044	2.72	6.0	2.4	6.7	23.6	0.5	8.1	28.1	10.4	39.1
MI ≤ 6 hours	11,539	1.49	10.4	2.6	8.6	31.2	0.6	10.6	36.8	13.3	33.5
MI-missing timing	4,064	0.52	3.6	1.6	4.6	11.3	0.5	5.6	17.5	7.2	43.9
Missing	1,249	0.16	2.1	1.1	2.4	6.6	0.2	3.8	10.2	6.7	49.4
Angina											
No	130,143	16.80	2.5	1.4	3.8	9.5	0.4	5.6	14.7	6.2	48.5
Yes	643,815	83.09	2.3	1.3	3.6	9.7	0.4	5.2	14.3	5.4	51.8
Missing	923	0.12	2.3	1.0	2.6	8.8	0.5	4.0	11.2	8.2	43.9
Cardiogenic shock											
No	758,766	97.92	2.0	1.3	3.4	8.9	0.4	5.0	13.6	5.2	51.9
Yes	14,919	1.93	18.0	3.6	14.6	49.6	1.0	15.3	55.7	23.1	18.3
Missing	1,196	0.15	2.7	1.1	3.2	8.0	0.4	4.4	12.0	7.4	44.7
Resuscitation											
No	766,674	98.94	2.2	1.3	3.5	9.4	0.4	5.1	14.1	5.4	51.5
Yes	6,939	0.90	17.1	3.0	11.4	37.5	0.9	14.0	46.1	18.2	24.3
Missing	1,268	0.16	2.2	0.8	3.4	8.0	0.6	3.9	11.5	7.3	45.0
Arrhythmia											
No arrhythmia	706,709	91.20	2.0	1.3	3.3	8.9	0.4	4.9	13.4	5.0	53.1
AFib/flutter	39,125	5.05	5.4	2.3	7.1	16.4	0.7	8.5	23.8	11.9	29.4
Heart block	10,026	1.29	5.8	1.9	6.8	16.8	0.6	9.2	24.2	9.4	36.4
Sustained VT/VF	14,336	1.85	8.2	2.0	6.8	23.8	0.6	11.1	31.5	12.0	33.0
Arrhythmia–other	1,853	0.24	3.8	1.6	5.3	12.9	0.6	6.5	19.1	7.3	39.2
Arrhythmia–missing type	1,344	0.17	3.9	1.7	4.4	11.9	0.7	7.3	17.5	8.3	37.9
Missing	1,488	0.19	2.7	1.0	3.0	7.1	0.5	3.8	11.1	6.5	45.5
Preoperative IABP											
No	714,824	92.25	2.0	1.3	3.3	8.0	0.4	4.9	12.8	4.9	52.8
Yes	58,134	7.50	6.9	2.2	7.7	30.8	0.6	9.6	34.4	12.9	32.0
Missing	1,923	0.25	4.2	1.7	4.3	10.9	0.6	5.8	16.0	7.2	45.7
NYHA class											
I	97,812	12.62	1.5	1.1	2.5	6.3	0.3	4.4	10.6	3.9	57.0
II	187,947	24.25	1.3	1.1	2.6	6.5	0.3	4.2	10.7	3.8	56.5
III	287,760	37.14	2.0	1.3	3.6	9.0	0.4	5.0	13.9	5.4	51.3
IV	165,325	21.34	4.5	1.9	5.5	16.6	0.6	7.1	21.9	8.8	42.5
Missing	36,037	4.65	2.4	1.2	3.6	8.9	0.3	5.3	13.6	5.8	47.7
Congestive heart failure											
No	666,592	86.03	1.8	1.2	2.9	7.9	0.3	4.7	12.2	4.3	54.7
Yes	106,700	13.77	5.9	2.4	8.5	21.0	0.9	8.7	28.0	13.2	29.5
Missing	1,589	0.21	3.3	1.4	3.1	9.2	0.4	4.5	13.0	7.4	49.6

Table 1. Continued

Variable	Number of Patients		Percent of Patients Experiencing Endpoint								
	N	%	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Number of diseased coronary vessels											
None	2,012	0.26	2.3	0.4	2.8	8.9	0.4	4.6	12.6	5.5	53.1
One	32,311	4.17	1.5	0.6	1.9	6.1	0.2	4.4	9.8	3.1	66.3
Two	150,881	19.47	1.8	1.0	2.7	8.0	0.4	4.5	12.0	4.5	56.3
Three	586,658	75.71	2.5	1.5	4.0	10.4	0.4	5.5	15.3	6.0	49.1
Missing	3,019	0.39	2.6	0.6	1.8	5.5	0.2	4.5	10.8	5.9	54.2
Left main disease $\geq 50\%$											
No	554,355	71.54	2.1	1.3	3.4	8.8	0.4	5.0	13.5	5.1	52.7
Yes	217,548	28.08	3.0	1.5	4.3	11.9	0.5	5.9	16.8	6.6	47.6
Missing	2,978	0.38	2.3	1.4	2.7	6.3	0.3	5.5	11.9	5.9	45.4
Ejection fraction (%)											
< 25	25,323	3.27	7.2	2.2	8.0	25.2	0.8	10.5	31.9	13.7	27.8
25–34	57,460	7.42	4.6	2.1	6.1	17.6	0.6	7.6	23.8	10.3	36.8
35–44	108,623	14.02	3.0	1.7	4.7	12.4	0.6	6.0	17.5	7.2	45.7
45–54	189,478	24.45	1.9	1.3	3.4	8.7	0.4	4.8	13.2	5.0	53.1
$\geq 55$	351,455	45.36	1.5	1.1	2.7	6.8	0.3	4.4	11.1	3.9	56.1
Missing	42,542	5.49	3.4	1.4	4.1	10.8	0.4	5.6	16.0	6.1	50.0
Mitral stenosis											
No	756,609	97.64	2.3	1.4	3.6	9.7	0.4	5.2	14.4	5.5	51.2
Yes	2,703	0.35	5.5	2.4	6.4	17.0	0.7	7.5	22.9	10.5	35.7
Missing	15,569	2.01	2.1	1.3	3.4	8.3	0.4	4.6	13.1	5.0	53.1
Aortic stenosis											
No	750,185	96.81	2.3	1.4	3.6	9.6	0.4	5.2	14.3	5.5	51.4
Yes	11,386	1.47	4.7	2.1	6.5	14.8	0.7	7.9	21.5	9.7	36.6
Missing	13,310	1.72	2.3	1.3	3.3	8.5	0.4	4.7	13.1	5.0	52.5
Tricuspid stenosis											
No	756,574	97.64	2.3	1.4	3.6	9.7	0.4	5.2	14.4	5.6	51.2
Yes	597	0.08	3.4	2.3	6.6	14.9	0.7	6.0	20.9	10.1	43.6
Missing	17,710	2.29	2.1	1.3	3.6	8.5	0.4	4.7	13.4	5.0	53.2
Pulmonic stenosis											
No	753,975	97.30	2.3	1.4	3.6	9.7	0.4	5.2	14.4	5.6	51.2
Yes	445	0.06	3.4	2.2	3.9	12.6	0.0	6.3	20.2	6.3	49.4
Missing	20,461	2.64	2.2	1.4	3.8	8.7	0.4	5.0	13.9	5.3	52.5
Mitral insufficiency											
None	622,173	80.29	2.1	1.2	3.3	8.9	0.4	4.9	13.4	5.0	53.2
Trivial	49,152	6.34	2.4	1.6	4.2	10.5	0.4	5.7	15.7	6.2	47.9
Mild	60,811	7.85	3.7	2.0	5.7	14.3	0.5	6.9	20.3	8.6	40.3
Moderate	16,723	2.16	6.7	2.7	7.9	20.1	0.7	9.6	28.0	12.5	30.3
Severe	2,143	0.28	8.7	3.1	8.9	24.1	0.6	11.0	32.6	15.1	28.2
Missing	23,879	3.08	2.1	1.2	3.0	7.5	0.4	4.7	12.0	5.2	51.5
Aortic insufficiency											
None	705,771	91.08	2.3	1.3	3.5	9.5	0.4	5.1	14.1	5.4	51.9
Trivial	17,988	2.32	3.6	2.1	5.6	13.4	0.5	7.0	19.4	8.3	40.9
Mild	18,571	2.40	4.1	2.2	5.9	14.3	0.4	7.3	20.8	9.0	37.9
Moderate	3,576	0.46	5.3	2.6	7.0	16.2	0.5	7.9	23.2	10.2	32.8
Severe	411	0.05	7.1	1.9	6.7	15.6	0.7	9.5	25.8	10.9	37.2
Missing	28,564	3.69	2.1	1.2	3.2	7.9	0.4	4.6	12.5	5.3	51.4
Tricuspid insufficiency											
None	675,778	87.21	2.2	1.3	3.4	9.4	0.4	5.1	13.9	5.3	52.1
Trivial	32,856	4.24	2.5	1.6	4.5	11.1	0.4	6.1	16.6	6.7	47.4
Mild	29,611	3.82	3.9	2.2	5.9	14.7	0.5	7.4	21.0	9.1	39.3
Moderate	5,753	0.74	7.6	3.0	9.0	22.7	0.5	9.8	30.2	13.7	26.9
Severe	728	0.09	9.1	2.9	10.5	24.2	0.4	10.9	33.0	17.2	26.2
Missing	30,155	3.89	2.2	1.2	3.2	7.9	0.4	4.6	12.6	5.2	51.7

Table 1. Continued

Variable	Number of Patients		Percent of Patients Experiencing Endpoint								
	N	%	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Pulmonic insufficiency											
None	724,258	93.47	2.3	1.4	3.6	9.7	0.4	5.2	14.3	5.5	51.4
Trivial	10,726	1.38	2.8	1.5	4.3	12.3	0.5	6.5	17.3	7.3	44.8
Mild	4,867	0.63	3.8	2.1	5.6	14.1	0.4	7.4	21.0	9.1	39.7
Moderate	546	0.07	6.6	3.1	7.8	17.6	0.2	7.9	24.4	11.5	29.7
Severe	217	0.03	5.1	0.5	5.1	9.7	0.5	6.5	15.7	8.8	50.7
Missing	34,267	4.42	2.2	1.3	3.5	8.3	0.4	4.8	13.2	5.5	50.7

AFib = atrial fibrillation; CABG = coronary artery bypass graft surgery; Comp = composite adverse outcome (any); CV = cardiovascular; CVA = cerebrovascular accident (stroke); DSWI = deep sternal wound infection; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; Na = not applicable; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation; VF = ventricular fibrillation; VT = ventricular tachycardia.

cholesterolemia, a decision was made not to include this variable in the new models.

### Predictor Frequency

For each variable, the QMTF explored the overall prevalence and missing data frequency per year. Predictor variables that are rarely present in the development sample are difficult to model. For this reason, mitral (0.35%), tricuspid (0.08%), and pulmonic stenosis (0.06%), pulmonic insufficiency (0.10%), and endocarditis (0.09%) were not considered as variables in the new isolated CABG models.

### Inconsistently Coded Variables

A few variables have been collected inconsistently or with questionable reliability, often for clinically unavoidable reasons. For example, pulmonary artery mean pressure data were missing for 70% of patients during 2002 to 2006. Furthermore, the value of this continuous variable may vary substantially depending on the clinical state and volume-loading status of the patient when the measurement is obtained. Because of these concerns, pulmonary artery pressure was not included in the models.

### Derived or Redundant Variables

Several derived variables were considered for inclusion in the models. For example, body mass index (BMI) is a useful measure of overall body habitus. However, because BMI is highly correlated with body surface area (BSA), the more commonly used anthropometric measure in most previous STS models, the latter was retained in the new models. Similarly, there is a theoretical superiority to inclusion of glomerular filtration rate (GFR) rather than serum creatinine as a measure of renal function. However, the Modification of Diet in Renal Disease formula for estimating GFR is a complex function of creatinine, race, sex, and age, and not all laboratories perform this calculation automatically. Furthermore, as age, sex, and race are already model covariates, using GFR would complicate the interpretation of their regression coefficients. Some of the prognostic value of GFR

comes from these variables that are already included in the model. Finally, previous studies suggest that various measures of renal function used in CABG mortality risk models have similar performance [19]. For all these reasons, serum creatinine was retained as the measure of renal function.

### Controversial Variables

RACE. Several variables raised particular clinical, statistical, or health policy issues. For example, race was an obvious candidate variable because it was a significant predictor ( $p < 0.001$ ) of each endpoint except mortality and because the proportion of nonwhite patients varied substantially across institutions. In exploratory analyses, the association between race and outcomes persisted after adjusting for hospital identity, suggesting that this association is not explained by differences in hospital quality.

However, general principles of risk model development complicated the decision as to whether or not to include race in the models. When the dominant purpose of a risk model is adjustment of provider results, it is advisable to include only biological and clinical patient variables that are present before a patient's first contact with the provider. In this context, race is clearly a fixed biological characteristic, but its impact on patient outcomes may be mediated through other mechanisms. It is possible that certain racial and ethnic groups have worse outcomes not because of inherent biological characteristics but because of differences in the quality of care delivered to them. In this case, including race and ethnicity in a risk model could essentially select out or obscure the very disparity issues that society wishes to identify and correct. Inclusion of race and ethnicity in a risk model would say, in effect, that we expect nonwhites to have inferior results and would make an allowance for providers who care for such patients, just as we would for providers who care for patients in cardiogenic shock.

After deliberation regarding the pros and cons, the QMTF ultimately elected to retain race and ethnicity in the new models because of their impact on outcomes,

Table 2. Initial List of Potential Candidate Variables

Demographics

1. Age
2. Sex
3. Race (black, Caucasian, Hispanic, Asian, Native American, other)

*Note: Data collection changed in v2.61. New version allows for multiple races (check all that apply). Added Hawaiian/Pacific Islander category. Hispanic ethnicity is a separate variable.*

Anthropometric

4. Height
5. Weight

Status

6. Status (elective, urgent, emergent, salvage)
7. Shock
8. Resuscitation

Cardiac variables

9. Angina, angina type (STS categories are unstable, stable, no angina)

*Note: Angina was removed on v2.61 data collection form. The new form has a variable called "cardiac presentation on admission." Angina is one of possible response categories to that field.*

10. New York Heart Association functional class

*Note: In v2.61, NYHA class is only collected if patient has congestive heart failure.*

11. Arrhythmia and arrhythmia type (sustained VT/VF; heart block; AFib/flutter, None)
12. Myocardial infarction timing: ( $\leq 6$ ,  $> 6$  and  $< 24$  hours; 1–7, 8–21,  $> 21$  days)

Hemodynamic/catheterization variables

13. Ejection fraction
14. Number of diseased vessels (0, 1, 2, 3)
15. Left main disease
16. Pulmonary artery mean pressure
17. Mitral stenosis
18. Aortic stenosis
19. Tricuspid stenosis
20. Pulmonic stenosis
21. Mitral insufficiency (none, trivial, mild, moderate, severe)
22. Aortic insufficiency (none, trivial, mild, moderate, severe)
23. Tricuspid insufficiency (none, trivial, mild, moderate, severe)
24. Pulmonic insufficiency (none, trivial, mild, moderate, severe)

Comorbidities

25. Serum creatinine
26. Dialysis
27. Renal failure
- Note: This variable was removed in v2.61.*
28. Endocarditis (active, treated, none)
29. Diabetes and treatment (insulin, oral, diet, untreated, no diabetes)
30. Chronic lung disease (none, mild, moderate, severe)
31. Congestive heart failure
32. Peripheral vascular disease
33. Cerebrovascular disease
34. CVA and CVA timing (recent, remote, none)

*Note: CVA is a child field of cerebrovascular disease in v2.61.*

35. Hypercholesterolemia (v2.35, v2.41) and Dyslipidemia (v2.52)

*Note: Data from all 3 versions were merged and analyzed under the variable name "hypercholesterolemia."*

36. Hypertension
37. Smoker

*Note: Major definition change in v2.61.*

Preoperative interventions

38. Preoperative intra-aortic balloon pump
39. Preoperative inotropes
40. Immunosuppressive treatment
41. Prior percutaneous coronary intervention and timing ( $\leq 6$  hours,  $> 6$  hours, none)

Previous Interventions

42. Prior coronary artery bypass graft surgery
43. Prior valve surgery
44. Prior other cardiac surgery
45. Number of previous cardiovascular surgeries



while recognizing the potential limitations of this decision.

**PREOPERATIVE INTRA-AORTIC BALLOON PUMP.** Preoperative intra-aortic balloon pump (IABP) is a proxy for more serious preoperative status of the patient (eg, unstable angina, ventricular dysfunction). It captures information that may not be present in other data elements, and it is associated with higher risk of postoperative morbidity and mortality. For these reasons, most CABG risk models include preoperative IABP as a risk predictor. However, placement of an IABP is also a highly discretionary care process the frequency of which varies widely among participating institutions. Indications are subjective and are often dictated by the cardiologist before even referring the patient for cardiac surgery. Based on CABG risk models, an institution that liberally utilizes IABPs will have a higher expected risk of morbidity and mortality (according to the model) compared with another institution with a similar case-mix but a more restrictive IABP policy. That would impact their relative O/E ratios and risk-adjusted outcomes.

Despite its discretionary nature (and the potential for gaming), the QMTF decided to retain IABP use in the models because it is such an important predictor. Ultimately, it was elected to model preoperative IABP as a joint variable with preoperative inotrope use as an overall measure of preoperative acuity/severity.

#### *Review of External Sources*

The QMTF also reviewed multiple external resources to aid in the selection of potential candidate variables [15, 16, 20]. First, all previous versions of the STS CABG risk models were reviewed. The QMTF also examined other CABG risk models including the European System for Cardiac Operative Risk Evaluation (EuroSCORE) [21], the New York Cardiac Surgery Reporting System [22], the Veterans Affairs Administration cardiac surgery models [23, 24], and the Northern New England Cardiovascular Disease Study Group model [25, 26]. We particularly wanted to identify variables that were found in some form across all the risk models. Subject to the constraints of version 2.61 data specifications, we made a special effort to include such variables in the new STS risk models, in some instances requiring us to “force” them into the models, as described in the section on the final variable selection procedure.

#### **Missing Data**

Missing data in the STS NCD are rare, having a frequency of less than 1% for most variables. Candidate predictor variables missing most commonly were ejection fraction (5.5%), New York Heart Association (NYHA) class (4.7%), tricuspid insufficiency (3.9%), aortic insufficiency (3.7%), mitral insufficiency (3.1%), aortic stenosis (1.7%), and creatinine/dialysis (1.5%).

Missing predictor values in the STS NCD were managed using imputation. Multiple imputation is the generally preferred statistical method [27], but single imputation was also considered based on the following

practical considerations: (a) the fraction of missing data in the STS NCD was small and, hence, single and multiple imputation would likely give similar point estimates; (b) a slight adjustment to the standard errors would not impact the study conclusions or the published risk algorithms; (c) the large sample size would make multiple imputation less practical to implement because of long computational times.

Prior to selecting an imputation strategy, exploratory analyses were performed using CABG data from 2002 to 2003 to compare single versus multiple imputation results for predicting mortality. These analyses confirmed that the choice between single versus multiple imputation would have only a slight impact on regression coefficients. For example, the estimated odds ratio for a 5-unit increase in ejection fraction was 0.90 (with a 95% confidence interval extending from 0.83 to 0.97) under single imputation and was 0.92 (with a confidence interval extending from 0.85 to 0.99) under multiple imputation. Other variables were missing less frequently than ejection fraction and were even less sensitive to the choice between single versus multiple imputation. Additional analyses of missing data consisted of reestimating the final model coefficients using single versus multiple imputation and comparing results. A summary of these investigations, as well as model coefficients and covariance matrices, are available at [www.sts.org/riskmodels](http://www.sts.org/riskmodels). For most patients, if risk were calculated using the multiple imputation model instead of single imputation, the relative change in their risk estimate would only be 1% to 2% (eg, 5% to 5.1% is a 2% change).

Based on the considerations described above, single imputation was used with the following specific rules: (1) binary (yes/no) risk factors were modeled as yes versus no or missing. Missing data for such variables usually implies their absence, and for most binary variables the composite event rates were similar for “no” and “missing” categories; (2) missing data on categorical predictor variables were imputed to the lowest risk value, which, in most instances, was the mode. In most instances, composite event rates for patients with missing data were among the lowest. It is the policy of the STS Data Warehouse and Analysis Center to discourage missing data through this default coding practice; and (3) missing data on continuous predictor variables were imputed to the conditional median. For ejection fraction, we conditioned on congestive heart failure (CHF) and sex. For BSA, we conditioned on sex. For serum creatinine, we conditioned on renal failure (although this approach will be modified when the model is ultimately applied to version 2.61 data, as renal failure has been removed).

For model endpoints (eg, mortality), missing data were handled by modeling yes versus no or missing. Thus, cases with missing data for an endpoint were analyzed as if the endpoint did not occur. Complete case analysis was not used because “missing” was not considered to be consistently coded for these variables. For example, some STS data managers have reported that they set complications to “no” unless there is explicit documentation in the medical record that the complication occurred. Other

data managers may leave the field missing unless there is explicit documentation that the complication did not occur. Thus, missing data may reflect differences in coding practices rather than truly unknown or missing data.

### Preliminary Analyses for Ordinal Categorical Variables and Continuous Variables

The QMTF conducted preliminary analyses to determine how best to model ordinal categorical variables and continuous variables. Categorical variables were entered into a logistic regression model by including a separate parameter for each category. Continuous variables were entered as piecewise linear functions (splines) with several changes of slope (knots). Terms were then removed one at a time using backward selection based on the Wald statistic. At each iteration, either two adjacent categories were collapsed into a single category or else two adjacent line segments were collapsed into a single line with no change of slope. The backward selection terminated when all adjacent categories and slopes were statistically different from one another at  $p < 0.001$ . This variable selection routine was performed separately for each endpoint. An expert panel determined the final coding based on the results of the backwards selection algorithm, supplemented by their clinical judgment and practical considerations. Table 3 summarizes these coding decisions.

### Specific Coding Decisions

**RACE AND ETHNICITY.** In versions 2.35, 2.41, and 2.52.1, race was collected by choosing one of the following mutually exclusive response categories: Caucasian, black, Hispanic, Asian, Native American, and other. In version 2.61, the data collection form was modified to conform to standards adopted by the US Census Bureau. It allows for selecting one or more races per patient (ie, select all that apply), and treats ethnicity (Hispanic versus non-Hispanic) as a separate variable. Because of these differences, the mapping of race among data versions is not straightforward.

Ultimately, the QMTF decided to model race as black, Asian, Hispanic, and Caucasian/other (collapsed). Initially, these categories will be mapped to version 2.61 as follows: (1) black will include all black patients, regardless of ethnicity or additional races; (2) Hispanic will include all nonblack Hispanic patients; (3) Asian will include all Asian patients who are not also identified as black or Hispanic; and (4) all remaining patients will be placed in the Caucasian/other category. The validity of this mapping will be assessed once 2.61 data become available and future versions could employ race “bridging” methodologies.

**BODY SURFACE AREA.** Height and weight were replaced by BSA, which was modeled as a quadratic trend to allow for a possible U-shaped relationship with outcomes (eg, extreme obesity and cachexia). This quadratic polynomial was modeled separately for males and females. Any BSA values below 1.4 or above 2.6 were mapped to these

values respectively, which represent the approximate 1st and 99th percentiles of the empirical distribution.

**ANGINA.** Version 2.61 of the data collection form eliminates angina and substitutes a new variable called “cardiac presentation on admission,” within which unstable angina is one of the possible response categories. The QMTF believed that unstable angina would be coded more consistently than any other angina class, and also that this was the most important type of angina presentation to include in the models. Angina coding was therefore restricted in the new risk models to “unstable angina without MI < 7 days (yes/no).” It was necessary to exclude patients with myocardial infarction less than 7 days because the new version 2.61 does not permit simultaneous coding of angina and acute myocardial infarction.

**REOPERATIVE STATUS.** The most important consideration with regard to reoperative status is the number of prior sternotomies, irrespective of the specific type of procedure performed. The revised models replaced prior CABG, prior valve, and prior “other” cardiac surgery with simply the number of previous cardiovascular surgeries.

**ACUITY STATUS.** The new models combine resuscitation with salvage status. By definition, all salvage patients should have resuscitation coded “yes.”

**NUMBER OF DISEASED CORONARY VESSELS.** Outcomes are modeled using the number of diseased vessels (grouped as 0 or 1 versus 2 versus 3), as a linear effect across the three categories. This approach is consistent with the previous STS CABG models and was supported by the data.

**NYHA CLASS.** Version 2.61 uses NYHA class as a subfield of CHF. The grouping of NYHA IV versus less than IV (I–III) classes is consistent with all existing STS models. The final categories were no CHF, CHF not NYHA IV, and CHF plus NYHA IV.

**AGE.** Age was modeled as a linear spline with knots at ages 50 and 60 years.

**EJECTION FRACTION.** Ejection fraction (EF) was modeled linearly, and EFs below 10% and above 50% were mapped to these values respectively. Only 0.03% of patients have EFs lower than 10%; such values are considered invalid and are treated like missing data. The coding decision regarding EF values above 50% was based on preliminary analyses in which the data were used to suggest the functional form of continuous variables.

**CREATININE.** Creatinine was modeled as a linear spline with knots at 1.0 and 1.5. Creatinine levels less than 0.5 or greater than 5.0 were mapped to these values respectively, which represent the approximate 1st and 99th percentiles of the empirical distribution.

**MORTALITY AND LENGTH OF STAY.** The QMTF changed the previous STS definition of the “short postoperative length of stay (SLOS)” endpoint. The original definition did not specifically exclude early postoperative deaths, and such patients could have been inappropriately included with the remaining SLOS patients who had a particularly short and uncomplicated postoperative course. In the new models, patients who die within 5 days of surgery are included in the analysis but are not counted as a short stay.

Table 3. Final List of Candidate Variables and Coding For STS Risk Models

Candidate Variables	Coding
<b>Continuous variables</b>	
Age <sup>a</sup>	Linear spline with knots at 50 and 60.
Ejection fraction <sup>a</sup>	Linear; values > 50 are mapped to 50. Only 0.03% of patients have ejection fraction < 10, and that is presumed to be a data entry error; these values are considered invalid and are treated like missing data. The decision to consolidate values > 50 was based on initial exploratory analyses in which data were used to suggest the functional form of continuous variables.
Body surface area <sup>a</sup>	Quadratic polynomial modeled separately for males and females. Note: body surface areas < 1.4 and > 2.6 were mapped to these values, respectively. <sup>c</sup>
Creatinine <sup>a</sup>	Linear spline with knots at 1.0 and 1.5. (Only for patients not on dialysis.) Note: Creatinine values < 0.5 and > 5.0 were mapped to these values, respectively. <sup>d</sup>
Time trend <sup>a</sup>	Ordinal categorical variable with separate category for each 6-month harvest interval.
<b>Binary variables</b>	
Dialysis <sup>a</sup>	Yes/no
Preoperative atrial fibrillation <sup>b</sup>	Yes/no
Shock	Yes/no
Female <sup>a</sup>	Yes/no
Hypertension	Yes/no
Immunosuppressive treatment	Yes/no
Percutaneous coronary intervention ≤ 6 hours	Yes/no
Preoperative intra-aortic balloon pump or inotropes	Yes/no
Peripheral vascular disease	Yes/no
Unstable angina (no myocardial infarction < 7 days)	Yes/no
Left main disease	Yes/no
Aortic stenosis	Yes/no
Aortic insufficiency	Defined as at least moderate (yes/no)
Mitral insufficiency	Defined as at least moderate (yes/no)
Tricuspid insufficiency	Defined as at least moderate (yes/no)
<b>Categorical variables</b>	
Chronic lung disease	4 groups: (1) none, (2) mild, (3) moderate, (4) severe
CVD/CVA	3 groups: (1) no CVD, (2) CVD no CVA, (3) CVD + CVA
Diabetes mellitus	3 groups: (1) insulin diabetes, (2) noninsulin diabetes, (3) other or no diabetes
Number diseased coronary vessels	3 groups: (1) fewer than 2 diseased vessels, (2) 2 disease vessels, (3) 3 diseased vessels; modeled as linear across the categories.
Myocardial infarction	4 groups: (1) ≤ 6 hours, (2) > 6 and < 24 hours, (3) 1 to 21 days, (4) > 21 days or no myocardial infarction.
Race	4 groups: (1) black, (2) Asian, (3) Hispanic, (4) other, including Caucasian
Status	4 groups: (1) elective, (2) urgent, (3) emergent, no resuscitation, (4) salvage or emergent with resuscitation
Previous cardiovascular operations	3 groups: 0 previous, 1 previous, 2 or more previous
CHF and NYHA class	3 groups: no CHF, CHF not NYHA IV, CHF + NYHA IV
<b>Interactions</b>	
Age by reoperation <sup>a</sup>	
Age by emergent status <sup>a</sup>	

<sup>a</sup> These variables were forced into each model. <sup>b</sup> Preoperative atrial fibrillation was forced into the model for stroke. <sup>c</sup> These are the approximate 1st and 99th percentiles of the empirical distribution. Values less than 1.4 were mapped to 1.4. Values greater than 2.6 were mapped to 2.6. Estimates in the extreme tails of the body surface area distribution are highly influenced by data from other regions of the body surface area distribution (owing to use of a parametric, quadratic model) and may not be reliable. <sup>d</sup> These are approximately the 1st and 99th percentiles of the empirical distribution. Although we used a flexible spline model, linear splines can have unreliable extreme results in the tails due to the assumption that the effect is linear above the largest knot and below the smallest knot.

CHF = congestive heart failure; CVA = cerebrovascular accident; CVD = cerebrovascular disease; NYHA = New York Heart Association.

Table 4. Discrimination of Models (C-Index)

New STS models—development sample (C-index)								
Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
0.810	0.716	0.795	0.756	0.706	0.657	0.724	0.769	0.727
New STS models—validation sample (C-index)								
Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
0.812	0.720	0.793	0.754	0.689	0.653	0.725	0.767	0.726
Old STS models—validation sample (C-index)								
Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
0.807	0.713	0.750	0.742	0.672	0.645	0.711	0.754	0.713

Comp = composite adverse outcome (any); CVA = stroke; DSWI = deep sternal wound infection; Mort = mortality; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; STS = The Society of Thoracic Surgeons; Vent = prolonged ventilation.

## Final Variable Selection Procedure

### Backward Selection

Using the remaining candidate variables and the coding schemes described previously, a supervised backward selection approach was then performed. Initial variable selection used the Wald  $\chi^2$  statistic with a significance criterion of 0.001. This high level of significance was chosen because of the very large sample size that resulted in quite small *p* values. An expert panel of cardiothoracic surgeons and biostatisticians then reviewed the selected variables and made several modifications. Measures of model performance (discrimination and calibration) were similar when all variables were retained in the models regardless of statistical significance or expert panel review.

### Forced Variables

Several variables were included in the models regardless of statistical significance. These included all of the continuous variables (age, BSA, date of surgery [in 6-month intervals], creatinine, ejection fraction), plus sex and dialysis. In addition, atrial fibrillation was included a priori in the model for permanent stroke.

The rationale for including surgery date, a nonmodifiable variable of no intrinsic interest, was to adjust for changes in the frequency of adverse outcomes over the 5-year study period. We adjusted for surgery date to reduce potential confounding by time trends when estimating regression coefficients for variables that are of primary interest, such as preoperative clinical characteristics. For example, temporal changes in the frequency of coding for dyslipidemia, if they occur coincidentally with a secular declining trend in mortality rates, may lead to the unwarranted causal inferences unless there is adjustment for surgery date.

Date of surgery was categorized by 6-month intervals (corresponding to STS data harvests) and modeled as a linear trend across the ordinal categories. Surgery date is not included in the final risk algorithm and a patient's predicted risk is not dependent upon it. The intercept

parameter published in the Appendix has been adjusted to incorporate the time trend, and it reflects the baseline risk for a reference period of July to December 2006.

### Interaction Terms

These models focused on main effects, and the final models included only four sets of preselected variable interactions: (1) sex by BSA; (2) sex by BSA squared; (3) age by reoperation; (4) age by emergent status. More extensive investigation for interactions was considered, including nonlinear, machine-learning approaches. However, the incremental value of such approaches remains uncertain [28], and interpretability can also become more problematic with numerous interaction terms.

Although multiple terms were allotted for modeling the main effects of age and reoperation, only a single degree of freedom was allotted for their interaction. The models defined a single variable interaction term for age and reoperation. It was equal to the patient's age minus 50 if the patient was at least 50 years old and had a previous CV surgery; otherwise it was equal to zero. This term represents the difference in the change of the slope of age at age 50 for patients who have had at least one previous CV surgery compared with patients who have not had a previous CV surgery. Similarly, only one degree of freedom was allotted for the interaction between age and status. The interaction represents the difference in the change of the slope of age at age 50 for patients with emergent or salvage status compared with patients with elective or urgent status. Although these interaction terms complicate the interpretation of other model variables, this was considered to be acceptable because the main focus of the analysis was prediction, not effect estimation.

## Results

### Model Performance

Table 4 presents the discrimination of each of the isolated CABG models as well as a comparison with the previous



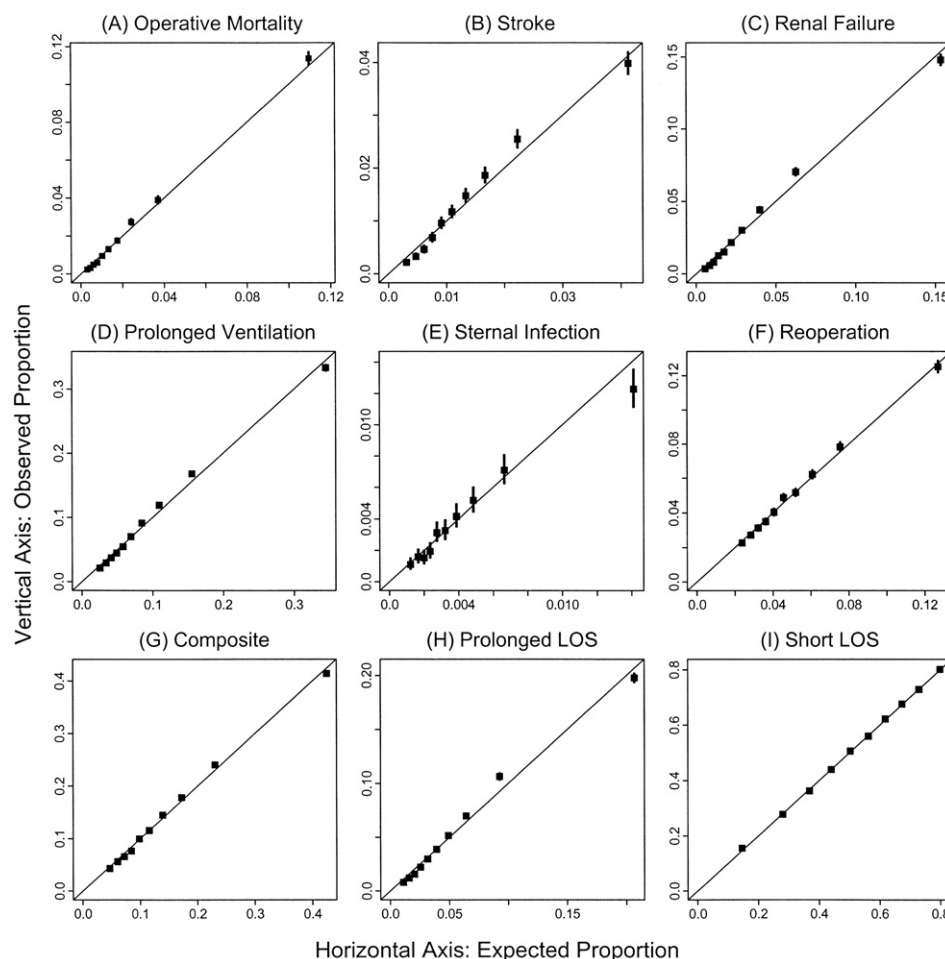


Fig 1. Plots of observed (O) versus expected (E) in validation sample

STS CABG risk models. For the new CABG models, discrimination ranged from 0.657 to 0.810 in the development sample and from 0.653 to 0.812 in the validation sample. The close agreement between c-indices from the development and validation samples reflects the large sample size and suggests that the models did not overfit the data. When the discrimination of the new and previous STS models were compared using the validation sample, the c-index of the new model was larger for each endpoint.

The Hosmer-Lemeshow test is not reported as an overall measure of calibration for these models because of its sensitivity to sample size. With samples as large as those used to develop these models, the null hypothesis will inevitably be proven false, given that all such models are only approximations [29]. As an alternative to such global measures of calibration, Figure 1 shows plots of observed versus expected event proportions within deciles of predicted risk for a variety of endpoints. For each endpoint, the absolute difference between the observed and expected proportions was less than 1.5% in each decile category. Additional analyses of model fit and discrimination are available online at [www.sts.org/riskmodels](http://www.sts.org/riskmodels).

### Final Models

After calculating these measures of model performance, the final regression coefficients were estimated from the combined training and validation samples. Odds ratios for each predictor variable and model endpoint are summarized in Table 5. “Not applicable” indicates that the specific predictor was not included in a particular risk model. These final models were estimated using generalized estimating equations with empirical (sandwich) standard error estimates to account for clustering of patients within institutions [30]. An independence working correlation matrix was used to apply the generalized estimating equations method. With this approach, the estimated regression coefficients were identical to those obtained using ordinary logistic regression, but the standard errors were adjusted to account for correlated observations within hospitals.

### Final Model Intercept and Coefficients

The Appendix contains the algorithm, intercept and coefficients for the final STS 2008 CABG risk models. The variance/covariance matrix is available on the web at [www.sts.org/riskmodels](http://www.sts.org/riskmodels). An on-line risk calculator is available at <http://209.220.160.181/STSTWebRiskCalc261/>.



Table 5. Estimated Odds Ratios for CABG Mortality, Morbidity, and Length of Stay Models

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50 (no reoperation, elective)	1.36 (1.24, 1.49)	1.78 (1.58, 1.99)	1.24 (1.16, 1.33)	1.06 (1.02, 1.10)	1.43 (1.23, 1.67)	1.14 (1.09, 1.19)	1.08 (1.05, 1.11)	1.35 (1.27, 1.43)	0.77 (0.75, 0.78)
Age 70 versus 50 (no reoperation, elective)	2.53 (2.31, 2.76)	2.43 (2.19, 2.71)	1.93 (1.81, 2.07)	1.42 (1.37, 1.47)	1.70 (1.47, 1.97)	1.45 (1.39, 1.51)	1.49 (1.44, 1.53)	2.17 (2.05, 2.29)	0.44 (0.43, 0.45)
Age 80 versus 50 (no reoperation, elective)	4.70 (4.29, 5.15)	3.34 (2.99, 3.72)	3.01 (2.80, 3.24)	1.90 (1.82, 1.99)	2.02 (1.73, 2.36)	1.85 (1.76, 1.94)	2.05 (1.98, 2.12)	3.48 (3.28, 3.69)	0.25 (0.24, 0.26)
BSA 1.6 versus 2.0 among females	1.26 (1.19, 1.32)	1.15 (1.08, 1.23)	0.84 (0.80, 0.89)	1.03 (1.00, 1.06)	0.49 (0.43, 0.57)	1.23 (1.18, 1.28)	1.03 (1.01, 1.06)	0.94 (0.90, 0.97)	1.17 (1.14, 1.20)
BSA 1.6 versus 2.0 among males	1.75 (1.64, 1.86)	1.19 (1.08, 1.31)	1.24 (1.17, 1.32)	1.40 (1.34, 1.46)	0.77 (0.63, 0.93)	1.40 (1.33, 1.46)	1.35 (1.30, 1.40)	1.43 (1.36, 1.50)	0.79 (0.77, 0.82)
BSA 1.8 versus 2.0 among females	1.02 (0.99, 1.05)	1.07 (1.03, 1.11)	0.86 (0.84, 0.88)	0.95 (0.94, 0.97)	0.67 (0.63, 0.71)	1.06 (1.04, 1.08)	0.96 (0.94, 0.97)	0.90 (0.88, 0.92)	1.14 (1.13, 1.16)
BSA 1.8 versus 2.0 among males	1.20 (1.17, 1.23)	1.09 (1.05, 1.13)	1.02 (1.00, 1.04)	1.09 (1.07, 1.10)	0.85 (0.79, 0.91)	1.13 (1.11, 1.15)	1.08 (1.07, 1.09)	1.10 (1.08, 1.12)	0.96 (0.95, 0.97)
BSA 2.2 versus 2.0 among females	1.20 (1.14, 1.27)	0.95 (0.88, 1.02)	1.32 (1.27, 1.37)	1.18 (1.15, 1.22)	1.62 (1.51, 1.74)	1.03 (1.00, 1.07)	1.19 (1.16, 1.21)	1.28 (1.24, 1.32)	0.78 (0.77, 0.80)
BSA 2.2 versus 2.0 among males	1.01 (0.99, 1.03)	0.92 (0.90, 0.95)	1.17 (1.15, 1.19)	1.10 (1.08, 1.11)	1.27 (1.22, 1.32)	0.97 (0.96, 0.99)	1.07 (1.06, 1.08)	1.08 (1.07, 1.10)	0.90 (0.89, 0.91)
Creatinine 1.5 versus 1.0	1.66 (1.57, 1.76)	1.39 (1.30, 1.49)	3.36 (3.16, 3.58)	1.56 (1.51, 1.62)	1.44 (1.28, 1.62)	1.33 (1.28, 1.38)	1.76 (1.70, 1.82)	1.65 (1.59, 1.72)	0.69 (0.67, 0.71)
Creatinine 2.0 versus 1.0	1.94 (1.84, 2.04)	1.49 (1.39, 1.58)	4.06 (3.83, 4.31)	1.73 (1.68, 1.79)	1.47 (1.30, 1.65)	1.44 (1.40, 1.49)	2.05 (1.98, 2.11)	1.92 (1.86, 2.00)	0.55 (0.53, 0.57)
Creatinine 2.5 versus 1.0	2.26 (2.14, 2.39)	1.59 (1.47, 1.71)	4.90 (4.61, 5.21)	1.92 (1.85, 1.99)	1.50 (1.30, 1.72)	1.57 (1.51, 1.64)	2.39 (2.30, 2.48)	2.24 (2.15, 2.34)	0.44 (0.42, 0.46)
Dialysis versus no dialysis and creatinine = 1.0	3.84 (3.54, 4.16)	1.67 (1.48, 1.88)	NA	2.85 (2.68, 3.03)	2.13 (1.78, 2.56)	1.86 (1.73, 2.00)	2.46 (2.33, 2.60)	2.80 (2.63, 2.98)	0.27 (0.25, 0.29)
EF per 10-unit decrease	1.19 (1.17, 1.22)	1.14 (1.11, 1.16)	1.08 (1.06, 1.10)	1.18 (1.16, 1.20)	1.11 (1.07, 1.16)	1.11 (1.09, 1.13)	1.16 (1.15, 1.18)	1.17 (1.15, 1.19)	0.84 (0.83, 0.85)
Preoperative atrial fibrillation	1.36 (1.28, 1.44)	1.21 (1.12, 1.30)	1.24 (1.18, 1.30)	1.20 (1.16, 1.24)	NA	1.26 (1.21, 1.31)	1.24 (1.21, 1.28)	1.42 (1.37, 1.48)	0.61 (0.59, 0.63)
CHF not NYHA IV	1.21 (1.15, 1.28)	NA	1.36 (1.30, 1.43)	1.31 (1.26, 1.35)	1.33 (1.19, 1.48)	1.16 (1.11, 1.21)	1.27 (1.23, 1.31)	1.43 (1.38, 1.48)	0.72 (0.70, 0.75)
CHF NYHA IV	1.39 (1.31, 1.47)	NA	1.35 (1.28, 1.42)	1.52 (1.45, 1.59)	1.45 (1.25, 1.67)	1.26 (1.20, 1.32)	1.48 (1.42, 1.54)	1.50 (1.44, 1.57)	0.65 (0.61, 0.68)
Chronic lung disease, mild	1.22 (1.16, 1.29)	NA	1.14 (1.08, 1.21)	1.36 (1.31, 1.41)	1.56 (1.40, 1.73)	1.11 (1.07, 1.15)	1.23 (1.19, 1.27)	1.34 (1.29, 1.39)	0.79 (0.76, 0.82)
Chronic lung disease, moderate	1.40 (1.32, 1.49)	NA	1.25 (1.18, 1.33)	1.65 (1.57, 1.73)	1.80 (1.58, 2.06)	1.20 (1.14, 1.26)	1.42 (1.36, 1.47)	1.65 (1.58, 1.73)	0.68 (0.65, 0.71)
Chronic lung disease, severe	2.35 (2.19, 2.52)	NA	1.66 (1.54, 1.79)	2.37 (2.24, 2.51)	2.40 (2.06, 2.79)	1.54 (1.44, 1.64)	1.98 (1.90, 2.07)	2.46 (2.34, 2.60)	0.48 (0.45, 0.51)
CVD with CVA	1.31 (1.24, 1.38)	2.09 (1.96, 2.22)	1.18 (1.12, 1.23)	1.35 (1.31, 1.39)	NA	1.21 (1.17, 1.26)	1.32 (1.29, 1.36)	1.45 (1.40, 1.51)	0.70 (0.68, 0.72)

Table 5. Continued

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
CVD without CVA	1.14 (1.08, 1.20)	1.65 (1.54, 1.75)	1.11 (1.06, 1.17)	1.15 (1.11, 1.18)	NA	1.12 (1.08, 1.17)	1.17 (1.14, 1.20)	1.14 (1.10, 1.18)	0.85 (0.81, 0.89)
Diabetes, insulin dependent	1.30 (1.24, 1.37)	1.19 (1.12, 1.27)	1.80 (1.72, 1.87)	1.22 (1.18, 1.26)	2.24 (2.02, 2.48)	1.14 (1.10, 1.18)	1.30 (1.27, 1.34)	1.59 (1.53, 1.64)	0.64 (0.62, 0.66)
Diabetes, noninsulin dependent	1.01 (0.97, 1.06)	1.16 (1.11, 1.22)	1.32 (1.28, 1.36)	1.04 (1.02, 1.07)	1.38 (1.27, 1.49)	0.98 (0.96, 1.01)	1.08 (1.06, 1.10)	1.15 (1.12, 1.17)	0.87 (0.86, 0.88)
Diseased vessels (2 versus 1, or 3 versus 2)	1.17 (1.12, 1.23)	1.35 (1.29, 1.42)	1.23 (1.19, 1.27)	1.19 (1.16, 1.22)	1.15 (1.07, 1.24)	1.07 (1.05, 1.10)	1.16 (1.14, 1.18)	1.15 (1.11, 1.18)	0.81 (0.80, 0.82)
Preoperative IABP/inotropes	1.41 (1.33, 1.49)	NA	1.43 (1.36, 1.51)	2.56 (2.42, 2.72)	NA	1.37 (1.31, 1.43)	1.96 (1.86, 2.06)	1.60 (1.53, 1.67)	0.60 (0.57, 0.63)
Shock	2.29 (2.12, 2.47)	1.38 (1.23, 1.55)	1.65 (1.54, 1.77)	2.08 (1.96, 2.21)	NA	1.43 (1.34, 1.52)	2.10 (1.99, 2.23)	1.73 (1.62, 1.84)	0.58 (0.54, 0.62)
Female versus male (at BSA = 1.8)	1.31 (1.25, 1.36)	1.32 (1.24, 1.39)	1.25 (1.21, 1.31)	1.33 (1.29, 1.36)	1.19 (1.06, 1.35)	0.90 (0.87, 0.93)	1.18 (1.15, 1.21)	1.24 (1.20, 1.28)	0.65 (0.63, 0.66)
Hypertension	NA	1.29 (1.22, 1.37)	1.25 (1.20, 1.30)	1.10 (1.08, 1.13)	NA	NA	1.12 (1.10, 1.15)	1.07 (1.04, 1.11)	0.92 (0.90, 0.94)
Immunosuppressive treatment	1.48 (1.37, 1.60)	NA	1.21 (1.12, 1.31)	1.11 (1.05, 1.18)	NA	1.32 (1.24, 1.41)	1.20 (1.14, 1.26)	1.28 (1.20, 1.37)	0.80 (0.76, 0.84)
Aortic insufficiency, moderate/severe	NA	NA	NA	NA	NA	NA	NA	NA	0.82 (0.75, 0.89)
Mitral insufficiency, moderate/severe	1.31 (1.21, 1.41)	NA	NA	1.12 (1.06, 1.18)	NA	1.24 (1.16, 1.32)	1.20 (1.15, 1.26)	1.15 (1.09, 1.22)	0.85 (0.80, 0.91)
Tricuspid insufficiency, moderate/severe	NA	NA	1.31 (1.17, 1.45)	1.28 (1.18, 1.39)	NA	NA	1.24 (1.16, 1.33)	NA	0.78 (0.71, 0.87)
PCI $\leq$ 6 hours	1.37 (1.24, 1.50)	NA	1.29 (1.16, 1.43)	1.21 (1.13, 1.29)	NA	1.30 (1.19, 1.42)	1.31 (1.23, 1.39)	1.17 (1.07, 1.27)	0.79 (0.74, 0.84)
Peripheral vascular disease	1.42 (1.36, 1.48)	1.32 (1.26, 1.39)	1.21 (1.17, 1.26)	1.22 (1.19, 1.26)	1.36 (1.24, 1.48)	1.24 (1.20, 1.28)	1.25 (1.22, 1.28)	1.31 (1.28, 1.35)	0.82 (0.81, 0.84)
Aortic stenosis	NA	NA	NA	1.18 (1.11, 1.26)	NA	NA	1.16 (1.10, 1.22)	1.15 (1.07, 1.23)	0.87 (0.82, 0.92)
Left main disease	NA	NA	NA	1.07 (1.04, 1.09)	NA	NA	1.04 (1.02, 1.06)	NA	NA
MI 1–21 days	1.37 (1.32, 1.44)	1.31 (1.25, 1.37)	1.27 (1.22, 1.32)	1.34 (1.29, 1.38)	NA	NA	1.23 (1.20, 1.25)	1.22 (1.18, 1.25)	0.88 (0.86, 0.90)
MI $>$ 6 and $<$ 24 hours	1.59 (1.46, 1.74)	1.59 (1.43, 1.76)	1.48 (1.36, 1.60)	1.59 (1.49, 1.68)	NA	NA	1.43 (1.37, 1.50)	1.31 (1.24, 1.39)	0.80 (0.76, 0.84)
MI $\leq$ 6 hours	1.70 (1.53, 1.89)	1.49 (1.31, 1.68)	1.43 (1.29, 1.57)	1.56 (1.45, 1.67)	NA	NA	1.44 (1.35, 1.53)	1.30 (1.21, 1.40)	0.82 (0.77, 0.87)
Time trend, per 6-month harvest interval	0.97 (0.97, 0.98)	0.97 (0.96, 0.98)	1.01 (1.00, 1.02)	1.01 (1.01, 1.02)	0.97 (0.95, 0.99)	0.99 (0.99, 1.00)	1.00 (1.00, 1.01)	1.00 (0.99, 1.01)	0.99 (0.98, 1.00)
Race Asian	NA	1.33 (1.14, 1.55)	1.08 (0.96, 1.22)	1.33 (1.21, 1.47)	1.00 (0.66, 1.51)	1.31 (1.17, 1.46)	1.23 (1.15, 1.31)	1.26 (1.13, 1.40)	0.70 (0.61, 0.81)
Race black	NA	1.41 (1.30, 1.54)	1.24 (1.16, 1.33)	1.37 (1.27, 1.48)	1.30 (1.13, 1.51)	1.21 (1.14, 1.30)	1.31 (1.24, 1.38)	1.43 (1.34, 1.51)	0.69 (0.65, 0.73)
Race Hispanic	NA	1.12 (0.98, 1.27)	1.24 (1.11, 1.39)	1.16 (1.07, 1.26)	1.30 (1.07, 1.58)	1.05 (0.97, 1.13)	1.12 (1.05, 1.19)	1.09 (0.99, 1.20)	0.85 (0.77, 0.94)

Table 5. Continued

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Reoperation, 1 previous operation <sup>a</sup>	3.13 (2.74, 3.57)	NA	1.52 (1.35, 1.71)	1.72 (1.58, 1.86)	NA	1.57 (1.43, 1.74)	1.61 (1.50, 1.72)	1.62 (1.47, 1.80)	0.72 (0.67, 0.78)
Reoperation, ≥ 2 previous operations <sup>a</sup>	4.19 (3.45, 5.09)	NA	1.58 (1.33, 1.87)	1.86 (1.62, 2.14)	NA	1.71 (1.44, 2.03)	1.84 (1.65, 2.05)	1.79 (1.53, 2.08)	0.64 (0.56, 0.73)
Status urgent <sup>a</sup>	1.16 (1.10, 1.22)	1.11 (1.06, 1.17)	1.12 (1.05, 1.19)	1.24 (1.18, 1.31)	1.20 (1.10, 1.32)	1.18 (1.13, 1.23)	1.18 (1.14, 1.22)	1.20 (1.15, 1.25)	0.86 (0.83, 0.90)
Status emergent, no resuscitation <sup>a</sup>	2.83 (2.52, 3.18)	2.12 (1.82, 2.48)	1.68 (1.49, 1.89)	2.14 (1.96, 2.34)	1.87 (1.46, 2.40)	1.83 (1.68, 1.99)	1.77 (1.64, 1.91)	2.12 (1.93, 2.32)	0.62 (0.58, 0.67)
Status emergent with resuscitation or salvage <sup>a</sup>	8.00 (6.91, 9.26)	2.51 (1.98, 3.18)	2.16 (1.82, 2.55)	3.01 (2.68, 3.38)	2.09 (1.45, 3.01)	2.34 (2.06, 2.65)	3.65 (3.26, 4.09)	2.39 (2.10, 2.72)	0.34 (0.30, 0.38)
Unstable angina	1.12 (1.07, 1.17)	NA	1.11 (1.05, 1.17)	1.05 (1.01, 1.10)	NA	NA	NA	NA	NA

<sup>a</sup> Variable interacts with age. Reported odds ratio represents effect of risk factor for patients aged 50 years old.

BSA = body surface area; CHF = congestive heart failure; Comp = composite adverse outcome (any); CVA = cerebrovascular accident, or stroke; CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; NA = not applicable; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; PLOS = prolonged length of stay; PVD = peripheral vascular disease; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

Previously, the STS risk models were completely upgraded every 3 years, with annual recalibration in the interim to assure that the benchmark O/E ratio is always 1. In the near future, annual upgrades of the models are planned.

## Limitations

Regardless of sample size or degree of statistical sophistication, all risk models are imperfect representations of reality. Although the STS risk models are based upon excellent clinical data and large sample sizes, there are some risk factors that are rare in the overall population but, when present, may be important predictors of outcome for specific patients. Some such variables, such as liver disease, are not included in the risk models, and the mortality risk for patients with these risk factors may be underestimated. Addition of a number of such variables will be considered at the next major specification upgrade.

There are other variables whose specifications undergo small but important changes over time, often in response to comments from STS database participants. These refinements are discussed on regular biweekly conference calls open to database participants, and suggested changes are regularly communicated to participants through a variety of means including FAQ's. With each major specification upgrade, they are incorporated into the new software specifications.

Audit is extremely important to assure the accuracy of any data registry. For the STS database and the risk models derived from it, robust audit is particularly critical as this registry is increasingly used for public reporting of outcomes and pay for performance. Studies suggest that the accuracy of the STS database is high for most important variables [31-35], although these audits are currently restricted to a limited number of sites annually because of budgetary constraints. In these audits, one of the most problematic variables has been 30-day mortality status (as opposed to in-hospital mortality). This is often a difficult endpoint to ascertain and may require more substantial investment of time and effort by participants, particularly for patients referred from outside their own institutions. Analysis of STS data suggests that approximately 90% of 30-day deaths occur in-hospital. Thus, if some patients recorded as being alive at 30 days have actually had their status ascertained only during the index hospitalization, the impact of this misclassification on the risk models should be negligible. This hypothesis was confirmed by comparing the odds ratios of all model variables for in-hospital versus 30-day mortality. Differences between the two were quite small, and these data are available on the web at [www.sts.org/riskmodels](http://www.sts.org/riskmodels). A new risk model for in-hospital mortality has been developed and placed on the same STS website. Furthermore, an aggressive program is in place to further enhance the accuracy of 30-day follow-up. In 2009, STS instituted a requirement that participants maintain documentation of the method by which they ascertained 30-day status, and that has become part of our routine audit. Linkage of the STS database with external death registries, such as the Social Security Death Master File, will

further support this capability. Finally, plans are being developed to expand the audit of certain key variables such as 30-day mortality to a significantly greater number of sites annually.

## Conclusions

Risk-adjustment models account for the effect of patient comorbidities on outcomes. STS risk models are based upon clinical data from the STS NCD, one of the oldest and largest of all specialty registries. The value of such clinical registries is particularly evident in today's health care environment, where accreditation, regulatory compliance, reimbursement, and referrals are increasingly based upon objective data. Organizations such as the AQA and the National Quality Forum that evaluate and endorse performance measures strongly advocate the use of risk-adjusted outcomes measures.

STS believes that clinical data are superior to those derived from administrative sources. Furthermore, given the substantial implications of risk-adjusted outcomes, we believe that all risk models used for profiling quality of care should be transparent to permit comprehensive peer review and to foster credibility among stakeholders.

We present a detailed exposition of the development and validation of the 2008 STS CABG risk model. This describes not only the statistical considerations but, just as importantly, the many clinical and pragmatic judgments that are always necessary in risk model development.

## References

1. Kouchoukos NT, Ebert PA, Grover FL, Lindesmith GG. Report of the Ad Hoc Committee on Risk Factors for Coronary Artery Bypass Surgery. *Ann Thorac Surg* 1988;45:348–9.
2. Clark RE. It is time for a national cardiothoracic surgical data base. *Ann Thorac Surg* 1989;48:755–6.
3. Edwards FH. Evolution of the Society of Thoracic Surgeons National Cardiac Surgery Database. *J Invasive Cardiol* 1998;10:485–8.
4. Grover FL, Shroyer AL, Hammermeister K, et al. A decade's experience with quality improvement in cardiac surgery using the Veterans Affairs and Society of Thoracic Surgeons national databases. *Ann Surg* 2001;234:464–72.
5. Shahian DM, Blackstone EH, Edwards FH, et al. Cardiac surgery risk models: a position article. *Ann Thorac Surg* 2004;78:1868–77.
6. Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: comprehensive review and statistical critique. *Ann Thorac Surg* 2001;72:2155–68.
7. Normand S-LT, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci* 2007;22:206–26.
8. Naftel DC. Do different investigators sometimes produce different multivariable equations from the same data? *J Thorac Cardiovasc Surg* 1994;107:1528–9.
9. Iezzoni LI. "Black box" medical information systems. A technology needing assessment. *JAMA* 1991;265:3006–7.
10. Shahian DM, Hutter MM, Torchiana DF, Iezzoni LI. Transparency: a mandatory requirement for risk models. *J Am Coll Surg* 2008;206:1240–2.
11. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association scientific statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. *Circulation* 2006;113:456–62.
12. Breiman L. Statistical modeling: the two cultures. *Stat Sci* 2001;16:199–231.
13. Harrell FE Jr. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag, 2001.
14. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics linear, logistic, survival, and repeated measures models. New York: Springer-Verlag, 2005.
15. Jones RH, Hannan EL, Hammermeister KE, et al. Identification of preoperative variables needed for risk adjustment of short-term mortality after coronary artery bypass graft surgery. The Working Group Panel on the Cooperative CABG Database Project. *J Am Coll Cardiol* 1996;28:1478–87.
16. Tu JV, Sykora K, Naylor CD. Assessing the outcomes of coronary artery bypass graft surgery: how many risk factors are enough? Steering Committee of the Cardiac Care Network of Ontario. *J Am Coll Cardiol* 1997;30:1317–23.
17. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928–35.
18. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
19. Cooper WA, O'Brien SM, Thourani VH, et al. Impact of renal dysfunction on outcomes of coronary artery bypass surgery: results from the Society of Thoracic Surgeons National Adult Cardiac Database. *Circulation* 2006;113:1063–70.
20. Grunkemeier GL, Zerr KJ, Jin R. Cardiac surgery report cards: making the grade. *Ann Thorac Surg* 2001;72:1845–8.
21. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9–13.
22. Hannan EL, Kilburn H Jr, O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. *JAMA* 1990;264:2768–74.
23. Grover FL, Johnson RR, Marshall G, Hammermeister KE. Factors predictive of operative mortality among coronary artery bypass subsets. *Ann Thorac Surg* 1993;56:1296–306.
24. Grover FL, Shroyer AL, Hammermeister KE. Calculating risk and outcome: the Veterans Affairs database. *Ann Thorac Surg* 1996;62(Suppl):6–11.
25. O'Connor GT, Plume SK, Olmstead EM, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. The Northern New England Cardiovascular Disease Study Group. *JAMA* 1991;266:803–9.
26. O'Connor GT, Plume SK, Olmstead EM, et al. Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Northern New England Cardiovascular Disease Study Group. *Circulation* 1992;85:2110–8.
27. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken: Wiley-Interscience, 2002.
28. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Stat Med* 2007;26:2937–57.
29. Marcin JP, Romano PS. Size matters to a model's fit. *Crit Care Med* 2007;35:2212–3.
30. Liang KY, Zeger SL. Longitudinal data-analysis using generalized linear models. *Biometrika* 1986;73:13–22.
31. Grover FL, Shroyer AL, Edwards FH, et al. Data quality review program: the Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg* 1996;62:1229–31.
32. Shroyer AL, Edwards FH, Grover FL. Updates to the Data Quality Review Program: the Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg* 1998;66:1494–7.
33. Herbert MA, Prince SL, Williams JL, Magee MJ, Mack MJ. Are unaudited records from an outcomes registry database accurate? *Ann Thorac Surg* 2004;77:1960–4.
34. Welke KF, Peterson ED, Vaughan-Sarrazin MS, et al. Comparison of cardiac surgery volumes and mortality rates

between the Society of Thoracic Surgeons and Medicare databases from 1993 through 2001. *Ann Thorac Surg* 2007;84:1538–46.

35. Welke KF, Ferguson TB Jr, Coombs LP, et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. *Ann Thorac Surg* 2004;77:1137–9.

## Appendix

### *Regression Coefficients and Variable Definitions for STS 2008 CABG Models*

For each endpoint, the formula for calculating a patient's predicted risk of the endpoint has the form:

$$\text{Predicted Risk} = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

where  $x_1, x_2, \dots, x_n$  denote patient preoperative risk factors (eg, quantitative variables such as age, and comorbidities coded as 1 = present, 0 = absent), and  $\beta_0, \beta_1, \dots, \beta_n$  denote regression coefficients (numerical constants). Regression coefficients for each endpoint are presented in [Appendix Table 1](#). The variables  $x_1, x_2, \dots, x_n$  are the same for each endpoint and are defined in [Appendix Table 2](#). The regression coefficient for the time trend is not presented. Instead, the intercept has been adjusted to incorporate the time trend. This adjusted intercept reflects the baseline risk for a reference period of July to December 2006.



Appendix Table 1. Regression Coefficients

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Intercept	-6.34090	-7.18174	-7.94605	-4.15175	-6.75378	-3.84861	-3.71671	-5.35975	2.84959
Atrial fibrillation	0.30830	0.18935	0.21351	0.17871	0.00000	0.23031	0.21565	0.35322	-0.49309
Age	-0.00259	0.00996	0.00678	0.00170	-0.00665	-0.00013	0.00247	0.00914	-0.01781
Age function 1	0.03325	0.04742	0.01496	0.00393	0.04270	0.01305	0.00515	0.02085	-0.00895
Age function 2	0.03140	-0.02582	0.02249	0.02366	-0.01895	0.01133	0.02441	0.01734	-0.02904
Age by reoperation function	-0.01714	-0.00098	-0.00291	-0.00459	0.00304	-0.00720	-0.00444	-0.00809	0.00449
Age by status function	-0.01366	-0.01363	-0.00022	-0.00106	-0.00352	-0.00435	0.00270	-0.00833	-0.00266
BSA function 1	-1.39342	-0.44041	-0.53672	-0.83950	0.65513	-0.83758	-0.75006	-0.89037	0.57952
BSA function 2	2.41303	0.06122	2.19879	2.15647	0.90025	1.16543	1.81770	2.15270	-1.83776
CHF but not NYHA IV	0.19229	0.00000	0.30971	0.26853	0.28272	0.14692	0.23695	0.35623	-0.32350
CHF and NYHA IV	0.32663	0.00000	0.30013	0.41599	0.36909	0.22846	0.39005	0.40757	-0.43827
Chronic lung disease mild	0.20273	0.00000	0.13488	0.30473	0.44371	0.10432	0.20878	0.29051	-0.23600
Chronic lung disease moderate	0.33843	0.00000	0.22530	0.50235	0.59021	0.18071	0.34720	0.50246	-0.39085
Chronic lung disease severe	0.85513	0.00000	0.50645	0.86175	0.87366	0.43034	0.68538	0.90211	-0.73862
Creatinine function 1	0.19353	0.02822	1.91934	-0.02712	-0.37465	0.01583	0.13361	-0.09060	0.00773
Creatinine function 2	0.82140	0.63174	0.50685	0.92120	1.09976	0.55107	0.99190	1.09571	-0.75781
Creatinine function 3	-0.70646	-0.52856	-2.04970	-0.68907	-0.68466	-0.39956	-0.81791	-0.70069	0.30449
CVD without prior CVA	0.13177	0.49807	0.10637	0.13792	0.00000	0.11403	0.15561	0.13271	-0.16385
CVD and prior CVA	0.26877	0.73600	0.16135	0.29946	0.00000	0.19208	0.28099	0.37248	-0.35706
Diabetes noninsulin dependent	0.01375	0.14992	0.27443	0.04283	0.31888	-0.01929	0.07453	0.13541	-0.13813
Diabetes insulin dependent	0.26312	0.17483	0.58581	0.19735	0.80627	0.12930	0.26525	0.46226	-0.44725
Dialysis	1.53777	0.54158	0.00000	1.01943	0.38312	0.63691	1.03466	0.93792	-1.30294
Ejection fraction function	0.01765	0.01274	0.00754	0.01669	0.01081	0.01063	0.01496	0.01542	-0.01756
Female	0.26801	0.27414	0.22704	0.28338	0.17792	-0.10270	0.16434	0.21488	-0.43658
Female by BSA function 1	0.82285	0.08974	0.96428	0.76954	1.11546	0.31901	0.66663	1.05623	-0.96846
Female by BSA function 2	0.05606	0.06490	-0.61086	-0.62558	0.17399	-0.02390	-0.25077	-0.35160	0.46088
Hypertension	0.00000	0.25718	0.22126	0.09930	0.00000	0.00000	0.11674	0.07200	-0.08155
IABP or inotropes	0.34193	0.00000	0.36023	0.94050	0.00000	0.31326	0.67253	0.47092	-0.51444
Immunosuppressive treatment	0.39159	0.00000	0.18881	0.10686	0.00000	0.27802	0.18030	0.24833	-0.22718
Insufficiency, aortic	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.19889
Insufficiency, mitral	0.26631	0.00000	0.00000	0.11169	0.00000	0.21170	0.18225	0.14174	-0.15962
Insufficiency, tricuspid	0.00000	0.00000	0.26729	0.24834	0.00000	0.00000	0.21893	0.00000	-0.24548
Left main disease	0.00000	0.00000	0.00000	0.06629	0.00000	0.00000	0.03570	0.00000	0.00000
MI 1 to 21 days	0.31810	0.27134	0.23962	0.28925	0.00000	0.00000	0.20524	0.19517	-0.12752
MI > 6 and < 24 hours	0.46614	0.46063	0.38917	0.46158	0.00000	0.00000	0.35859	0.27109	-0.22557
MI ≤ 6 hours	0.53242	0.39601	0.35421	0.44230	0.00000	0.00000	0.36337	0.26311	-0.19946
No. diseased vessel function	0.16120	0.30339	0.20729	0.17622	0.13869	0.06895	0.15075	0.13589	-0.21043
PCI ≤ 6 hours	0.31149	0.00000	0.25189	0.18695	0.00000	0.26256	0.26774	0.15633	-0.23860
Peripheral vascular disease	0.34951	0.27985	0.19308	0.20240	0.30529	0.21306	0.22277	0.27380	-0.19321
Race black	0.00000	0.34423	0.21696	0.31563	0.26572	0.19456	0.26634	0.35426	-0.37515
Race Hispanic	0.00000	0.11002	0.21645	0.14802	0.26330	0.04798	0.11289	0.08968	-0.16091
Race Asian	0.00000	0.28567	0.07579	0.28561	-0.00145	0.26855	0.20484	0.23064	-0.35049
Reop, 1 previous operation	1.13997	0.00000	0.41962	0.53987	0.00000	0.45372	0.47614	0.48534	-0.32375
Reop, ≥ 2 previous operations	1.43250	0.00000	0.45592	0.62211	0.00000	0.53695	0.61014	0.57945	-0.44745
Shock	0.82667	0.32434	0.50003	0.73290	0.00000	0.35800	0.74320	0.54575	-0.54475
Status urgent	0.14608	0.10671	0.11226	0.21738	0.18496	0.16500	0.16492	0.18202	-0.14608
Status emergent	1.04010	0.75216	0.51857	0.76090	0.62665	0.60549	0.56983	0.75083	-0.47745
Status salvage	2.07934	0.91950	0.76808	1.10085	0.73651	0.84873	1.29422	0.87072	-1.08265
Stenosis aortic	0.00000	0.00000	0.00000	0.16529	0.00000	0.00000	0.14706	0.13988	-0.14173
Unstable angina	0.11217	0.00000	0.10287	0.05060	0.00000	0.00000	0.00000	0.00000	0.00000

BSA = body surface area; CHF = congestive heart failure; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); CVD = cerebrovascular disease; DSWI = deep sternal wound infection; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

Appendix Table 2. Definition of Variables Appearing in STS 2008 CABG Models

Variable	Definition
Intercept	= 1 for all patients
Atrial fibrillation	= 1 if patient has history of preoperative atrial fibrillation, = 0 otherwise
Age	= Patient age in years
Age function 1	= max (age–50, 0)
Age function 2	= max (age–60, 0)
Age by reop function	= Age function 1 if surgery is a reoperation, = 0 otherwise
Age by status function	= Age function 1 if status is emergent or salvage, = 0 otherwise
BSA function 1	= max (1.4, min [2.6, BSA]) – 1.8
BSA function 2	= (BSA function 1) <sup>2</sup>
CHF but not NYHA IV	= 1 if patient has CHF and is not NYHA class IV, = 0 otherwise
CHF and NYHA IV	= 1 if patient has CHF and is NYHA class IV, = 0 otherwise
CLD mild	= 1 if patient has mild chronic lung disease, = 0 otherwise
CLD moderate	= 1 if patient has moderate chronic lung disease, = 0 otherwise
CLD severe	= 1 if patient has severe chronic lung disease, = 0 otherwise
Creatinine function 1	= max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise
Creatinine function 2	= max ([creatinine function 1] – 1.0, 0)
Creatinine function 3	= max ([creatinine function 1] – 1.5, 0)
CVD without prior CVA	= 1 if patient has history of CVD and no prior CVA, = 0 otherwise
CVD and prior CVA	= 1 if patient has history of CVD and a prior CVA, = 0 otherwise
Diabetes, noninsulin	= 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin	= 1 if patient has diabetes treated with insulin, = 0 otherwise
Dialysis	= 1 if patient requires dialysis preoperatively, = 0 otherwise
Ejection fraction function	= max (50 – ejection fraction, 0)
Female	= 1 if patient is female, = 0 otherwise
Female by BSA function 1	= BSA function 1 if female, = 0 otherwise
Female by BSA function 2	= BSA function 2 if female, = 0 otherwise
Hypertension	= 1 if patient has hypertension, = 0 otherwise
IABP or inotropes	= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
Immunosuppressive treatment	= 1 if patient given immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency, aortic	= 1 if patient has at least moderate aortic insufficiency, = 0 otherwise
Insufficiency, mitral	= 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency, tricuspid	= 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease	= 1 if patient has left main disease, = 0 otherwise
MI 1 to 21 days	= 1 if history of MI 1 to 21 days prior to surgery, = 0 otherwise
MI > 6 and < 24 hours	= 1 if history of MI >6 and <24 hours prior to surgery, = 0 otherwise
MI ≤ 6 hours	= 1 if history of MI ≤ 6 hours prior to surgery, = 0 otherwise
No. diseased vessel function	= 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise
PCI ≤ 6 hours	= 1 if patient had PCI ≤ 6 hours prior to surgery, = 0 otherwise
Peripheral vascular disease	= 1 if patient has peripheral vascular disease, = 0 otherwise
Race black	= 1 if patient is black, = 0 otherwise
Race Hispanic	= 1 if patient is nonblack Hispanic, = 0 otherwise
Race Asian	= 1 if patient is nonblack, non-Hispanic, and is Asian, = 0 otherwise
Reop, 1 previous operation	= 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, ≥ 2 previous operations	= 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Shock	= 1 if patient was in shock at time of procedure, = 0 otherwise
Status urgent	= 1 if status is urgent, = 0 otherwise
Status emergent	= 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage	= 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Stenosis aortic	= 1 if patient has aortic stenosis, = 0 otherwise
Unstable angina	= 1 if patient has unstable angina, no MI within 7 days of surgery, = 0 otherwise

BSA = body surface area; CHF = congestive heart failure; CLD = chronic lung disease; CVA = cerebrovascular accident, or stroke; CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; PLOS = prolonged length of stay; Reop = reoperation; Comp = composite adverse event (any); RF = renal failure; SLOS = short length of stay; STS = The Society of Thoracic Surgeons; Vent = prolonged ventilation.

# The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 2—Isolated Valve Surgery

Sean M. O'Brien, PhD,<sup>a</sup> David M. Shahian, MD,<sup>b</sup> Giovanni Filardo, PhD, MPH,<sup>c</sup> Victor A. Ferraris, MD,<sup>d</sup> Constance K. Haan, MD,<sup>e</sup> Jeffrey B. Rich, MD,<sup>f</sup> Sharon-Lise T. Normand, PhD,<sup>g</sup> Elizabeth R. DeLong, PhD,<sup>a</sup> Cynthia M. Shewan, PhD,<sup>h</sup> Rachel S. Dokholyan, MPH,<sup>a</sup> Eric D. Peterson, MD, MPH,<sup>a</sup> Fred H. Edwards, MD,<sup>e</sup> and Richard P. Anderson, MD<sup>i†</sup>

<sup>a</sup>Duke Clinical Research Institute, Durham, North Carolina; <sup>b</sup>Massachusetts General Hospital, Boston, Massachusetts; <sup>c</sup>Institute for Health Care Research and Improvement, Baylor Health Care System, Dallas, Texas; <sup>d</sup>University of Kentucky Chandler Medical Center, Division of Cardiovascular and Thoracic Surgery, Lexington, Kentucky; <sup>e</sup>University of Florida, Division of Cardiothoracic Surgery, Jacksonville, Florida; <sup>f</sup>Sentara Cardiovascular Research Institute, Norfolk, Virginia; <sup>g</sup>Department of Health Care Policy, Harvard Medical School, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; <sup>h</sup>The Society of Thoracic Surgeons, Chicago, Illinois; and <sup>i</sup>Seattle, Washington

**Background.** Adjustment for case-mix is essential when using observational data to compare surgical techniques or providers. That is most often accomplished through the use of risk models that account for preoperative patient factors that may impact outcomes. The Society of Thoracic Surgeons (STS) uses such risk models to create risk-adjusted performance reports for participants in the STS National Adult Cardiac Surgery Database (NCD). Although risk models were initially developed for coronary artery bypass surgery, similar models have now been developed for use with heart valve surgery, particularly as the proportion of such procedures has increased. The last published STS model for isolated valve surgery was based on data from 1994 to 1997 and did not include patients undergoing mitral valve repair. STS has developed new valve surgery models using contemporary data that include both valve repair as well as replacement. Expanding upon existing valve models, the new STS models include several nonfatal complications in addition to mortality.

**Methods.** Using STS data from 2002 to 2006, isolated valve surgery risk models were developed for operative mortality, permanent stroke, renal failure, prolonged ventilation (> 24 hours), deep sternal wound infection, reoperation for any reason, a major morbidity or mortality composite endpoint, prolonged postoperative length of stay, and short postoperative length of stay. The study population consisted of adult patients who underwent one of three types of valve surgery: isolated aortic valve replacement (n = 67,292), isolated mitral valve replacement (n = 21,229), or isolated mitral valve repair (n = 21,238). The

population was divided into a 60% development sample and a 40% validation sample. After an initial empirical investigation, the three surgery groups were combined into a single logistic regression model with numerous interactions to allow the covariate effects to differ across these groups. Variables were selected based on a combination of automated stepwise selection and expert panel review.

**Results.** Unadjusted operative mortality (in-hospital regardless of timing, and 30-day regardless of venue) for all isolated valve procedures was 3.4%, and unadjusted in-hospital morbidity rates ranged from 0.3% for deep sternal wound infection to 11.8% for prolonged ventilation. The number of predictors in each model ranged from 10 covariates in the sternal infection model to 24 covariates in the composite mortality plus morbidity model. Discrimination as measured by the c-index ranged from 0.639 for reoperation to 0.799 for mortality. When patients in the validation sample were grouped into 10 categories based on deciles of predicted risk, the average absolute difference between observed versus predicted events within these groups ranged from 0.06% for deep sternal wound infection to 1.06% for prolonged postoperative stay.

**Conclusions.** The new STS risk models for valve surgery include mitral valve repair as well as multiple endpoints other than mortality. Model coefficients are provided and an online risk calculator is publicly available from The Society of Thoracic Surgeons website.

(Ann Thorac Surg 2009;88:S23–42)

© 2009 by The Society of Thoracic Surgeons

Models for predicting surgical outcomes on the basis of patient preoperative characteristics are valuable tools for research, quality improvement, and clinical prac-

tice. Such models are used by The Society of Thoracic Surgeons (STS) to produce risk-adjusted performance re-

†This author is deceased. Former Chair, Quality, Research and Patient Safety Council, The Society of Thoracic Surgeons, Chicago, IL.

Address correspondence to Dr Shahian, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114; e-mail: dshahian@partners.org.

Drs O'Brien, Shahian, Filardo, Ferraris, Haan, Rich, Normand, DeLong, Shewan, Peterson, Edwards, Anderson, and Ms Dokholyan, have no conflicts of interest to declare regarding this work.

#### Abbreviations and Acronyms

AVR	= aortic valve replacement
CABG	= coronary artery bypass graft surgery
CI	= confidence interval
MI	= myocardial infarction
MVR	= mitral valve replacement
MVRepair	= mitral valve repair
NCD	= National Adult Cardiac Surgery Database
QMTF	= Quality Measurement Task Force
STS	= The Society for Thoracic Surgeons

ports for providers participating in the STS National Adult Cardiac Surgery Database (NCD). They are also used by STS surgeons and other physicians for counseling patients about the risk of surgery.

The earliest STS risk models were developed nearly 2 decades ago for isolated coronary artery bypass graft surgery (CABG). Subsequently, similar models have been developed for isolated valve replacement and combined CABG plus valve replacement. Because surgical practice and outcomes are changing rapidly, these models are updated periodically to reflect contemporary experience.

The last published STS model for isolated valve surgery was based on STS data from 1994 to 1997. The reference population included aortic and mitral valve replacements but excluded mitral valve repair, and the endpoint was operative mortality. In the decade since this model was published, many aspects of heart surgery have changed. First, as CABG volumes have decreased with the introduction of coronary stents, valve surgery as a proportion of overall heart surgery volume has increased in most practices. Between 2000 and 2006, the percentage of isolated CABG procedures decreased from 73% to 60% and the percentage of isolated valve procedures increased from 18% to 22%. Thus, in assessing provider performance, it is no longer sufficient only to consider isolated CABG surgery. Second, the frequency of mitral repair as a percentage of all isolated mitral operations in the STS NCD increased from 35% in 2000 to 53% in 2006. Third, during the same time period, the average mortality rate for isolated aortic or mitral surgery also decreased. Finally, efforts to measure and compare surgical performance have intensified and expanded. In addition to measuring operative mortality, performance reports increasingly focus on nonfatal complications as well as resource utilization and efficiency. Such outcomes have not historically been risk-adjusted for valve surgery.

The STS Quality Measurement Task Force (QMTF) has undertaken a complete revision of all STS risk models for adult cardiac surgery, and these new models were implemented in January 2008. This report, Part 2 of 3, describes the new STS models for isolated valve surgery (Part 1 describes the STS isolated CABG models, and Part 3 describes the models for CABG plus valve surgery). Authors of this report are the QMTF members who were involved in this initiative.

Two important features have been incorporated into these new models. First, the population includes mitral valve repair as well as aortic and mitral valve replacement. Second, in addition to operative mortality, the new models include six nonfatal in-hospital morbidity endpoints and two length-of-stay endpoints. In comparison with several other valve models that have recently been published [1–6], the STS models are distinguished by the large size of the development population and the broad spectrum of endpoints included.

#### Study Population and Endpoints

The population for this analysis consisted of operations on adult patients aged 20 to 100 years who underwent isolated single aortic or mitral valve surgery between January 1, 2002, and December 31, 2006. Only patients undergoing one of the following procedures were included: (1) isolated aortic valve replacement (AVR); (2) isolated mitral valve replacement (MVR); and (3) isolated mitral valve repair (MVRepair).

Because of the relatively small number of pulmonic, tricuspid, multiple valve procedures, and aortic repairs, these cases were not included in the current models. Patients undergoing concomitant CABG were excluded from the current analysis, but these were included in the separate STS valve plus CABG models described in Part 3 of this series. Records with missing data on sex ( $n = 44$ ) were excluded because missing sex is not allowed in the analysis dataset used for creating STS database participant feedback reports. This left a final study population of 109,759 patient operations performed at 809 STS NCD participating groups. Patients on dialysis preoperatively ( $n = 2,699$ ) were not included when developing the risk model for prediction of postoperative renal failure.

Patient characteristics in the study population are presented in Table 1.

#### Training and Validation Samples

The study population was randomly divided into a 60% training (development) sample and a 40% test (validation) sample. The development sample was used to identify predictor variables and estimate model coefficients. Data from the validation sample were used to assess model fit, discrimination, and calibration. After choosing variables and assessing model fit, the development and validation samples were subsequently combined, and the final model coefficients were estimated using the combined (development plus validation) data.

#### Endpoints

Risk models were developed for nine endpoints, identical to those in the STS CABG models. In contrast with the definition of operative mortality, which includes hospital deaths as well as deaths that occur after discharge within 30 days of surgery, the morbidity endpoints only include events that occurred before discharge. However, beginning with version 2.61, sternal infection data will be recorded for as long as 30 days postoperatively. The nine endpoints are as follows: (1) operative mortality: death during the same



Table 1. Distribution of Risk Factors in Overall Study Population Isolated Valve (2002–2006)

Variable	Overall Valve (n = 109,759)		AVR (n = 67,292)		MVR (n = 21,229)		MVRepair (n = 21,238)	
	N	%	N	%	N	%	N	%
<b>Demographics</b>								
Age, years								
< 55	28,147	25.6	13,227	19.66	6,601	31.09	8,319	39.17
55–64	23,258	21.2	12,987	19.30	4,833	22.77	5,438	25.61
65–74	28,145	25.6	18,299	27.19	5,294	24.94	4,552	21.43
≥75	30,209	27.5	22,779	33.85	4,501	21.20	2,929	13.79
Sex								
Male	60,752	55.4	39,209	58.27	9,055	42.65	12,488	58.80
Female	49,007	44.6	28,083	41.73	12,174	57.35	8,750	41.20
Race								
Caucasian	93,522	85.2	58,656	87.17	16,810	79.18	18,056	85.02
Black	7,630	7.0	3,555	5.28	2,383	11.23	1,692	7.97
Hispanic	3,680	3.4	2,344	3.48	889	4.19	447	2.10
Asian	1,538	1.4	719	1.07	437	2.06	382	1.80
Other	2,493	2.3	1,508	2.24	505	2.38	480	2.26
Missing	896	0.8	510	0.76	205	0.97	181	0.85
<b>Risk factors</b>								
Body surface area, m <sup>2</sup>								
< 1.50	4,351	4.0	2,341	3.48	1,234	5.81	776	3.65
1.50–1.74	24,577	22.4	13,713	20.38	6,151	28.97	4,713	22.19
1.75–1.99	40,548	36.9	24,744	36.77	7,914	37.28	7,890	37.15
≥ 2.00	39,517	36.0	26,007	38.65	5,768	27.17	7,742	36.45
Missing	766	0.7	487	0.72	162	0.76	117	0.55
Body mass index, kg/m <sup>2</sup>								
< 25	35,526	32.4	18,509	27.51	8,447	39.79	8,570	40.35
25–29	39,074	35.6	24,035	35.72	6,992	32.94	8,047	37.89
30–34	20,534	18.7	14,142	21.02	3,318	15.63	3,074	14.47
≥ 35	13,682	12.5	10,008	14.87	2,280	10.74	1,394	6.56
Missing	943	0.9	598	0.89	192	0.90	153	0.72
Diabetes mellitus								
No diabetes	88,709	80.8	52,052	77.35	17,535	82.60	19,122	90.04
Diabetes, noninsulin	14,900	13.6	11,026	16.39	2,412	11.36	1,462	6.88
Diabetes, insulin	5,788	5.3	3,974	5.91	1,216	5.73	598	2.82
Diabetes missing	138	0.1	91	0.14	34	0.16	13	0.06
Treatment missing	224	0.2	149	0.22	32	0.15	43	0.20
Hypertension								
No	41,649	37.9	22,338	33.20	8,859	41.73	10,452	49.21
Yes	67,886	61.9	44,816	66.60	12,326	58.06	10,744	50.59
Missing	224	0.2	138	0.21	44	0.21	42	0.20
Hypercholesterolemia								
No	59,003	53.8	33,156	49.27	12,857	60.56	12,990	61.16
Yes	50,328	45.9	33,865	50.33	8,286	39.03	8,177	38.50
Missing	428	0.4	271	0.40	86	0.41	71	0.33
Past or present smoker								
No	57,609	52.5	33,953	50.46	11,075	52.17	12,581	59.24
Yes	51,910	47.3	33,191	49.32	10,109	47.62	8,610	40.54
Missing	240	0.2	148	0.22	45	0.21	47	0.22
Chronic lung disease								
None	87,826	80.0	53,503	79.51	16,125	75.96	18,198	85.69
Mild	11,184	10.2	6,991	10.39	2,520	11.87	1,673	7.88
Moderate	6,346	5.8	4,022	5.98	1,494	7.04	830	3.91
Severe	3,332	3.0	2,110	3.14	853	4.02	369	1.74
Missing	1,071	1.0	666	0.99	237	1.12	168	0.79



Table 1. Continued

Variable	Overall Valve (n = 109,759)		AVR (n = 67,292)		MVR (n = 21,229)		MVRepair (n = 21,238)	
	N	%	N	%	N	%	N	%
Peripheral vascular disease								
No	101,129	92.1	61,222	90.98	19,550	92.09	20,357	95.85
Yes	8,381	7.6	5,909	8.78	1,641	7.73	831	3.91
Missing	249	0.2	161	0.24	38	0.18	50	0.24
Cerebrovascular disease								
No	96,852	88.2	58,983	87.65	18,158	85.53	19,711	92.81
Yes	12,661	11.5	8,147	12.11	3,033	14.29	1,481	6.97
Missing	246	0.2	162	0.24	38	0.18	46	0.22
CVA								
No CVA	101,631	92.6	62,518	92.91	18,833	88.71	20,280	95.49
Remote CVA (> 2 weeks)	6,926	6.3	4,203	6.25	1,912	9.01	811	3.82
Recent CVA (≤ 2 weeks)	818	0.7	325	0.48	409	1.93	84	0.40
CVA—missing timing	100	0.1	60	0.09	29	0.14	11	0.05
Missing	284	0.3	186	0.28	46	0.22	52	0.24
Endocarditis								
No endocarditis	100,998	92.0	63,257	94.00	17,926	84.44	19,815	93.30
Treated endocarditis	4,197	3.8	1,761	2.62	1,445	6.81	991	4.67
Active endocarditis	4,238	3.9	2,068	3.07	1,791	8.44	379	1.78
Endocarditis—missing type	63	0.1	30	0.04	27	0.13	6	0.03
Missing	263	0.2	176	0.26	40	0.19	47	0.22
Renal failure								
No	102,205	93.1	62,873	93.43	19,016	89.58	20,316	95.66
Yes	7,305	6.7	4,251	6.32	2,173	10.24	881	4.15
Missing	249	0.2	168	0.25	40	0.19	41	0.19
Renal function								
Creatinine < 1.00 mg/dL	42,028	38.3	25,679	38.16	7,754	36.53	8,595	40.47
Creatinine 1–1.49 mg/dL	51,939	47.3	32,058	47.64	9,372	44.15	10,509	49.48
Creatinine 1.50–1.99 mg/dL	8,081	7.4	5,078	7.55	1,875	8.83	1,128	5.31
Creatinine 2.00–2.49 mg/dL	1,946	1.8	1,192	1.77	512	2.41	242	1.14
Creatinine ≥ 2.50 mg/dL	1,294	1.2	750	1.11	390	1.84	154	0.73
Dialysis	2,699	2.5	1,464	2.18	900	4.24	335	1.58
Missing	1,772	1.6	1,071	1.59	426	2.01	275	1.29
Immunosuppressive treatment								
No	106,037	96.6	64,953	96.52	20,356	95.89	20,728	97.60
Yes	3,336	3.0	2,074	3.08	819	3.86	443	2.09
Missing	386	0.4	265	0.39	54	0.25	67	0.32
Previous CV interventions								
Previous coronary artery bypass surgery								
No	98,978	90.2	60,351	89.69	18,564	87.45	20,063	94.47
Yes	10,399	9.5	6,713	9.98	2,569	12.10	1,117	5.26
Missing	382	0.3	228	0.34	96	0.45	58	0.27
Previous valve surgery								
No	100,179	91.3	62,898	93.47	16,857	79.41	20,424	96.17
Yes	9,227	8.4	4,186	6.22	4,285	20.18	756	3.56
Missing	353	0.3	208	0.31	87	0.41	58	0.27
Previous other cardiac surgery								
No	105,686	96.3	65,084	96.72	20,034	94.37	20,568	96.85
Yes	3,662	3.3	1,975	2.93	1,077	5.07	610	2.87
Missing	411	0.4	233	0.35	118	0.56	60	0.28
Number of previous CV surgeries								
No prior CV surgery	91,196	83.1	56,629	84.15	15,239	71.78	19,328	91.01
1 prior CV surgery	15,399	14.0	9,122	13.56	4,775	22.49	1,502	7.07
2 or more prior CV surgeries	2,653	2.4	1,260	1.87	1,069	5.04	324	1.53
Missing	511	0.5	281	0.42	146	0.69	84	0.40

Table 1. Continued

Variable	Overall Valve (n = 109,759)		AVR (n = 67,292)		MVR (n = 21,229)		MVRepair (n = 21,238)	
	N	%	N	%	N	%	N	%
Prior PCI								
No PCI	101,878	92.8	62,145	92.35	19,573	92.20	20,160	94.92
PCI within 6 hours	122	0.1	58	0.09	51	0.24	13	0.06
PCI not within 6 hours	7,100	6.5	4,678	6.95	1,447	6.82	975	4.59
PCI-missing timing	133	0.1	90	0.13	28	0.13	15	0.07
Missing	526	0.5	321	0.48	130	0.61	75	0.35
Preoperative cardiac status								
Acuity status								
Elective	84,052	76.6	51,734	76.88	14,293	67.33	18,025	84.87
Urgent	23,795	21.7	14,670	21.80	6,071	28.60	3,054	14.38
Emergent	1,555	1.4	685	1.02	747	3.52	123	0.58
Emergent salvage	154	0.1	70	0.10	78	0.37	6	0.03
Missing	203	0.2	133	0.20	40	0.19	30	0.14
MI								
No prior MI	99,416	90.6	60,850	90.43	18,716	88.16	19,850	93.46
MI > 21 days	7,785	7.1	4,770	7.09	1,848	8.71	1,167	5.49
MI 8–21 days	719	0.7	480	0.71	170	0.80	69	0.32
MI 1–7 days	1,247	1.1	863	1.28	315	1.48	69	0.32
MI > 6 and < 24 hours	142	0.1	61	0.09	66	0.31	15	0.07
MI ≤ 6 hours	90	0.1	42	0.06	40	0.19	8	0.04
MI-missing timing	127	0.1	79	0.12	33	0.16	15	0.07
Missing	233	0.2	147	0.22	41	0.19	45	0.21
Angina								
No	85,364	77.8	49,573	73.67	17,598	82.90	18,193	85.66
Yes	24,164	22.0	17,577	26.12	3,591	16.92	2,996	14.11
Missing	231	0.2	142	0.21	40	0.19	49	0.23
Cardiogenic shock								
No	108,163	98.5	66,646	99.04	20,460	96.38	21,057	99.15
Yes	1,329	1.2	485	0.72	725	3.42	119	0.56
Missing	267	0.2	161	0.24	44	0.21	62	0.29
Resuscitation								
No	108,958	99.3	66,832	99.32	20,992	98.88	21,134	99.51
Yes	533	0.5	297	0.44	186	0.88	50	0.24
Missing	268	0.2	163	0.24	51	0.24	54	0.25
Arrhythmia								
No arrhythmia	89,779	81.8	57,451	85.38	14,604	68.79	17,724	83.45
AFib/flutter	16,124	14.7	7,569	11.25	5,721	26.95	2,834	13.34
Heart block	1,598	1.5	1,109	1.65	315	1.48	174	0.82
Sustained VT/VF	984	0.9	486	0.72	290	1.37	208	0.98
Arrhythmia—other	688	0.6	324	0.48	175	0.82	189	0.89
Arrhythmia-missing type	312	0.3	175	0.26	74	0.35	63	0.30
Missing	274	0.2	178	0.26	50	0.24	46	0.22
Preoperative IABP								
No	107,945	98.3	66,733	99.17	20,332	95.77	20,880	98.31
Yes	1,431	1.3	342	0.51	809	3.81	280	1.32
Missing	383	0.3	217	0.32	88	0.41	78	0.37
NYHA class								
I	17,413	15.9	10,222	15.19	2,706	12.75	4,485	21.12
II	32,360	29.5	20,295	30.16	4,915	23.15	7,150	33.67
III	40,321	36.7	25,483	37.87	8,205	38.65	6,633	31.23
IV	14,324	13.1	8,104	12.04	4,256	20.05	1,964	9.25
Missing	5,341	4.9	3,188	4.74	1,147	5.40	1,006	4.74

Table 1. Continued

Variable	Overall Valve (n = 109,759)		AVR (n = 67,292)		MVR (n = 21,229)		MVRepair (n = 21,238)	
	N	%	N	%	N	%	N	%
Congestive heart failure								
No	64,608	58.9	41,972	62.37	9,341	44.00	13,295	62.60
Yes	44,934	40.9	25,185	37.43	11,849	55.82	7,900	37.20
Missing	217	0.2	135	0.20	39	0.18	43	0.20
Number of diseased coronary vessels								
None	90,281	82.3	55,072	81.84	17,525	82.55	17,684	83.27
One	8,947	8.2	5,393	8.01	1,498	7.06	2,056	9.68
Two	3,386	3.1	2,180	3.24	735	3.46	471	2.22
Three	5,611	5.1	3,766	5.60	1,147	5.40	698	3.29
Missing	1,534	1.4	881	1.31	324	1.53	329	1.55
Left main disease $\geq$ 50%								
No	106,462	97.0	65,328	97.08	20,495	96.54	20,639	97.18
Yes	1,625	1.5	1,127	1.67	289	1.36	209	0.98
Missing	1,672	1.5	837	1.24	445	2.10	390	1.84
Ejection fraction, %								
< 25	2,694	2.5	1,774	2.64	341	1.61	579	2.73
25–34	5,900	5.4	3,810	5.66	1,052	4.96	1,038	4.89
35–44	10,035	9.1	6,181	9.19	2,208	10.40	1,646	7.75
45–54	20,481	18.7	12,411	18.44	4,382	20.64	3,688	17.37
$\geq$ 55	60,890	55.5	36,584	54.37	11,308	53.27	12,998	61.20
Missing	9,759	8.9	6,532	9.71	1,938	9.13	1,289	6.07
Aortic stenosis								
No	54,457	49.6	13,309	19.78	20,303	95.64	20,845	98.15
Yes	54,681	49.8	53,722	79.83	696	3.28	263	1.24
Missing	621	0.6	261	0.39	230	1.08	130	0.61
Mitral stenosis								
No	100,609	91.7	65,186	96.87	15,383	72.46	20,040	94.36
Yes	8,155	7.4	1,401	2.08	5,676	26.74	1,078	5.08
Missing	995	0.9	705	1.05	170	0.80	120	0.57
Tricuspid stenosis								
No	108,073	98.5	66,243	98.44	20,821	98.08	21,009	98.92
Yes	331	0.3	152	0.23	120	0.57	59	0.28
Missing	1,355	1.2	897	1.33	288	1.36	170	0.80
Pulmonic stenosis								
No	107,512	98.0	65,842	97.85	20,783	97.90	20,887	98.35
Yes	141	0.1	91	0.14	29	0.14	21	0.10
Missing	2,106	1.9	1,359	2.02	417	1.96	330	1.55
Aortic insufficiency								
None	59,905	54.6	25,861	38.43	16,701	78.67	17,343	81.66
Trivial	9,191	8.4	5,916	8.79	1,661	7.82	1,614	7.60
Mild	13,282	12.1	10,014	14.88	1,798	8.47	1,470	6.92
Moderate	9,501	8.7	8,815	13.10	382	1.80	304	1.43
Severe	15,722	14.3	15,529	23.08	109	0.51	84	0.40
Missing	2,158	2.0	1,157	1.72	578	2.72	423	1.99
Mitral insufficiency								
None	43,731	39.8	40,453	60.12	2,283	10.75	995	4.68
Trivial	7,743	7.1	7,285	10.83	388	1.83	70	0.33
Mild	14,455	13.2	13,066	19.42	1,089	5.13	300	1.41
Moderate	10,224	9.3	4,438	6.60	3,246	15.29	2,540	11.96
Severe	31,813	29.0	573	0.85	14,045	66.16	17,195	80.96
Missing	1,793	1.6	1,477	2.19	178	0.84	138	0.65

Table 1. Continued

Variable	Overall Valve (n = 109,759)		AVR (n = 67,292)		MVR (n = 21,229)		MVRepair (n = 21,238)	
	N	%	N	%	N	%	N	%
Tricuspid insufficiency								
None	78,472	71.5	49,976	74.27	14,266	67.20	14,230	67.00
Trivial	8,856	8.1	5,612	8.34	1,381	6.51	1,863	8.77
Mild	13,346	12.2	7,333	10.90	2,788	13.13	3,225	15.19
Moderate	5,167	4.7	2,126	3.16	1,753	8.26	1,288	6.06
Severe	974	0.9	297	0.44	460	2.17	217	1.02
Missing	2,944	2.7	1,948	2.89	581	2.74	415	1.95
Pulmonic insufficiency								
None	97,954	89.2	60,463	89.85	18,837	88.73	18,654	87.83
Trivial	4,161	3.8	2,370	3.52	779	3.67	1,012	4.77
Mild	2,541	2.3	1,340	1.99	573	2.70	628	2.96
Moderate	441	0.4	209	0.31	144	0.68	88	0.41
Severe	76	0.1	34	0.05	30	0.14	12	0.06
Missing	4,586	4.2	2,876	4.27	866	4.08	844	3.97

AFib = atrial fibrillation; AVR = aortic valve replacement; CV = cardiovascular; CVA = cerebrovascular accident (stroke); IABP = intra-aortic balloon pump; MI = myocardial infarction; MVR = mitral valve replacement; MVRepair = mitral valve repair; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; VF = ventricular fibrillation; VT = ventricular tachycardia.

hospitalization as surgery, regardless of timing, or within 30 days of surgery regardless of venue; (2) permanent stroke (cerebrovascular accident [CVA]): a central neurologic deficit persisting longer than 72 hours; (3) renal failure: a new requirement for dialysis or an increase of the serum creatinine to greater than 2.0 mg/dL and double the most recent preoperative creatinine level; (4) prolonged ventilation (longer than 24 hours); (5) deep sternal wound infection; (6) reoperation for any reason; (7) major morbidity or mortality: a composite defined as the occurrence of any of the above endpoints; (8) prolonged postoperative length of stay (PLOS): length of stay (LOS) more than 14 days (alive or

dead); and (9) short postoperative LOS (SLOS): LOS less than 6 days and patient alive at discharge.

Table 2 summarizes the endpoint frequencies in the study population.

### Single Versus Multiple Models

Two issues required particularly careful consideration: whether to construct separate models for the AVR and MVR populations, and how best to further subdivide the mitral population into repair versus replacement.

Because of the large size of the STS NCD, separate

Table 2. Frequency of Endpoints in Overall Study Population 2002 to 2006

	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
All isolated valve (AVR, MVR, MVRepair)									
N	109,759	109,759	107,060	109,759	109,759	109,759	109,759	109,759	109,759
Events	3,706	1,751	4,673	12,892	307	9,164	20,074	9,718	41,214
%	3.4	1.6	4.3	11.8	0.3	8.4	18.3	8.9	37.6
AVR									
N	67,292	67,292	65,828	67,292	67,292	67,292	67,292	67,292	67,292
Events	2,157	1,007	2,774	7,323	197	5,369	11,706	5,308	26,144
%	3.2	1.5	4.1	10.9	0.3	8.0	17.4	7.9	38.9
MVR									
N	21,229	21,229	20,329	21,229	21,229	21,229	21,229	21,229	21,229
Events	1,210	447	1,348	4,015	71	2,450	5,675	3,244	4,727
%	5.7	2.1	6.4	18.9	0.3	11.5	26.7	15.3	22.3
MVRepair									
N	21,238	21,238	20,903	21,238	21,238	21,238	21,238	21,238	21,238
Events	339	297	551	1,554	39	1,345	2,693	1,166	10,343
%	1.6	1.4	2.6	7.3	0.2	6.3	12.7	5.5	48.7

AVR = aortic valve replacement; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); DS WI = deep sternal wound infection; Mort = mortality; MVR = mitral valve replacement; MVRepair = mitral valve repair; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

models for AVR, MVR, and MVRrepair initially seemed both feasible and appropriate. However, because the endpoints of interest are rare events, we recognized the possibility that the number of such events would be too small to support reliable estimation of the model coefficients.

To assess this tradeoff, we conducted a pilot study to compare two alternative strategies for developing risk models for isolated valve surgery. The first strategy involved developing models separately for three subpopulations (AVR, MVR, and MVRrepair). The second strategy involved modeling all three subpopulations together in a single model; several interaction terms were included to allow the effect of selected risk factors to differ across the subpopulations. Both strategies were pilot tested by developing risk models for two endpoints: operative mortality and permanent stroke. These pilot models were developed in a 60% development sample and tested in a separate 40% validation sample. Each model was assessed by calculating the c-index and the generalized  $R^2$  index of Nagelkerke [7] in the validation sample for each combination of subpopulation and endpoint (3 subpopulations  $\times$  2 endpoints = 6 combinations). With the exception of AVR operative mortality, the combined model with interactions resulted in better discrimination. With the exception of MVR and MVRrepair operative mortality, the combined model also captured more variation as measured by the generalized  $R^2$  statistic.

Because the combined model strategy performed better in the majority of cases, and because a single combined model was consistent with the previous STS valve model, the combined model strategy was selected. To avoid assuming that the weighting of each risk factor was exactly constant across the three populations, we included interactions between surgery type and several key predictor variables. In principle, fitting a single model with several interactions is advantageous because it allows for pooling information across related groups without making an a priori assumption that all of the covariate effects are exactly constant across groups.

### Selection of Candidate Predictor Variables

Our general approach to variable selection is discussed in Part 1 of this series describing the development of the 2008 STS isolated CABG risk models. Briefly, we initially identified potential candidate variables by reviewing four versions of the STS data collection instrument (data versions 2.35, 2.41, 2.52.1, and 2.61) as well as previously published STS and similar cardiac risk models [1–6]. A panel of cardiac surgeons and health policy experts reviewed the initial variables for face validity and to be certain that no important predictor variables available in (or mappable to) STS NCD data version 2.61 had been excluded.

Final candidate explanatory variables and their coding are summarized in Table 3. The variables were identical to the CABG model candidate variables with the following differences: (1) percutaneous coronary intervention conducted within 6 hours or less of surgery was not a candidate variable because it was present in only 122 patients (0.1%) in the valve model population; (2) infec-

tious endocarditis was included. This risk factor was rarely present among isolated CABG patients (0.09%), but was not uncommon (7.7%) among patients undergoing valve surgery; (3) mitral stenosis was included; this risk factor was rarely present among isolated CABG patients (0.35%) but was common (7.4%) among patients undergoing valve surgery; and (4) an indicator for surgery type (AVR, MVR, MVRrepair) was included in the valve models.

### Coding of Explanatory Variables

The coding of continuous and categorical variables was identical to the CABG models, except for the following differences: (1) age was modeled as a linear spline truncated from below at 50 years and with a change of slope at 75; (2) creatinine was modeled as a linear term with values less than 0.5 and greater than 5.0 mapped to those values respectively (approximately the 1st and 99th percentiles of the empirical distribution); (3) previous myocardial infarction (MI) was modeled as three categories (< 24 hours, 1 to 21 days, and > 21 days or no MI); the first two categories were subsequently combined after expert panel review; (4) race was modeled as three categories: black, Hispanic, Caucasian/other; and (5) chronic lung disease was modeled as linear across four categories (none, mild, moderate, severe).

In general, these differences reflect a slightly simpler coding scheme (fewer parameters) for the valve models compared with the isolated CABG models.

### Repair Versus Replacement

In addition to a number of variables whose inclusion or coding were noted to be problematic during development of the 2008 STS isolated CABG models (Part 1 of this series), the approach to modeling mitral valve repair versus replacement was of some concern in the valve models. From a methodologic perspective, models used for risk-adjustment should include all patient preoperative risk factors that vary in prevalence between institutions and that substantially impact the probability of an adverse outcome. Such models should include variables that reflect the patient's baseline condition but should not include intraoperative events (eg, unexpected hemorrhage) or discretionary care processes (eg, use of a mechanical versus bioprosthetic valve). Adjusting for intraoperative events is not appropriate because these may be a reflection of the surgeon's performance. Adjusting for discretionary care processes may likewise mask differences in performance if the surgeon's choice of procedures has a substantial impact on outcomes. The same patient may receive valve repair if treated by one surgeon and replacement if treated by another. Adjusting for repair versus replacement will potentially conceal the outcomes of surgeons who achieve excellent results by repairing technically challenging valves that might otherwise be replaced if treated by a surgeon with less skill or tenacity. Importantly, there is considerable evidence to suggest the superiority of valve repair whenever feasible.

However, in addition to such discretionary factors, the decision to repair rather than replace the mitral valve is



Table 3. List of Final Candidate Variables and Their Coding for STS Valve Models

Candidate Variables	Coding
<b>Continuous variables</b>	
Age <sup>a</sup>	Linear spline truncated from below at 50 and with knot at 75
Ejection fraction	Linear, values > 50 mapped to 50
Body surface area <sup>a</sup>	Quadratic polynomial modeled separately for males and females. Note: body surface area < 1.4 and > 2.6 mapped to those values, respectively.
Creatinine	Linear (only for patients not on dialysis). Note: creatinine < 0.5 and > 5.0 mapped to those values, respectively.
Time trend <sup>a</sup>	Ordinal categorical variable with separate category for each 6-month harvest interval. Modeled as linear across categories.
<b>Binary variables</b>	
Active infectious endocarditis	Yes/no
Dialysis	Yes/no
Preoperative atrial fibrillation	Yes/no
Shock	Yes/no
Female <sup>a</sup>	Yes/no
Hypertension	Yes/no
Immunosuppressive treatment	Yes/no
Preoperative IABP or inotropes	Yes/no
Peripheral vascular disease	Yes/no
Unstable angina (no MI < 7 days)	Yes/no
Left main disease	Yes/no
Aortic stenosis	Yes/no
Mitral stenosis	Yes/no
Aortic insufficiency	Defined as at least moderate (yes/no)
Mitral insufficiency	Defined as at least moderate (yes/no)
Tricuspid insufficiency	Defined as at least moderate (yes/no)
<b>Categorical variables</b>	
Chronic lung disease	Modeled as linear across categories (none, mild, moderate, severe)
CVD/CVA	3 groups: no CVD, CVD no CVA, CVD + CVA
Diabetes mellitus	3 groups: insulin diabetes, noninsulin diabetes, other or no diabetes
Number diseased coronary vessels	3 groups: < 2, 2, 3. Modeled as linear across the categories
MI	3 groups: < 24 hr, 1–21 days, > 21 days or no MI (groups 1 and 2 were subsequently collapsed)
Race	3 groups: Black; Hispanic; Other including Caucasian
Status	4 groups: elective, urgent, emergent—no resuscitation, salvage or emergent with resuscitation
Previous cardiovascular operations	3 groups: 0 previous, 1 previous, ≥2 previous
CHF and NYHA class	3 groups: no CHF, CHF not NYHA IV, CHF+NYHA IV
Surgery type	3 groups: AVR, MVR, MVRRepair
<b>Interaction terms</b>	
Age by reoperation <sup>a</sup>	
Age by emergent status <sup>a</sup>	
Surgery type by each of the following:	Age, diabetes, dialysis, creatinine, reoperation, endocarditis, emergent status, CLD, CHF, EF, sex, shock, IABP/inotropes, mitral insufficiency, aortic insufficiency, mitral stenosis, aortic stenosis

<sup>a</sup> These variables were forced into each model.

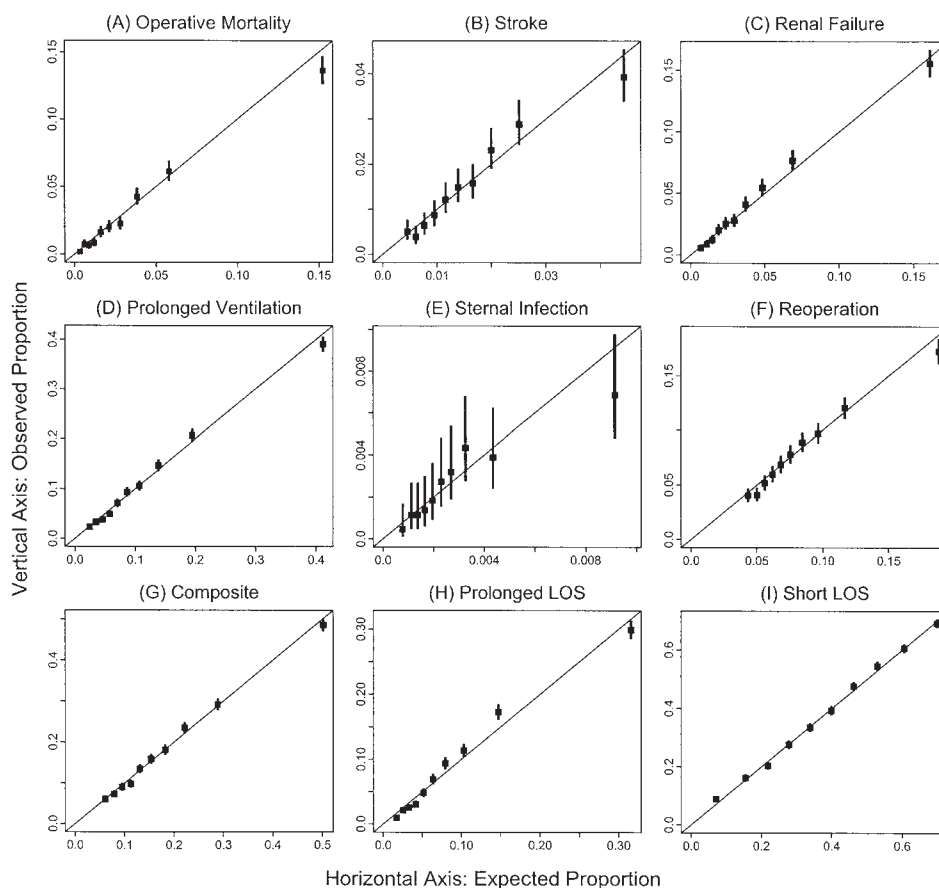
AVR = aortic valve replacement; CHF = congestive heart failure; CLD = chronic lung disease; CVA = cerebrovascular accident (stroke); CVD = cardiovascular disease; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; MVR = mitral valve replacement; MVRRepair = mitral valve repair; NYHA = New York Heart Association.

also dependent upon the patient's preoperative valve disease etiology, anatomy, and pathophysiology. On average, patients amenable to valve repair have less extensive valve pathology and a relatively favorable postoperative prognosis (the mortality rate for valve repair is 1.6%

compared with 5.7% for replacement). Ignoring these anatomical differences can introduce bias when comparing institutions, especially because these variables are not captured elsewhere on the STS data collection form.

A related difficulty in adjusting for repair versus re-

Fig 1. Plots of observed (O) versus expected (E) in validation sample



placement is that the former approach may sometimes be abandoned intraoperatively by the surgeon and converted to MVR. That may sometimes occur because of unforeseen technical problems that would prevent most surgeons from completing the repair, but in other instances, a more skilled surgeon might persist and achieve successful valve repair. Effectively separating these two scenarios is problematic from available data.

Ultimately, it was elected to include an indicator for mitral valve repair versus replacement in the valve risk models, consistent with the approach in a number of existing valve surgery models. We acknowledge that available data make it impossible to determine whether patient differences or surgical skill and judgment are the most important factors in determining between-provider variation in the proportion of valves repaired.

Recognizing the potential limitations of this modeling approach, the decision to adjust for repair versus replacement may be reassessed in future versions of the STS risk models. Beginning with data in version 2.61, the database will capture whether or not repair was attempted, and repair versus replacement may be analyzed based on an intention-to-treat principle.

### Missing Data

Model variables with more than 1% missing data in the study sample were ejection fraction (8.9%), NYHA class

(4.9%), tricuspid insufficiency (2.7%), aortic insufficiency (2.0%), mitral insufficiency (1.6%), left main disease (1.5%), creatinine/dialysis (1.6%), and number of diseased vessels (1.4%). The method of imputing missing data was identical to that employed in the isolated CABG models and described in Part 1 of this series. Briefly, binary risk factors were modeled as yes versus no or missing (ie, missing values were analyzed as if the endpoint did not occur). Missing data on categorical variables were imputed to the lowest risk value, typically the mode, and outcomes were typically similar for missing data and lowest risk patients. Missing data on continuous variables were imputed by grouping patients into strata and assigning the stratum-specific median value. For example, ejection fraction was imputed by grouping on sex and congestive heart failure and calculating the median ejection fraction among patients with nonmissing ejection fraction in each group.

Although multiple imputation is generally preferable to single imputation [8], single imputation was chosen for this analysis mainly because of practical considerations. Furthermore, because of the small fraction of missing data, the impact of single versus multiple imputation was considered to be inconsequential. Subsequent sensitivity analyses confirmed that the choice between single versus multiple imputation had little impact on the final regression coefficients, risk estimates, and confidence intervals. A summary of these sensitivity analyses, including coef-

Table 4. Discrimination of Models in Development and Validation Samples

	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Overall									
Development sample	0.805	0.694	0.782	0.770	0.704	0.643	0.721	0.770	0.738
Validation sample	0.799	0.691	0.762	0.762	0.659	0.639	0.718	0.773	0.734
AVR									
Development sample	0.779	0.679	0.766	0.748	0.710	0.630	0.698	0.752	0.713
Validation sample	0.759	0.689	0.749	0.736	0.637	0.619	0.694	0.759	0.713
MVR									
Development sample	0.794	0.679	0.767	0.772	0.591	0.642	0.735	0.748	0.726
Validation sample	0.802	0.702	0.748	0.772	0.656	0.634	0.738	0.729	0.710
MVRepair									
Development sample	0.855	0.736	0.813	0.765	0.774	0.616	0.703	0.777	0.733
Validation sample	0.844	0.672	0.788	0.773	0.714	0.646	0.712	0.800	0.725

AVR = aortic valve replacement; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); DS WI = deep sternal wound infection; Mort = mortality; MVR = mitral valve replacement; MVRepair = mitral valve repair; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

ficients and covariance matrices, is available at [www.sts.org/riskmodels](http://www.sts.org/riskmodels).

### Final Variable Selection Procedure

Variables were initially selected using an automated stepwise model selection algorithm. The stepwise procedure began with a model that included all of the final candidate variables except for interaction terms. Age, sex, body surface area, and month of surgery were forced into each model. Other variables were selected in a stepwise fashion using a significance criterion of 0.05 for entry and removal. This criterion was less stringent than that employed in development of the CABG models, because the sample size in the former was so much larger than that which was used for the valve models. The stepwise procedure was performed separately for each endpoint. The results were then reviewed by an expert panel of surgeons, and the following changes were made based on their feedback: (1) “MI less than 24 hours” and “MI 1 to 21 days” were collapsed into a single category; (2) preoperative atrial fibrillation was forced into the model for stroke (CVA); and (3) an indicator variable for dialysis was forced into any model that included creatinine level.

### Interaction Terms

In addition to including main effects, we tested the interaction between surgery group (AVR, MVR, MVRepair) and each of the following variables: age, diabetes mellitus, dialysis, creatinine, reoperation, endocarditis, emergent status, chronic lung disease, congestive heart failure, ejection fraction, sex, shock, intra-aortic balloon pump/inotropes, mitral insufficiency, aortic insufficiency, mitral stenosis, and aortic stenosis. These interaction terms allowed the effect of these selected risk factors to differ across the surgery populations.

Four additional sets of interactions were also included in the models: (1) sex by body surface area (BSA); (2) sex by BSA<sup>2</sup>; (3) age by reoperation; and (4) age by emergent status. These interaction terms were preselected and were

not tested as part of the backward selection algorithm. Additional technical details are provided in the Appendix. For reasons described in Part 1 of this series (isolated CABG risk models), an extensive automated search for additional interaction terms was not conducted.

### Adjustment for Time Trends

Surgery date was included in each model to adjust for changes in the frequency of adverse outcomes over the 5-year study period. Although surgery date is not itself a variable of interest, we adjusted for it to reduce potential confounding by time trends when estimating regression coefficients for the variables that are of primary interest (ie, patient preoperative risk factors). An example is provided in Part 1 of this series.

Surgery date was categorized into 6-month intervals (corresponding to the biannual STS data harvests) and modeled as a linear trend across the ordinal categories. Because it is a nuisance variable, surgery date is not included in the final risk prediction algorithm. Thus, a patient's predicted risk does not depend on the patient's surgery date. As described in the Appendix, the published intercept parameter has been adjusted to incorporate the time trend. The adjusted intercept reflects the baseline risk for a reference period of July to December 2006.

## Results

### Assessment of Model Fit and Discrimination

Because of the relatively large size of our sample, the Hosmer-Lemeshow test is uninformative and would invariably result in a significant *p* value [9]. As an alternative, model fit was assessed graphically by plotting observed versus predicted rates of each endpoint across deciles of predicted risk in the development and validation samples. This was done in the overall population and in subgroups based on surgery type (AVR, MVR, MVRepair); age (< 60, 60 to 79, ≥ 80 years); sex (male, female); diabetes mellitus (yes/no); status (elective, nonelective); and ejection fraction

Table 5. Odds Ratios (95% Confidence Intervals) for the Final Selected Models

A. Odds ratios for variables that do not interact with surgery group									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Preoperative AFib	1.20 (1.10, 1.31)	1.06 (0.93, 1.20)	NA	1.18 (1.11, 1.25)	NA	1.11 (1.04, 1.18)	1.12 (1.07, 1.18)	1.17 (1.10, 1.24)	0.74 (0.70, 0.78)
BSA 1.6 versus 2.0 among females	1.19 (1.09, 1.30)	1.18 (1.03, 1.35)	0.95 (0.87, 1.04)	1.15 (1.08, 1.22)	0.42 (0.27, 0.68)	1.26 (1.18, 1.34)	1.17 (1.12, 1.23)	1.11 (1.04, 1.17)	0.99 (0.95, 1.04)
BSA 1.6 versus 2.0 among males	1.75 (1.48, 2.07)	1.17 (0.92, 1.47)	1.33 (1.12, 1.58)	1.56 (1.41, 1.74)	0.94 (0.49, 1.84)	1.34 (1.21, 1.49)	1.44 (1.33, 1.57)	1.39 (1.25, 1.56)	0.73 (0.68, 0.79)
BSA 1.8 versus 2.0 among females	0.99 (0.95, 1.04)	1.08 (0.99, 1.17)	0.90 (0.86, 0.94)	1.00 (0.97, 1.03)	0.65 (0.54, 0.77)	1.07 (1.03, 1.11)	1.02 (0.99, 1.04)	0.99 (0.96, 1.02)	1.05 (1.03, 1.08)
BSA 1.8 versus 2.0 among males	1.21 (1.14, 1.29)	1.07 (0.98, 1.16)	1.07 (1.00, 1.14)	1.14 (1.10, 1.19)	0.90 (0.70, 1.14)	1.12 (1.08, 1.16)	1.12 (1.09, 1.16)	1.10 (1.06, 1.15)	0.92 (0.89, 0.94)
BSA 2.2 versus 2.0 among females	1.21 (1.11, 1.33)	0.94 (0.80, 1.10)	1.30 (1.21, 1.41)	1.15 (1.09, 1.21)	1.57 (1.26, 1.96)	1.02 (0.95, 1.09)	1.12 (1.07, 1.16)	1.14 (1.08, 1.21)	0.85 (0.81, 0.88)
BSA 2.2 versus 2.0 among males	0.98 (0.93, 1.03)	0.95 (0.88, 1.03)	1.09 (1.05, 1.14)	1.05 (1.02, 1.08)	1.32 (1.17, 1.48)	0.95 (0.93, 0.98)	1.02 (0.99, 1.04)	1.03 (1.00, 1.07)	0.94 (0.93, 0.96)
Creatinine per 1 unit	1.55 (1.46, 1.64)	1.34 (1.22, 1.47)	2.04 (1.93, 2.16)	1.58 (1.51, 1.65)	NA	1.27 (1.20, 1.33)	1.64 (1.57, 1.71)	1.58 (1.51, 1.65)	0.64 (0.61, 0.68)
CVD with CVA	NA	1.81 (1.56, 2.10)	1.22 (1.09, 1.37)	1.28 (1.18, 1.38)	NA	1.14 (1.05, 1.24)	1.20 (1.12, 1.28)	1.40 (1.29, 1.52)	0.77 (0.72, 0.83)
CVD without CVA	NA	1.32 (1.11, 1.57)	1.23 (1.10, 1.37)	1.14 (1.05, 1.23)	NA	1.06 (0.96, 1.17)	1.08 (1.01, 1.15)	NA	0.80 (0.73, 0.88)
No. diseased coronary vessels (2 versus 1 or 3 versus 2)	NA	1.10 (1.01, 1.20)	NA	1.07 (1.02, 1.11)	NA	NA	1.04 (1.00, 1.08)	1.03 (0.98, 1.08)	0.90 (0.86, 0.94)
EF per 10-unit decrease	1.09 (1.05, 1.14)	NA	1.04 (1.00, 1.09)	1.12 (1.09, 1.15)	1.26 (1.12, 1.41)	1.08 (1.04, 1.11)	1.10 (1.07, 1.12)	1.12 (1.08, 1.15)	0.87 (0.85, 0.90)
Hypertension	1.12 (1.03, 1.22)	1.19 (1.07, 1.33)	1.35 (1.25, 1.45)	1.11 (1.06, 1.17)	NA	NA	1.11 (1.07, 1.15)	NA	0.94 (0.91, 0.97)
Immunosuppressive treatment	1.42 (1.21, 1.67)	NA	1.39 (1.19, 1.62)	NA	NA	NA	1.16 (1.06, 1.27)	1.31 (1.17, 1.47)	NA
Left main disease	1.19 (0.98, 1.46)	NA	1.19 (0.98, 1.44)	NA	2.17 (1.13, 4.16)	NA	NA	NA	NA
Active infectious endocarditis	1.95 (1.68, 2.27)	1.87 (1.52, 2.29)	2.17 (1.88, 2.50)	2.15 (1.95, 2.36)	NA	1.55 (1.39, 1.73)	1.97 (1.80, 2.15)	2.79 (2.51, 3.09)	0.34 (0.30, 0.38)
Mitral insufficiency, moderate/severe	NA	1.26 (1.14, 1.39)	NA	NA	NA	NA	NA	NA	NA
Tricuspid insufficiency, moderate/severe	NA	NA	1.14 (1.01, 1.29)	1.14 (1.04, 1.25)	NA	1.09 (1.00, 1.20)	1.21 (1.12, 1.30)	1.17 (1.05, 1.31)	0.82 (0.73, 0.92)
Peripheral vascular disease	1.25 (1.12, 1.38)	1.29 (1.11, 1.49)	NA	NA	NA	1.22 (1.12, 1.32)	1.14 (1.07, 1.21)	1.17 (1.09, 1.25)	0.83 (0.78, 0.88)
Aortic stenosis		NA	NA	0.90 (0.83, 0.97)	NA	0.90 (0.84, 0.96)	0.93 (0.87, 0.98)	0.86 (0.80, 0.92)	1.07 (1.02, 1.13)
Mitral stenosis	1.24 (1.08, 1.41)	NA	NA	NA	NA	NA	NA	NA	NA
MI $\leq$ 21 days	1.14 (0.98, 1.34)	NA	NA	1.37 (1.22, 1.55)	NA	1.04 (0.91, 1.18)	1.28 (1.16, 1.41)	1.21 (1.06, 1.37)	0.81 (0.72, 0.91)
Time trend, per 6-month harvest interval	0.98 (0.97, 0.99)	0.98 (0.96, 1.00)	1.01 (0.99, 1.02)	1.02 (1.01, 1.03)	0.97 (0.93, 1.01)	1.00 (0.99, 1.01)	1.01 (1.00, 1.02)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Race black	NA	1.33 (1.13, 1.57)	1.51 (1.34, 1.69)	1.42 (1.27, 1.58)	NA	1.27 (1.15, 1.40)	1.37 (1.27, 1.49)	1.45 (1.31, 1.60)	0.64 (0.59, 0.70)
Race Hispanic	NA	0.87 (0.64, 1.19)	1.16 (0.97, 1.38)	1.07 (0.94, 1.22)	NA	1.14 (1.00, 1.30)	1.09 (0.98, 1.22)	1.16 (0.98, 1.38)	0.82 (0.72, 0.93)
Status urgent	1.29 (1.19, 1.40)	NA	1.21 (1.11, 1.33)	1.29 (1.20, 1.39)	NA	1.17 (1.10, 1.25)	1.22 (1.15, 1.29)	1.42 (1.33, 1.51)	0.70 (0.66, 0.74)
Unstable angina	1.21 (1.04, 1.41)	NA	NA	NA	NA	NA	NA	NA	NA

Table 5. Continued

B. Odds ratios for aortic valve replacement									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50	1.43 (1.34, 1.52)	1.48 (1.38, 1.59)	1.38 (1.30, 1.47)	1.31 (1.26, 1.36)	1.52 (1.31, 1.76)	1.16 (1.12, 1.21)	1.23 (1.19, 1.26)	1.31 (1.25, 1.37)	0.75 (0.73, 0.77)
Age 70 versus 50	2.04 (1.79, 2.32)	2.19 (1.90, 2.52)	1.90 (1.68, 2.16)	1.71 (1.59, 1.84)	2.31 (1.72, 3.10)	1.35 (1.25, 1.46)	1.50 (1.42, 1.59)	1.71 (1.55, 1.87)	0.57 (0.54, 0.60)
Age 80 versus 50	3.34 (2.84, 3.93)	3.21 (2.70, 3.81)	2.88 (2.46, 3.37)	2.31 (2.12, 2.52)	2.73 (1.95, 3.80)	1.59 (1.44, 1.76)	1.97 (1.82, 2.12)	2.50 (2.24, 2.79)	0.34 (0.32, 0.36)
CHF, not NYHA IV	1.29 (1.18, 1.42)	NA	1.24 (1.14, 1.34)	1.33 (1.24, 1.43)	NA	NA	1.20 (1.13, 1.27)	1.25 (1.17, 1.34)	0.86 (0.81, 0.91)
CHF, NYHA IV	1.83 (1.62, 2.07)	NA	1.61 (1.44, 1.81)	1.92 (1.77, 2.08)	NA	1.25 (1.17, 1.35)	1.62 (1.51, 1.73)	1.54 (1.40, 1.68)	0.72 (0.65, 0.79)
Diabetes, insulin	1.62 (1.43, 1.83)	NA	1.91 (1.70, 2.14)	1.42 (1.31, 1.55)	1.56 (1.05, 2.31)	1.20 (1.10, 1.31)	1.39 (1.29, 1.50)	1.68 (1.55, 1.83)	0.64 (0.59, 0.69)
Diabetes, noninsulin	1.27 (1.15, 1.39)	NA	1.45 (1.34, 1.57)	1.12 (1.04, 1.20)	NA	NA	1.12 (1.06, 1.18)	1.22 (1.15, 1.30)	0.85 (0.81, 0.88)
Dialysis versus no dialysis and creatinine = 1.0	2.85 (2.35, 3.45)	1.65 (1.34, 2.03)	NA	3.07 (2.74, 3.43)	NA	1.79 (1.60, 2.01)	2.42 (2.21, 2.66)	2.94 (2.64, 3.27)	0.29 (0.24, 0.34)
Preoperative IABP/ inotropes	1.47 (1.26, 1.71)	NA	1.34 (1.15, 1.57)	1.78 (1.55, 2.05)	1.69 (1.08, 2.65)	1.14 (1.02, 1.29)	1.75 (1.59, 1.94)	1.46 (1.30, 1.63)	0.56 (0.48, 0.66)
Shock	1.62 (1.29, 2.03)	1.65 (1.21, 2.25)	NA	2.09 (1.77, 2.47)	NA	1.32 (1.11, 1.58)	2.11 (1.80, 2.49)	1.74 (1.37, 2.21)	NA
Female versus male (at BSA = 1.8)	1.23 (1.10, 1.36)	1.25 (1.09, 1.43)	0.97 (0.88, 1.07)	1.29 (1.21, 1.38)	0.98 (0.72, 1.33)	0.86 (0.81, 0.93)	1.03 (0.98, 1.08)	1.25 (1.16, 1.35)	0.69 (0.66, 0.73)
CLD (moderate versus mild, or severe versus moderate)	1.27 (1.21, 1.33)	NA	1.18 (1.13, 1.23)	1.26 (1.22, 1.30)	1.27 (1.13, 1.42)	1.09 (1.06, 1.12)	1.17 (1.14, 1.20)	1.29 (1.24, 1.34)	0.81 (0.79, 0.83)
Reoperation, 1 previous operation <sup>a</sup>	2.11 (1.78, 2.49)	2.09 (1.64, 2.65)	1.55 (1.31, 1.84)	1.83 (1.64, 2.05)	NA	1.31 (1.16, 1.49)	1.55 (1.42, 1.70)	1.42 (1.27, 1.59)	0.67 (0.62, 0.72)
Reoperation, ≥ 2 previous operations <sup>a</sup>	2.48 (1.99, 3.08)	2.36 (1.76, 3.16)	1.66 (1.33, 2.07)	2.49 (2.14, 2.90)	NA	1.41 (1.19, 1.67)	1.96 (1.73, 2.22)	1.76 (1.52, 2.03)	0.50 (0.43, 0.58)
Status emergent, no resuscitation <sup>a</sup>	3.77 (2.75, 5.16)	2.78 (1.85, 4.17)	3.10 (2.21, 4.35)	4.54 (3.54, 5.83)	NA	1.63 (1.31, 2.03)	3.23 (2.66, 3.93)	2.45 (2.02, 2.97)	0.33 (0.25, 0.42)
Status emergent, with resuscitation or salvage <sup>a</sup>	7.94 (5.40, 11.66)	2.11 (1.06, 4.19)	3.47 (2.19, 5.51)	3.50 (2.41, 5.08)	NA	NA	3.38 (2.36, 4.84)	NA	0.32 (0.19, 0.54)



Table 5. Continued

C. Odds ratios for mitral valve replacement									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50	1.65 (1.53, 1.78)	1.48 (1.38, 1.59)	1.35 (1.26, 1.44)	1.31 (1.26, 1.36)	1.52 (1.31, 1.76)	1.25 (1.19, 1.31)	1.33 (1.29, 1.39)	1.26 (1.21, 1.33)	0.71 (0.68, 0.74)
Age 70 versus 50	2.71 (2.33, 3.17)	2.19 (1.90, 2.52)	1.81 (1.60, 2.06)	1.71 (1.59, 1.84)	2.31 (1.72, 3.10)	1.56 (1.42, 1.71)	1.78 (1.65, 1.92)	1.60 (1.45, 1.76)	0.50 (0.46, 0.55)
Age 80 versus 50	5.14 (4.15, 6.37)	3.21 (2.70, 3.81)	2.67 (2.23, 3.20)	2.31 (2.12, 2.52)	2.73 (1.95, 3.80)	1.97 (1.72, 2.26)	2.54 (2.27, 2.84)	2.27 (2.00, 2.58)	0.28 (0.25, 0.32)
CHF, not NYHA IV	1.29 (1.18, 1.42)	NA	1.24 (1.14, 1.34)	1.19 (1.07, 1.32)	NA	NA	1.11 (1.01, 1.21)	1.25 (1.17, 1.34)	0.96 (0.87, 1.06)
CHF, NYHA IV	1.83 (1.62, 2.07)	NA	1.61 (1.44, 1.81)	1.72 (1.55, 1.91)	NA	1.25 (1.17, 1.35)	1.49 (1.36, 1.64)	1.54 (1.40, 1.68)	0.80 (0.71, 0.91)
Diabetes, insulin	1.62 (1.43, 1.83)	NA	1.91 (1.70, 2.14)	1.66 (1.47, 1.86)	1.56 (1.05, 2.31)	1.20 (1.10, 1.31)	1.67 (1.52, 1.83)	1.68 (1.55, 1.83)	0.64 (0.59, 0.69)
Diabetes, noninsulin	1.27 (1.15, 1.39)	NA	1.45 (1.34, 1.57)	1.30 (1.16, 1.45)	NA	NA	1.34 (1.22, 1.47)	1.22 (1.15, 1.30)	0.85 (0.81, 0.88)
Dialysis versus no dialysis and creatinine = 1.0	4.59 (3.65, 5.77)	1.65 (1.34, 2.03)	NA	3.07 (2.74, 3.43)	NA	1.79 (1.60, 2.01)	2.42 (2.21, 2.66)	2.94 (2.64, 3.27)	0.23 (0.16, 0.33)
Preoperative IABP/ inotropes	1.47 (1.26, 1.71)	NA	1.34 (1.15, 1.57)	2.21 (1.90, 2.56)	1.69 (1.08, 2.65)	1.14 (1.02, 1.29)	1.75 (1.59, 1.94)	1.46 (1.30, 1.63)	0.63 (0.51, 0.77)
Shock	1.62 (1.29, 2.03)	1.65 (1.21, 2.25)	NA	2.09 (1.77, 2.47)	NA	1.32 (1.11, 1.58)	2.11 (1.80, 2.49)	1.05 (0.85, 1.31)	NA
Female versus male (at BSA=1.8)	1.11 (0.97, 1.27)	1.25 (1.09, 1.43)	0.97 (0.88, 1.07)	1.06 (0.98, 1.16)	0.98 (0.72, 1.33)	0.79 (0.72, 0.87)	1.03 (0.98, 1.08)	1.09 (0.99, 1.19)	0.69 (0.66, 0.73)
CLD (moderate versus mild, or severe versus moderate)	1.08 (1.01, 1.16)	NA	1.18 (1.13, 1.23)	1.26 (1.22, 1.30)	1.27 (1.13, 1.42)	1.09 (1.06, 1.12)	1.17 (1.14, 1.20)	1.16 (1.11, 1.22)	0.81 (0.79, 0.83)
Reoperation, 1 previous operation <sup>a</sup>	2.11 (1.78, 2.49)	2.09 (1.64, 2.65)	1.55 (1.31, 1.84)	1.50 (1.34, 1.67)	NA	1.31 (1.16, 1.49)	1.55 (1.42, 1.70)	1.42 (1.27, 1.59)	0.67 (0.62, 0.72)
Reoperation, ≥ 2 previous operations <sup>a</sup>	2.48 (1.99, 3.08)	2.36 (1.76, 3.16)	1.66 (1.33, 2.07)	2.03 (1.76, 2.35)	NA	1.41 (1.19, 1.67)	1.96 (1.73, 2.22)	1.76 (1.52, 2.03)	0.50 (0.43, 0.58)
Status emergent, no resuscitation <sup>a</sup>	2.74 (1.99, 3.78)	2.78 (1.85, 4.17)	2.20 (1.59, 3.05)	3.19 (2.41, 4.23)	NA	1.63 (1.31, 2.03)	3.23 (2.66, 3.93)	2.45 (2.02, 2.97)	0.33 (0.25, 0.42)
Status emergent, with resuscitation or salvage <sup>a</sup>	5.78 (3.77, 8.85)	2.11 (1.06, 4.19)	2.46 (1.56, 3.88)	2.46 (1.66, 3.65)	NA	NA	3.38 (2.36, 4.84)	NA	0.32 (0.19, 0.54)

Table 5. Continued

D. Odds ratios for mitral valve repair									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50	1.80 (1.62, 2.00)	1.48 (1.38, 1.59)	1.55 (1.41, 1.71)	1.31 (1.26, 1.36)	1.52 (1.31, 1.76)	1.20 (1.13, 1.27)	1.31 (1.26, 1.37)	1.50 (1.41, 1.60)	0.62 (0.60, 0.65)
Age 70 versus 50	3.24 (2.63, 4.00)	2.19 (1.90, 2.52)	2.42 (2.00, 2.92)	1.71 (1.59, 1.84)	2.31 (1.72, 3.10)	1.44 (1.29, 1.62)	1.73 (1.58, 1.89)	2.25 (1.98, 2.55)	0.39 (0.36, 0.42)
Age 80 versus 50	6.72 (5.00, 9.04)	3.21 (2.70, 3.81)	4.11 (3.14, 5.38)	2.31 (2.12, 2.52)	2.73 (1.95, 3.80)	1.75 (1.48, 2.07)	2.42 (2.12, 2.76)	3.78 (3.17, 4.51)	0.19 (0.17, 0.22)
CHF, not NYHA IV	1.29 (1.18, 1.42)	NA	1.24 (1.14, 1.34)	1.16 (0.99, 1.35)	NA	NA	1.11 (0.99, 1.24)	1.25 (1.17, 1.34)	0.92 (0.80, 1.05)
CHF, NYHA IV	1.83 (1.62, 2.07)	NA	1.61 (1.44, 1.81)	1.67 (1.43, 1.95)	NA	1.25 (1.17, 1.35)	1.50 (1.33, 1.68)	1.54 (1.40, 1.68)	0.76 (0.65, 0.90)
Diabetes, insulin	1.62 (1.43, 1.83)	NA	1.91 (1.70, 2.14)	1.68 (1.42, 1.97)	1.56 (1.05, 2.31)	1.20 (1.10, 1.31)	1.57 (1.36, 1.81)	1.68 (1.55, 1.83)	0.64 (0.59, 0.69)
Diabetes, noninsulin	1.27 (1.15, 1.39)	NA	1.45 (1.34, 1.57)	1.31 (1.11, 1.55)	NA	NA	1.26 (1.10, 1.45)	1.22 (1.15, 1.30)	0.85 (0.81, 0.88)
Dialysis versus no dialysis and creatinine = 1.0	6.24 (4.19, 9.30)	1.65 (1.34, 2.03)	NA	3.07 (2.74, 3.43)	NA	1.79 (1.60, 2.01)	2.42 (2.21, 2.66)	2.94 (2.64, 3.27)	0.26 (0.19, 0.37)
Preoperative IABP/ inotropes	1.47 (1.26, 1.71)	NA	1.34 (1.15, 1.57)	2.90 (2.28, 3.70)	1.69 (1.08, 2.65)	1.14 (1.02, 1.29)	1.75 (1.59, 1.94)	1.46 (1.30, 1.63)	0.49 (0.38, 0.64)
Shock	1.62 (1.29, 2.03)	1.65 (1.21, 2.25)	NA	2.09 (1.77, 2.47)	NA	1.32 (1.11, 1.58)	2.11 (1.80, 2.49)	2.50 (1.51, 4.12)	NA
Female versus male (at BSA = 1.8)	0.97 (0.77, 1.21)	1.25 (1.09, 1.43)	0.97 (0.88, 1.07)	1.23 (1.10, 1.38)	0.98 (0.72, 1.33)	0.90 (0.80, 1.02)	1.03 (0.98, 1.08)	1.28 (1.12, 1.47)	0.69 (0.66, 0.73)
CLD (moderate versus mild, or severe versus moderate)	1.23 (1.09, 1.39)	NA	1.18 (1.13, 1.23)	1.26 (1.22, 1.30)	1.27 (1.13, 1.42)	1.09 (1.06, 1.12)	1.17 (1.14, 1.20)	1.26 (1.15, 1.40)	0.81 (0.79, 0.83)
Reoperation, 1 previous operation <sup>a</sup>	2.11 (1.78, 2.49)	2.09 (1.64, 2.65)	1.55 (1.31, 1.84)	2.06 (1.73, 2.45)	NA	1.31 (1.16, 1.49)	1.55 (1.42, 1.70)	1.42 (1.27, 1.59)	0.67 (0.62, 0.72)
Reoperation ≥ 2 previous operations <sup>a</sup>	2.48 (1.99, 3.08)	2.36 (1.76, 3.16)	1.66 (1.33, 2.07)	2.80 (2.32, 3.37)	NA	1.41 (1.19, 1.67)	1.96 (1.73, 2.22)	1.76 (1.52, 2.03)	0.50 (0.43, 0.58)
Status emergent, no resuscitation <sup>a</sup>	8.73 (4.84, 15.74)	2.78 (1.85, 4.17)	3.03 (1.69, 5.43)	6.12 (3.96, 9.46)	NA	1.63 (1.31, 2.03)	3.23 (2.66, 3.93)	2.45 (2.02, 2.97)	0.33 (0.25, 0.42)
Status emergent, with resuscitation or salvage <sup>a</sup>	18.39 (9.68, 34.96)	2.11 (1.06, 4.19)	3.39 (1.76, 6.54)	4.72 (2.71, 8.23)	NA	NA	3.38 (2.36, 4.84)	NA	0.32 (0.19, 0.54)

<sup>a</sup> Variable interacts with age. Reported odds ratio represents effect of risk factor for patients aged 50 years old.

BSA = body surface area; CHF = congestive heart failure; CLD = chronic lung disease; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; NA = not applicable; NYHA = New York Heart Association; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

( $\leq 40$ ,  $> 40$ ). Calibration plots (observed versus expected) based on the overall validation sample are presented in Figure 1. The average absolute difference between observed versus predicted event rates within deciles of predicted risk ranged from 0.06% for deep sternal wound infection to 1.06% for prolonged postoperative stay. Analogous figures were produced for specific valve procedures and numerous subgroups, and these are available at [www.sts.org/riskmodels](http://www.sts.org/riskmodels).

Model fit appeared to be adequate for each endpoint with the possible exception of deep sternal wound infection, which revealed some overfitting within certain subgroups. A modest degree of overfitting was expected for this endpoint given the relatively small number of infections and large number of candidate predictors.

Discrimination was assessed by the c-statistic, also known as the area under the receiver operating characteristic (ROC) curve. Table 4 presents the discrimination of each model in the development and validation samples for all patients combined and for subgroups consisting of AVR, MVR, and MVRepair. In the validation sample, c-statistics for the operative mortality model were 0.799 (overall), 0.759 (AVR), 0.802 (MVR), and 0.844 (MVRepair). C-statistics in the validation sample for other endpoints ranged from 0.619 for reoperation in the AVR subgroup to 0.800 for prolonged length of stay in the MVRepair subgroup.

### Final Models

After validating the models in the 40% validation sample, the development and validation samples were then combined, and the final model coefficients were estimated using the overall 100% combined sample. The final logistic regressions were estimated using generalized estimating equations with empirical (sandwich) standard error estimates to account for clustering of patients within institutions [10]. An independence working correlation matrix was used to apply the generalized estimating equations methodology. With this approach, the estimated regression coefficients were identical to those obtained using ordinary logistic regression, but the standard errors were adjusted to account for the clustered data structure.

### Odds Ratios

Odds ratios and 95% confidence intervals (CI) for the final selected models are presented in Table 5. "Not applicable" indicates that the specific predictor was not included in a particular risk model. Because several variables interact with surgery type, the odds ratios for these variables differ depending on the type of surgery (AVR, MVR, MVRepair). For example, in the operative mortality model, the odds ratio for emergent status is 3.77 (95% CI: 2.75, 5.16) for AVR, 2.74 (95% CI: 1.99, 3.78) for MVR, and 8.73 (95% CI: 4.84, 15.74) for MVRepair. Odds ratios that do not interact with surgery type are summarized in Table 5, Part A. Odds ratios that differ by surgery type for at least one endpoint are presented in Table 5, Parts B, C, and D.

### Final Model Intercept and Coefficients

The final risk prediction algorithms, including all coefficients and intercepts, are presented in the Appendix.

### Limitations

The limitations for these valve models are similar to those for the CABG models and are thoroughly discussed in Part 1 of this series (2008 STS CABG risk models).

### Conclusion

The STS Quality Measurement Task Force has developed and tested nine new risk-adjustment models for isolated valve surgery using the STS NCD. This report includes a detailed exposition of the model development process, including not only statistical issues but also the many clinical and pragmatic judgments that were required. An online risk calculator is also available through a link from the STS website.

### References

1. Ambler G, Omar RZ, Royston P, Kinsman R, Keogh BE, Taylor KM. Generic, simple risk stratification model for heart valve surgery. *Circulation* 2005;112:224–31.
2. Edwards FH, Peterson ED, Coombs LP, et al. Prediction of operative mortality after valve replacement surgery. *J Am Coll Cardiol* 2001;37:885–92.
3. Hannan EL, Wu C, Bennett EV, et al. Risk index for predicting in-hospital mortality for cardiac valve surgery. *Ann Thorac Surg* 2007;83:921–9.
4. Jin R, Grunkemeier GL, Starr A. Validation and refinement of mortality risk models for heart valve surgery. *Ann Thorac Surg* 2005;80:471–9.
5. Kuduvali M, Grayson AD, Au J, Grotte G, Bridgewater B, Fabri BM. A multi-centre additive and logistic risk model for in-hospital mortality following aortic valve replacement. *Eur J Cardiothorac Surg* 2007;31:607–13.
6. Nowicki ER, Birkmeyer NJ, Weintraub RW, et al. Multivariable prediction of in-hospital mortality associated with aortic and mitral valve surgery in Northern New England. *Ann Thorac Surg* 2004;77:1966–77.
7. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691–2.
8. Little RJA, Rubin DB. *Statistical analysis with missing data*. 2nd ed. Hoboken, NJ: Wiley-Interscience, 2002.
9. Marcin JP, Romano PS. Size matters to a model's fit. *Crit Care Med* 2007;35:2212–3.
10. Liang KY, Zeger SL. Longitudinal data-analysis using generalized linear-models. *Biometrika* 1986;73:13–22.

### Appendix

#### Regression Coefficients and Variable Definitions for STS 2008 Valve Models

For each endpoint, the formula for calculating a patient's predicted risk of the endpoint has the form:

$$\text{Predicted Risk} = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

where  $x_1, x_2, \dots, x_n$  denote patient preoperative risk factors (eg, quantitative variables such as age, and comorbidities coded as 1 = present, 0 = absent); and  $\beta_0, \beta_1, \dots, \beta_n$  denote regression coefficients (numerical constants). Regression coefficients for each endpoint are presented in Appendix Table 1. The variables  $x_1, x_2, \dots, x_n$  are the same for each endpoint and are defined in Appendix Table 2. The regression coefficient for the time trend is not presented. Instead, the intercept has been adjusted to incorporate the time trend. This adjusted intercept reflects the baseline risk for a reference period of July to December 2006.

Appendix Table 1. Regression Coefficients

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Intercept	−5.78680	−5.83957	−5.52789	−3.96796	−7.11095	−3.08816	−3.06527	−4.30676	1.25115
Atrial fibrillation	0.18074	0.05524	0.00000	0.16527	0.00000	0.10305	0.11403	0.15530	−0.30247
Age function 1	0.03557	0.03909	0.03219	0.02683	0.04180	0.01512	0.02041	0.02670	−0.02834
Age function 3	0.02804	−0.00132	0.01809	0.00629	−0.05024	0.00218	0.01282	0.02315	−0.04637
Age by reoperation function	−0.01308	−0.02043	−0.00551	−0.00840	−0.00939	−0.00697	−0.00684	−0.00485	0.00927
Age by status function	−0.02495	−0.02987	−0.00721	−0.01377	0.00277	0.00102	−0.00677	−0.00379	−0.00795
Age by MVR function	0.01436	0.00000	−0.00245	0.00000	0.00000	0.00715	0.00848	−0.00324	−0.00603
Age by MVRepair function	0.02326	0.00000	0.01190	0.00000	0.00000	0.00315	0.00685	0.01378	−0.01883
BSA function 1	−1.40168	−0.38619	−0.71012	−1.11750	0.14188	−0.73553	−0.91858	−0.82801	0.77317
BSA function 2	2.16782	0.23148	1.92875	2.29127	2.04603	0.83644	1.65638	1.65423	−1.76728
CHF but not NYHA IV	0.25590	0.00000	0.21233	0.28353	0.00000	0.00000	0.17974	0.22508	−0.15108
CHF and NYHA IV	0.60544	0.00000	0.47812	0.65056	0.00000	0.22686	0.48025	0.42957	−0.33521
CHF by MVR function	0.00000	0.00000	0.00000	−0.11007	0.00000	0.00000	−0.07864	0.00000	0.11503
CHF by MVRepair function	0.00000	0.00000	0.00000	−0.13792	0.00000	0.00000	−0.07731	0.00000	0.06468
CLD function	0.23846	0.00000	0.16629	0.22816	0.23817	0.08406	0.16044	0.25263	−0.21022
CLD by MVR function	−0.15906	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.10092	0.00000
CLD by MVRepair function	−0.03243	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.01795	0.00000
Creatinine function 1	0.43909	0.29230	0.71439	0.45646	0.00000	0.23562	0.49230	0.45631	−0.44178
CVD without prior CVA	0.00000	0.27837	0.20531	0.12726	0.00000	0.05830	0.07684	0.00000	−0.22223
CVD and prior CVA	0.00000	0.59220	0.20018	0.24512	0.00000	0.13200	0.18343	0.33480	−0.25595
Diabetes, noninsulin	0.23563	0.00000	0.37172	0.11040	0.00000	0.00000	0.11355	0.19843	−0.16630
Diabetes, insulin	0.48368	0.00000	0.64648	0.35367	0.44389	0.18293	0.33165	0.51913	−0.45093
Diabetes by MVR function	0.00000	0.00000	0.00000	0.15051	0.00000	0.00000	0.17990	0.00000	0.00000
Diabetes by MVRepair function	0.00000	0.00000	0.00000	0.16260	0.00000	0.00000	0.11734	0.00000	0.00000
Dialysis	1.48666	0.79199	0.00000	1.57690	1.19109	0.81972	1.37741	1.53351	−1.69019
Dialysis by MVR function	0.47550	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.20998
Dialysis by MVRepair function	0.78385	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.07964
Ejection fraction function	0.00904	0.00000	0.00407	0.01107	0.02308	0.00734	0.00925	0.01111	−0.01348
Endocarditis, active	0.66737	0.62434	0.77276	0.76318	0.00000	0.43876	0.67810	1.02521	−1.08299
Female	0.20372	0.21925	−0.03031	0.25668	−0.02355	−0.14567	0.03066	0.22437	−0.36400
Female by MVR function	−0.10089	0.00000	0.00000	−0.19465	0.00000	−0.08773	0.00000	−0.14211	0.00000
Female by MVRepair function	−0.23812	0.00000	0.00000	−0.04564	0.00000	0.04424	0.00000	0.02470	0.00000
Female by BSA function 1	0.96491	−0.02257	0.83074	0.77598	2.00214	0.16707	0.52716	0.57195	−0.75434
Female by BSA function 2	0.18084	−0.07419	0.08397	−0.58460	−1.87036	0.25158	−0.09063	−0.12289	0.35123
Hypertension	0.11372	0.17789	0.29770	0.10799	0.00000	0.00000	0.10361	0.00000	−0.06504
IABP or inotropes	0.38682	0.00000	0.29606	0.57608	0.52474	0.13432	0.56046	0.37621	−0.57115
IABP by MVR function	0.00000	0.00000	0.00000	0.21517	0.00000	0.00000	0.00000	0.00000	0.10760
IABP by MVRepair function	0.00000	0.00000	0.00000	0.48870	0.00000	0.00000	0.00000	0.00000	−0.13850

Appendix Table 1. Continued

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Immunosuppressive treatment	0.35022	0.00000	0.32828	0.00000	0.00000	0.00000	0.14887	0.27152	0.00000
Insufficiency mitral	0.00000	0.23253	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Insufficiency tricuspid	0.00000	0.00000	0.13159	0.12973	0.00000	0.08969	0.18929	0.15846	−0.20027
Left main disease	0.17593	0.00000	0.17280	0.00000	0.77557	0.00000	0.00000	0.00000	0.00000
MI ≤ 21 days	0.13276	0.00000	0.00000	0.31706	0.00000	0.03495	0.24687	0.18812	−0.20961
MVR	0.10284	0.00000	0.40455	0.44639	0.00000	0.12852	0.13795	0.58004	−0.61402
MVRepair	−0.65440	0.00000	−0.23666	−0.19726	0.00000	−0.22398	−0.23002	−0.37618	0.25710
No. diseased vessel function	0.00000	0.09556	0.00000	0.06299	0.00000	0.00000	0.03700	0.03312	−0.10126
Peripheral vascular disease	0.21980	0.25236	0.00000	0.00000	0.00000	0.19758	0.13174	0.15342	−0.18903
Race black	0.00000	0.28378	0.40941	0.34795	0.00000	0.23856	0.31567	0.37161	−0.44177
Race Hispanic	0.00000	−0.13774	0.14968	0.06720	0.00000	0.12816	0.08581	0.15128	−0.20068
Reop, 1 previous operation	0.74484	0.73489	0.43804	0.60704	0.00000	0.27365	0.44052	0.35252	−0.40042
Reop, ≥ 2 previous operations	0.90625	0.85841	0.50595	0.91229	0.00000	0.34233	0.67201	0.56294	−0.69765
Reop by MVR function	0.00000	0.00000	0.00000	−0.20333	0.00000	0.00000	0.00000	0.00000	0.00000
Reop by MVRepair function	0.00000	0.00000	0.00000	0.11559	0.00000	0.00000	0.00000	0.00000	0.00000
Shock	0.47961	0.50213	0.00000	0.73670	0.00000	0.28068	0.74786	0.55376	0.00000
Shock by MVR function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.50071	0.00000
Shock by MVRepair function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.36096	0.00000
Status urgent	0.25552	0.00000	0.19344	0.25714	0.00000	0.15548	0.19858	0.35184	−0.36106
Status emergent	1.32597	1.02109	1.13199	1.51294	0.00000	0.49075	1.17360	0.89480	−1.12373
Status salvage	2.07144	0.74530	1.24544	1.25342	0.00000	0.00000	1.21823	0.00000	−1.13785
Status by MVR function	−0.31729	0.00000	−0.34380	−0.35206	0.00000	0.00000	0.00000	0.00000	0.00000
Status by MVRepair function	0.84051	0.00000	−0.02373	0.29927	0.00000	0.00000	0.00000	0.00000	0.00000
Stenosis aortic	0.00000	0.00000	0.00000	−0.10782	0.00000	−0.10852	−0.07479	−0.15434	0.06873
Stenosis mitral	0.21309	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Unstable angina	0.18950	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

BSA = body surface area; CHF = congestive heart failure; CLD = chronic lung disease; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; Mort = mortality; MVR = mitral valve replacement; MVRepair = mitral valve repair; NYHA = New York Heart Association; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.



Appendix Table 2. Definition of Variables Appearing in STS 2008 Valve Models

Variable	Definition
Intercept	= 1 for all patients
Atrial fibrillation	= 1 if patient has history of preop atrial fibrillation, = 0 otherwise
Age function 1	= max (age – 50, 0)
Age function 3	= max (age – 75, 0)
Age by reoperation function	= Age function 1 if surgery is a reoperation, = 0 otherwise
Age by status function	= Age function 1 if status is emergent or salvage, = 0 otherwise
Age by MVR function	= Age function 1 if operation is MVR, = 0 otherwise
Age by MVRepair function	= Age function 1 if operation is MVRepair, = 0 otherwise
BSA function 1	= max (1.4, min [2.6, BSA]) – 1.8
BSA function 2	= (BSA function 1) <sup>2</sup>
CHF but not NYHA IV	= 1 if patient has CHF and is not NYHA class IV, = 0 otherwise
CHF and NYHA IV	= 1 if patient has CHF and is NYHA class IV, = 0 otherwise
CHF by MVR function	= 1 if patient has CHF and operation is MVR, = 0 otherwise
CHF by MVRepair function	= 1 if patient has CHF and operation is MVRepair, = 0 otherwise
CLD function	= 0 if no CLD, = 1 if mild CLD, = 2 if moderate CLD, = 3 if severe CLD
CLD by MVR function	= CLD function if operation is MVR, = 0 otherwise
CLD by MVRepair function	= CLD function if operation is MVRepair, = 0 otherwise
Creatinine function 1	= max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise
CVD without prior CVA	= 1 if patient has history of CVD and no prior CVA, = 0 otherwise
CVD and prior CVA	= 1 if patient has history of CVD and a prior CVA, = 0 otherwise
Diabetes, noninsulin	= 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin	= 1 if patient has diabetes treated with insulin, = 0 otherwise
Diabetes by MVR function	= 1 if patient has diabetes and operation is MVR, = 0 otherwise
Diabetes by MVRepair function	= 1 if patient has diabetes and operation is MVRepair, = 0 otherwise
Dialysis	= 1 if patient requires dialysis preoperatively, = 0 otherwise
Dialysis by MVR function	= 1 if patient has history of dialysis and operation is MVR, = 0 otherwise
Dialysis by MVRepair function	= 1 if patient has history of dialysis and operation is MVRepair, = 0 otherwise
Ejection fraction function	= max (50–ejection fraction, 0)
Endocarditis, active	= 1 if patient has active endocarditis, = 0 otherwise
Female	= 1 if patient is female, = 0 otherwise
Female by MVR function	= 1 if female and operation is MVR, = 0 otherwise
Female by MVRepair function	= 1 if female and operation is MVRepair, = 0 otherwise
Female by BSA function 1	= BSA function 1 if female, = 0 otherwise
Female by BSA function 2	= BSA function 2 if female, = 0 otherwise
Hypertension	= 1 if patient has hypertension, = 0 otherwise
IABP or inotropes	= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
IABP by MVR function	= 1 if patient requires preop IABP/inotropes and operation is MVR, = 0 otherwise
IABP by MVRepair function	= 1 if patient requires preop IABP/inotropes and operation is MVRepair, = 0 otherwise
Immunosuppressive treatment	= 1 if patient received immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency mitral	= 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency tricuspid	= 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease	= 1 if patient has left main disease, = 0 otherwise
MI ≤ 21 days	= 1 if patient has history of MI within 21 days of surgery, = 0 otherwise
MVR	= 1 if valve operation is mitral valve replacement, = 0 otherwise
MVRepair	= 1 if valve operation is mitral valve repair, = 0 otherwise
No. diseased vessel function	= 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise
Peripheral vascular disease	= 1 if patient has peripheral vascular disease, = 0 otherwise
Race black	= 1 if patient is black, = 0 otherwise
Race Hispanic	= 1 if patient is nonblack Hispanic, = 0 otherwise
Reop, 1 prior operation	= 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, ≥ 2 prior operations	= 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Reop by MVR function	= 1 if surgery is a reoperation and operation is MVR, = 0 otherwise
Reop by MVRepair function	= 1 if surgery is a reoperation and operation is MVRepair, = 0 otherwise

Appendix Table 2. Continued

Variable	Definition
Shock	= 1 if patient was in shock at time of procedure, = 0 otherwise
Shock by MVR function	= 1 if shock and operation is MVR, = 0 otherwise
Shock by MVRepair function	= 1 if shock and operation is MVRepair, = 0 otherwise
Status urgent	= 1 if status is urgent, = 0 otherwise
Status emergent	= 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage	= 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Status by MVR function	= 1 if status is emergent or salvage and operation is MVR, = 0 otherwise
Status by MVRepair function	= 1 if status is emergent or salvage and operation is MVRepair, = 0 otherwise
Stenosis aortic	= 1 if patient has aortic stenosis, = 0 otherwise
Stenosis mitral	= 1 if patient has mitral stenosis, = 0 otherwise
Unstable angina	= 1 if patient has unstable angina, no MI within 7 days of surgery, = 0 otherwise

Note: See [www.sts.org](http://www.sts.org) for exact definitions of terms used above.

BSA = body surface area; CHF = congestive heart failure; CLD = chronic lung disease; CVA = cerebrovascular accident, or stroke; CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; MVR = mitral valve replacement; MVRepair = mitral valve repair; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; PLOS = prolonged length of stay; Preop = preoperative; Reop = reoperation; Comp = composite adverse event (any); RF = renal failure; SLOS = short length of stay; STS = The Society of Thoracic Surgeons; Vent = prolonged ventilation.

# The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 3—Valve Plus Coronary Artery Bypass Grafting Surgery

David M. Shahian, MD,<sup>a</sup> Sean M. O'Brien, PhD,<sup>b</sup> Giovanni Filardo, PhD, MPH,<sup>c</sup> Victor A. Ferraris, MD,<sup>d</sup> Constance K. Haan, MD,<sup>e</sup> Jeffrey B. Rich, MD,<sup>f</sup> Sharon-Lise T. Normand, PhD,<sup>g</sup> Elizabeth R. DeLong, PhD,<sup>b</sup> Cynthia M. Shewan, PhD,<sup>h</sup> Rachel S. Dokholyan, MPH,<sup>b</sup> Eric D. Peterson, MD, MPH,<sup>b</sup> Fred H. Edwards, MD,<sup>e</sup> and Richard P. Anderson, MD<sup>i†</sup>

<sup>a</sup>Massachusetts General Hospital, Boston, Massachusetts; <sup>b</sup>Duke Clinical Research Institute, Durham, North Carolina; <sup>c</sup>Institute for Health Care Research and Improvement, Baylor Health Care System, Dallas, Texas; <sup>d</sup>University of Kentucky Chandler Medical Center, Division of Cardiovascular and Thoracic Surgery, Lexington, Kentucky; <sup>e</sup>University of Florida, Division of Cardiothoracic Surgery, Jacksonville, Florida; <sup>f</sup>Sentara Cardiovascular Research Institute, Norfolk, Virginia; <sup>g</sup>Department of Health Care Policy, Harvard Medical School, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; <sup>h</sup>The Society of Thoracic Surgeons, Chicago, Illinois; and <sup>i</sup>Seattle, Washington

**Background.** Since 1999, The Society of Thoracic Surgeons (STS) has published two risk models that can be used to adjust the results of valve surgery combined with coronary artery bypass graft surgery (CABG). The most recent was developed from data for patients who had surgery between 1994 and 1997 using operative mortality as the only endpoint. Furthermore, this model did not specifically consider mitral valve repair plus CABG, an increasingly common procedure. Consistent with STS policy of periodically updating and improving its risk models, new models for valve surgery combined with CABG have been developed. These models specifically address both perioperative morbidity and mitral valve repair, and they are based on contemporary data.

**Methods.** The final study population consisted of 101,661 procedures, including aortic valve replacement (AVR) plus CABG, mitral valve replacement (MVR) plus CABG, or mitral valve repair (MVRRepair) plus CABG between January 1, 2002, and December 31, 2006. Model outcomes included operative mortality, stroke, deep sternal wound infection, reoperation, prolonged ventilation, renal failure, composite major morbidity or mortality, prolonged postoperative length of stay, and short postoperative length of stay. Candidate variables were screened for frequency of missing data, and imputation techniques were used where appropriate. Stepwise variable selection was employed, supplemented by advice from an expert panel of cardiac surgeons and biostatisticians. Several variables were forced into models to insure face validity (eg, atrial

fibrillation for the permanent stroke model, sex for all models). Based on preliminary analyses of the data, a single model was employed for valve plus CABG, with indicator variables for the specific type of procedure. Interaction terms were included to allow for differential impact of predictor variables depending on procedure type. After validating the model in the 40% validation sample, the development and validation samples were then combined, and the final model coefficients were estimated using the overall 100% combined sample. The final logistic regression model was estimated using generalized estimating equations to account for clustering of patients within institutions.

**Results.** The c-index for mortality prediction for the overall valve plus CABG population was 0.75. Morbidity model c-indices for specific complications (permanent stroke, renal failure, prolonged ventilation > 24 hours, deep sternal wound infection, reoperation for any reason, major morbidity or mortality composite, and prolonged postoperative length of stay) for the overall group of valve plus CABG procedures ranged from 0.622 to 0.724, and calibration was excellent.

**Conclusions.** New STS risk models have been developed for heart valve surgery combined with CABG. These are the first valve plus CABG models that also include risk prediction for individual major morbidities, composite major morbidity or mortality, and short and prolonged length of stay.

(Ann Thorac Surg 2009;88:S43–62)

© 2009 by The Society of Thoracic Surgeons

Risk models for cardiac surgery were first developed almost 2 decades ago, and most of these early models focused on isolated coronary artery bypass graft surgery (CABG) [1–4]. The results of this frequently performed

surgical procedure have often been used as the sole marker to assess the quality of care delivered by cardiac surgical programs. Risk-adjusted results for CABG have been used

†This author is deceased. Former Chair, Quality, Research and Patient Safety Council, The Society of Thoracic Surgeons, Chicago, IL.

Address correspondence to Dr Shahian, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114; e-mail: dshahian@partners.org.

Drs Shahian, O'Brien, Filardo, Ferraris, Haan, Rich, Normand, DeLong, Shewan, Peterson, Edwards, Anderson, and Ms Dokholyan, have no conflicts of interest to declare regarding this work.

**Abbreviations and Acronyms**

AVR	= aortic valve replacement
CABG	= coronary artery bypass graft surgery
MI	= myocardial infarction
MVR	= mitral valve replacement
MVRepair	= mitral valve repair
NCD	= National Adult Cardiac Surgery Database
QMTF	= Quality Measurement Task Force
STS	= The Society of Thoracic Surgeons

for hospital and regional quality improvement initiatives, public reporting, pay for performance reimbursement programs, decision support, patient counseling, and clinical research. Earlier models focused primarily on mortality prediction, but subsequent models have been developed for both risk-adjusted morbidity and length of stay [5].

The other commonly performed category of cardiac surgery consists of operations on the heart valves, either alone or in combination with CABG. Relative to isolated CABG procedures, which are declining in frequency, the proportion of valve cases is steadily increasing. To better assess the overall performance of cardiac surgery programs, to discern the factors that are most significantly related to patient outcomes, and to aid in physician and patient decision-making, risk models have now also been developed for heart valve surgery [6–18].

Unlike risk models for isolated CABG, a relatively standardized procedure, valve surgery encompasses a much more diverse group of operations. There are four cardiac valves, and they may malfunction in a number of quite different ways (eg, stenosis, regurgitation, infection, and so forth). The valves may be repaired or replaced with a wide range of techniques and prosthetics. In some cases, procedures may be performed on multiple valves, or the valve procedure may be combined with CABG.

Given the heterogeneity of heart valve surgery, it is not surprising that a variety of risk-modeling techniques has been applied. At one extreme, the European System for Cardiac Operative Risk Evaluation (EuroSCORE) algorithm, developed by a European consortium, groups all cardiac operations together in a single risk model with indicator variables included to account for valve procedures [14, 18]. Although this approach is simple and easy to apply, recent studies by van Gameren and associates [19] have suggested that a dedicated valve risk model may have better discrimination and calibration than the EuroSCORE algorithm when applied to valve surgery patients. Combined models for aortic and mitral valve procedures with or without CABG have been developed by Jin and colleagues [12] and by Ambler and associates [13]. The 2001 valve models developed by The Society of Thoracic Surgeons (STS) [6] consisted of one model for all isolated valve procedures and one model for valve procedures combined with CABG, and a 2007 risk model derived from the New York Cardiac Surgery Reporting System used a similar stratification [8].

Unified valve models reflect the fact that many risk factors are common to both aortic and mitral valve surgery. They offer simplicity, and they also permit larger sample sizes for development and validation [12]. However, there are significant differences between aortic and mitral valvular disease in both pathophysiology and outcomes, and both also differ substantially from isolated CABG [11]. Some investigators advocate separate aortic and mitral valve models to have more homogeneous patient populations. Examples include models developed by STS, the New York Cardiac Surgery Reporting System, and the Northern New England Cardiovascular Disease Study Group [7, 9, 10]. Some of these models have been developed solely for isolated valve replacement, some have included CABG as a separate predictor variable in the isolated valve model, and some models have focused specifically on valve plus CABG. All these decisions involve a tradeoff—the more homogeneous the study group, the fewer patients are available for model development and validation [12].

Because of the large number of valve surgery patients available for analysis in the STS National Adult Cardiac Surgery Database (NCD), our approach has favored separate models for valve plus CABG versus isolated valve surgery. The STS Quality Measurement Task Force (QMTF) presumes that when adequate numbers of patients are available for study, relatively homogeneous operative categories result in more accurate risk prediction. Furthermore, recent studies by van Gameren and colleagues [19] suggest that the valve plus CABG group may be the most difficult to model accurately, thus meriting its own algorithm.

Several new features were added to the 2008 valve plus CABG models described in this report. First, recognizing that mitral valve repair is often different in both etiology and outcomes than replacement, the QMTF has included interactions between surgery type and several key predictor variables. Fitting a single model with several such interactions is useful. It allows for pooling information across related groups of valve procedures without making an *a priori* assumption that the effect of key risk factors is constant across these groups. Finally, new models have been developed for specific major complications of each valve plus CABG procedure, as well as for composite morbidity, mortality, and for both short and prolonged postoperative length of stay.

The authors of this report are members of the STS QMTF who were involved in this risk model development project.

## Study Population and Endpoints

Our general approaches to variable selection and risk model development have been described in the companion articles on isolated CABG (Part 1) and isolated valve surgery (Part 2). Details specific to the valve plus CABG models are included in this report.

### Study Population

The study population for this analysis consisted of single aortic or mitral valve surgical procedures combined with



CABG performed on adult patients between January 1, 2002, and December 31, 2006. Only the following procedures were included: (1) isolated aortic valve replacement (AVR) plus CABG; (2) isolated mitral valve replacement (MVR) plus CABG; and (3) isolated mitral valve repair (MVRRepair) plus CABG.

Because of the relatively small number of pulmonic, tricuspid, multiple valve procedures, and aortic repairs, these cases were not included in the current models. Patients undergoing isolated valve surgery without CABG were excluded from the current analysis, but these cases are the focus of a separate model described in Part 2 of this three-part series. Patients with missing sex data ( $n = 17$ ) were excluded because these patients are not allowed in the analysis dataset used for creating STS database participant feedback reports. Patients on dialysis preoperatively ( $n = 2,443$ ) were excluded when developing the risk model for prediction of postoperative renal failure. The final study population comprised 101,661 patient operations (66,074 AVR plus CABG; 13,663 MVR plus CABG; and 21,924 MVRRepair plus CABG) from 814 STS NCD participating groups.

Characteristics of the study population are summarized in [Table 1](#).

### *Training and Validation Samples*

The study population was randomly divided into a 60% training (development) sample and a 40% test (validation) sample. The development sample was used to identify predictor variables and estimate model coefficients. Data from the validation sample were used to assess model fit, discrimination, and calibration. After choosing variables and assessing model fit, the development and validation samples were subsequently combined, and the final model coefficients were estimated using the combined (development plus validation) data.

### *Endpoints*

In developing the valve plus CABG risk models, we used the same nine endpoints that were analyzed in the STS isolated CABG (Part 1) and the STS isolated valve (Part 2) models. Morbidities in all three models are recorded only in-hospital, in contrast to the operative mortality endpoint defined below (although beginning with version 2.61, sternal infection will be recorded at 30 days): (1) operative mortality: death during the same hospitalization as surgery, regardless of timing or within 30 days of surgery regardless of venue; (2) permanent stroke (CVA): a central neurologic deficit persisting longer than 72 hours; (3) renal failure: a new requirement for dialysis or an increase of the serum creatinine to more than 2.0 mg/dL and double the most recent preoperative creatinine level; (4) prolonged ventilation ( $> 24$  hours); (5) deep sternal wound infection; (6) reoperation for any reason; (7) major morbidity or mortality, a composite defined as the occurrence of any of the above endpoints; (8) prolonged postoperative length of stay (PLOS): length of stay (LOS) more than 14 days (alive or dead); and (4) short postoperative length of stay (SLOS): LOS less than 6 days and patient alive at discharge.

Endpoint frequencies in the study population are presented in [Table 2](#).

## **Separate Versus Combined Models**

Given the variety of approaches used in previous models by STS and other developers, we investigated the option of developing separate models for the AVR plus CABG and MVR plus CABG populations, and we also studied how best to subdivide the mitral plus CABG population into repair versus replacement. Although we had a large study population available, many of the individual outcomes were relatively rare. We were concerned that the number of events would be too small to permit reliable estimation of the model coefficients in separate models for each valve. Thus, in theory, the development of separate custom models for each valve type could be inferior to a single combined model because the custom models would have a smaller sample size and hence larger variance.

As described in detail in Part 2 of this series (isolated valve surgery), we performed preliminary empirical analyses to compare two alternative strategies (separate versus combined AVR plus CABG and MVR/Repair plus CABG) for developing these risk models. We first developed separate models for the three subpopulations (AVR plus CABG, MVR plus CABG, and MVRRepair plus CABG), then modeled all three subpopulations together in a single model. In the latter approach, we included several interaction terms to allow the effect of certain risk factors to differ across the specific valve subpopulations. These strategies were used to develop risk models for operative mortality and permanent stroke, using a 60% development sample and a separate 40% validation sample. The performance of the combined model was then assessed separately within each subpopulation and compared to the model that was developed specifically for that subpopulation. In the case of mortality, the combined model had better discrimination (larger c-index) than the corresponding custom model in each of the three subpopulations (AVR plus CABG, MVR plus CABG, MVRRepair plus CABG). For stroke, the combined model had better discrimination in two of the three populations (all except AVR plus CABG). Finally, when explained variation was quantified by the generalized  $R^2$  index of Nagelkerke [20], the combined model had greater explained variation than the custom model in each subpopulation for each endpoint. These results provide empirical support for the use of a single model with several interactions, which allows pooling of information across valve groups without assuming that the effect of risk factors is constant.

## **Selection of Candidate Predictor Variables**

The candidate variables for the STS valve plus CABG models were identical to those in the STS isolated valve models, described in Part 2 of this series. They differed from the isolated CABG model variables in the following specific areas: (1) Percutaneous coronary intervention (PCI) occurring 6 hours or less before surgery was present in only 315 patients (0.3%) in the valve plus CABG study population, and was not included as a candidate variable. (2) Infectious endocarditis was not included in the isolated CABG model but was considered for the valve plus CABG model. Although this risk factor was rarely present (0.8% active



Table 1. Distribution of Risk Factors in Overall Study Population 2002 to 2006

Variable	Overall Valve + CABG (n = 101,661)		AVR + CABG (n = 66,074)		MVR + CABG (n = 13,663)		MVRRepair + CABG (n = 21,924)	
	N	%	N	%	N	%	N	%
<b>Demographics</b>								
Age, years								
< 55	6,693	6.6	2,983	4.51	1,309	9.58	2,401	10.95
55–64	17,188	16.9	9,132	13.82	2,790	20.42	5,266	24.02
65–74	33,628	33.1	21,313	32.26	4,667	34.16	7,648	34.88
≥ 75	44,152	43.4	32,646	49.41	4,897	35.84	6,609	30.15
Sex								
Male	65,588	64.5	44,619	67.53	7,348	53.78	13,621	62.13
Female	36,073	35.5	21,455	32.47	6,315	46.22	8,303	37.87
Race								
Caucasian	90,572	89.1	60,121	90.99	11,765	86.11	18,686	85.23
Black	4,534	4.5	2,094	3.17	914	6.69	1,526	6.96
Hispanic	2,487	2.4	1,487	2.25	354	2.59	646	2.95
Asian	1,083	1.1	542	0.82	191	1.40	350	1.60
Other	2,295	2.3	1,402	2.12	331	2.42	562	2.56
Missing	690	0.7	428	0.65	108	0.79	154	0.70
<b>Risk factors</b>								
Body surface area, m <sup>2</sup>								
< 1.50	3,340	3.3	1,985	3.00	638	4.67	717	3.27
1.50–1.74	20,779	20.4	12,580	19.04	3,500	25.62	4,699	21.43
1.75–1.99	40,017	39.4	25,814	39.07	5,440	39.82	8,763	39.97
≥ 2.00	36,956	36.4	25,361	38.38	3,996	29.25	7,599	34.66
Missing	569	0.6	334	0.51	89	0.65	146	0.67
Body mass index, kg/m <sup>2</sup>								
< 25	29,353	28.9	17,712	26.81	4,787	35.04	6,854	31.26
25–29	39,345	38.7	25,692	38.88	4,951	36.24	8,702	39.69
30–34	21,063	20.7	14,447	21.86	2,507	18.35	4,109	18.74
≥ 35	11,165	11.0	7,785	11.78	1,299	9.51	2,081	9.49
Missing	735	0.7	438	0.66	119	0.87	178	0.81
Diabetes mellitus								
No diabetes	68,112	67.0	44,489	67.33	9,517	69.66	14,106	64.34
Diabetes, noninsulin	23,383	23.0	15,705	23.77	2,642	19.34	5,036	22.97
Diabetes, insulin	9,848	9.7	5,677	8.59	1,463	10.71	2,708	12.35
Diabetes, missing treatment	167	0.2	105	0.16	20	0.15	42	0.19
Missing	151	0.1	98	0.15	21	0.15	32	0.15
Hypertension								
No	22,709	22.3	13,944	21.10	3,482	25.48	5,283	24.10
Yes	78,823	77.5	52,050	78.78	10,163	74.38	16,610	75.76
Missing	129	0.1	80	0.12	18	0.13	31	0.14
Hypercholesterolemia								
No	33,759	33.2	21,248	32.16	5,324	38.97	7,187	32.78
Yes	67,613	66.5	44,649	67.57	8,280	60.60	14,684	66.98
Missing	289	0.3	177	0.27	59	0.43	53	0.24
Past or present smoker								
No	43,687	43.0	29,123	44.08	5,835	42.71	8,729	39.81
Yes	57,813	56.9	36,849	55.77	7,797	57.07	13,167	60.06
Missing	161	0.2	102	0.15	31	0.23	28	0.13
Chronic lung disease								
None	76,803	75.5	50,632	76.63	9,756	71.40	16,415	74.87
Mild	12,157	12.0	7,658	11.59	1,853	13.56	2,646	12.07
Moderate	7,797	7.7	4,720	7.14	1,269	9.29	1,808	8.25
Severe	4,005	3.9	2,463	3.73	658	4.82	884	4.03
Missing	899	0.9	601	0.91	127	0.93	171	0.78

Table 1. Continued

Variable	Overall Valve + CABG (n = 101,661)		AVR + CABG (n = 66,074)		MVR + CABG (n = 13,663)		MVRRepair + CABG (n = 21,924)	
	N	%	N	%	N	%	N	%
Peripheral vascular disease								
No	84,183	82.8	54,658	82.72	11,373	83.24	18,152	82.80
Yes	17,294	17.0	11,296	17.10	2,267	16.59	3,731	17.02
Missing	184	0.2	120	0.18	23	0.17	41	0.19
Cerebrovascular disease								
No	83,284	81.9	53,509	80.98	11,304	82.73	18,471	84.25
Yes	18,202	17.9	12,449	18.84	2,335	17.09	3,418	15.59
Missing	175	0.2	116	0.18	24	0.18	35	0.16
CVA								
No CVA	92,527	91.0	60,141	91.02	12,283	89.90	20,103	91.69
Remote CVA (> 2 weeks)	8,461	8.3	5,545	8.39	1,240	9.08	1,676	7.64
Recent CVA ( $\leq$ 2 weeks)	348	0.3	184	0.28	88	0.64	76	0.35
CVA, missing timing	114	0.1	62	0.09	23	0.17	29	0.13
Missing	211	0.2	142	0.21	29	0.21	40	0.18
Endocarditis								
No endocarditis	99,517	97.9	65,023	98.41	12,914	94.52	21,580	98.43
Treated endocarditis	1,091	1.1	525	0.79	356	2.61	210	0.96
Active endocarditis	827	0.8	387	0.59	356	2.61	84	0.38
Endocarditis, missing type	24	0.0	11	0.02	8	0.06	5	0.02
Missing	202	0.2	128	0.19	29	0.21	45	0.21
Renal failure								
No	92,592	91.1	60,880	92.14	12,037	88.10	19,675	89.74
Yes	8,888	8.7	5,072	7.68	1,605	11.75	2,211	10.08
Missing	181	0.2	122	0.18	21	0.15	38	0.17
Renal function								
Creatinine < 1.0 mg/dL	30,178	29.7	20,297	30.72	3,672	26.88	6,209	28.32
Creatinine 1.00–1.49 mg/dL	52,008	51.2	34,054	51.54	6,758	49.46	11,196	51.07
Creatinine 1.50–1.99 mg/dL	11,469	11.3	7,151	10.82	1,732	12.68	2,586	11.80
Creatinine 2.00–2.49 mg/dL	2,711	2.7	1,554	2.35	498	3.64	659	3.01
Creatinine $\geq$ 2.5 mg/dL	1,602	1.6	844	1.28	319	2.33	439	2.00
Dialysis	2,443	2.4	1,364	2.06	482	3.53	597	2.72
Missing	1,250	1.2	810	1.23	202	1.48	238	1.09
Immunosuppressive treatment								
No	98,421	96.8	63,984	96.84	13,211	96.69	21,226	96.82
Yes	2,975	2.9	1,904	2.88	427	3.13	644	2.94
Missing	265	0.3	186	0.28	25	0.18	54	0.25
Previous CV interventions								
Previous CABG surgery								
No	91,657	90.2	59,583	90.18	12,057	88.25	20,017	91.30
Yes	9,615	9.5	6,257	9.47	1,540	11.27	1,818	8.29
Missing	389	0.4	234	0.35	66	0.48	89	0.41
Previous valve surgery								
No	98,737	97.1	64,265	97.26	12,794	93.64	21,678	98.88
Yes	2,540	2.5	1,567	2.37	813	5.95	160	0.73
Missing	384	0.4	242	0.37	56	0.41	86	0.39
Previous other cardiac surgery								
No	98,538	96.9	64,166	97.11	13,181	96.47	21,191	96.66
Yes	2,683	2.6	1,634	2.47	407	2.98	642	2.93
Missing	440	0.4	274	0.41	75	0.55	91	0.42
Number of previous CV surgeries								
No previous CV surgery	89,419	88.0	58,161	88.02	11,530	84.39	19,728	89.98
1 prior CV surgery	10,453	10.3	6,796	10.29	1,799	13.17	1,858	8.47
$\geq$ 2 prior CV surgeries	1,200	1.2	766	1.16	231	1.69	203	0.93
Missing	589	0.6	351	0.53	103	0.75	135	0.62

Table 1. Continued

Variable	Overall Valve + CABG (n = 101,661)		AVR + CABG (n = 66,074)		MVR + CABG (n = 13,663)		MVRRepair + CABG (n = 21,924)	
	N	%	N	%	N	%	N	%
Prior PCI								
No PCI	84,553	83.2	55,581	84.12	11,152	81.62	17,820	81.28
PCI ≤ 6 hours	315	0.3	151	0.23	89	0.65	75	0.34
PCI > 6 hours	16,158	15.9	9,946	15.05	2,321	16.99	3,891	17.75
PCI, missing timing	234	0.2	145	0.22	45	0.33	44	0.20
Missing	401	0.4	251	0.38	56	0.41	94	0.43
Preoperative cardiac status								
Acuity status								
Elective	62,298	61.3	43,682	66.11	7,277	53.26	11,339	51.72
Urgent	36,454	35.9	21,414	32.41	5,315	38.90	9,725	44.36
Emergent	2,479	2.4	763	1.15	945	6.92	771	3.52
Emergent salvage	258	0.3	97	0.15	104	0.76	57	0.26
Missing	172	0.2	118	0.18	22	0.16	32	0.15
MI								
No prior MI	68,332	67.2	49,673	75.18	8,056	58.96	10,603	48.36
MI ≤ 21 days	16,934	16.7	9,308	14.09	2,621	19.18	5,005	22.83
MI 8–21 days	3,751	3.7	1,725	2.61	624	4.57	1,402	6.39
MI 1–7 days	10,458	10.3	4,514	6.83	1,741	12.74	4,203	19.17
MI > 6 and < 24 hours	1,113	1.1	367	0.56	341	2.50	405	1.85
MI ≤ 6 hours	531	0.5	178	0.27	192	1.41	161	0.73
MI, missing timing	355	0.3	184	0.28	59	0.43	112	0.51
Missing	187	0.2	125	0.19	29	0.21	33	0.15
Angina								
No	42,542	41.8	28,032	42.43	6,248	45.73	8,262	37.68
Yes	58,967	58.0	37,945	57.43	7,394	54.12	13,628	62.16
Missing	152	0.1	97	0.15	21	0.15	34	0.16
Cardiogenic shock								
No	98,743	97.1	65,219	98.71	12,590	92.15	20,934	95.48
Yes	2,719	2.7	720	1.09	1,055	7.72	944	4.31
Missing	199	0.2	135	0.20	18	0.13	46	0.21
Resuscitation								
No	100,474	98.8	65,522	99.16	13,359	97.78	21,593	98.49
Yes	971	1.0	405	0.61	281	2.06	285	1.30
Missing	216	0.2	147	0.22	23	0.17	46	0.21
Arrhythmia								
No arrhythmia	83,856	82.5	56,040	84.81	9,992	73.13	17,824	81.30
AFib/flutter	13,386	13.2	7,533	11.40	2,940	21.52	2,913	13.29
Heart block	1,975	1.9	1,311	1.98	289	2.12	375	1.71
Sustained VT/VF	1,513	1.5	614	0.93	299	2.19	600	2.74
Arrhythmia, other	483	0.5	305	0.46	63	0.46	115	0.52
Arrhythmia, missing type	242	0.2	135	0.20	59	0.43	48	0.22
Missing	206	0.2	136	0.21	21	0.15	49	0.22
Preoperative IABP								
No	96,136	94.6	64,597	97.76	11,957	87.51	19,582	89.32
Yes	5,205	5.1	1,275	1.93	1,655	12.11	2,275	10.38
Missing	320	0.3	202	0.31	51	0.37	67	0.31
NYHA class								
I	9,839	9.7	6,934	10.49	1,103	8.07	1,802	8.22
II	24,830	24.4	17,808	26.95	2,524	18.47	4,498	20.52
III	42,593	41.9	28,079	42.50	5,458	39.95	9,056	41.31
IV	20,571	20.2	10,808	16.36	3,882	28.41	5,881	26.82
Missing	3,828	3.8	2,445	3.70	696	5.09	687	3.13

Table 1. Continued

Variable	Overall Valve + CABG (n = 101,661)		AVR + CABG (n = 66,074)		MVR + CABG (n = 13,663)		MVRRepair + CABG (n = 21,924)	
	N	%	N	%	N	%	N	%
Congestive heart failure								
No	58,086	57.1	41,984	63.54	5,797	42.43	10,305	47.00
Yes	43,377	42.7	23,953	36.25	7,845	57.42	11,579	52.81
Missing	198	0.2	137	0.21	21	0.15	40	0.18
Number of diseased coronary vessels								
None	2,362	2.3	1,786	2.70	281	2.06	295	1.35
One	22,718	22.3	16,934	25.63	3,040	22.25	2,744	12.52
Two	27,144	26.7	19,014	28.78	3,655	26.75	4,475	20.41
Three	49,060	48.3	28,107	42.54	6,623	48.47	14,330	65.36
Missing	377	0.4	233	0.35	64	0.47	80	0.36
Left main disease $\geq$ 50%								
No	84,025	82.7	55,292	83.68	11,503	84.19	17,230	78.59
Yes	17,175	16.9	10,512	15.91	2,072	15.17	4,591	20.94
Missing	461	0.5	270	0.41	88	0.64	103	0.47
Ejection fraction, %								
< 25	5,805	5.7	2,199	3.33	640	4.68	2,966	13.53
25–34	10,988	10.8	4,877	7.38	1,566	11.46	4,545	20.73
35–44	14,928	14.7	8,064	12.20	2,487	18.20	4,377	19.96
45–54	20,398	20.1	13,424	20.32	3,048	22.31	3,926	17.91
$\geq$ 55	43,556	42.8	32,973	49.90	5,209	38.12	5,374	24.51
Missing	5,986	5.9	4,537	6.87	713	5.22	736	3.36
Aortic stenosis								
No	42,831	42.1	8,527	12.91	12,974	94.96	21,330	97.29
Yes	58,317	57.4	57,319	86.75	535	3.92	463	2.11
Missing	513	0.5	228	0.35	154	1.13	131	0.60
Mitral stenosis								
No	95,696	94.1	63,862	96.65	11,166	81.72	20,668	94.27
Yes	4,993	4.9	1,542	2.33	2,366	17.32	1,085	4.95
Missing	972	1.0	670	1.01	131	0.96	171	0.78
Tricuspid stenosis								
No	100,093	98.5	65,060	98.47	13,402	98.09	21,631	98.66
Yes	275	0.3	154	0.23	57	0.42	64	0.29
Missing	1,293	1.3	860	1.30	204	1.49	229	1.04
Pulmonic stenosis								
No	99,484	97.9	64,693	97.91	13,348	97.69	21,443	97.81
Yes	122	0.1	85	0.13	14	0.10	23	0.10
Missing	2,055	2.0	1,296	1.96	301	2.20	458	2.09
Aortic insufficiency								
None	57,561	56.6	28,972	43.85	10,821	79.20	17,768	81.04
Trivial	9,243	9.1	6,573	9.95	1,023	7.49	1,647	7.51
Mild	13,828	13.6	11,082	16.77	1,156	8.46	1,590	7.25
Moderate	10,195	10.0	9,581	14.50	232	1.70	382	1.74
Severe	8,686	8.5	8,580	12.99	49	0.36	57	0.26
Missing	2,148	2.1	1,286	1.95	382	2.80	480	2.19
Mitral insufficiency								
None	41,756	41.1	38,790	58.71	1,297	9.49	1,669	7.61
Trivial	7,467	7.3	7,139	10.80	147	1.08	181	0.83
Mild	15,407	15.2	13,485	20.41	584	4.27	1,338	6.10
Moderate	14,987	14.7	4,842	7.33	2,790	20.42	7,355	33.55
Severe	20,516	20.2	527	0.80	8,743	63.99	11,246	51.30
Missing	1,528	1.5	1,291	1.95	102	0.75	135	0.62

Table 1. Continued

Variable	Overall Valve + CABG (n = 101,661)		AVR + CABG (n = 66,074)		MVR + CABG (n = 13,663)		MVRRepair + CABG (n = 21,924)	
	N	%	N	%	N	%	N	%
Tricuspid insufficiency								
None	74,774	73.6	49,614	75.09	9,758	71.42	15,402	70.25
Trivial	7,972	7.8	5,454	8.25	839	6.14	1,679	7.66
Mild	11,505	11.3	7,060	10.68	1,631	11.94	2,814	12.84
Moderate	4,119	4.1	1,919	2.90	874	6.40	1,326	6.05
Severe	636	0.6	237	0.36	186	1.36	213	0.97
Missing	2,655	2.6	1,790	2.71	375	2.74	490	2.23
Pulmonic insufficiency								
None	91,715	90.2	59,891	90.64	12,275	89.84	19,549	89.17
Trivial	3,411	3.4	2,122	3.21	442	3.24	847	3.86
Mild	2,065	2.0	1,215	1.84	306	2.24	544	2.48
Moderate	326	0.3	165	0.25	70	0.51	91	0.42
Severe	49	0.0	25	0.04	11	0.08	13	0.06
Missing	4,095	4.0	2,656	4.02	559	4.09	880	4.01

AFib = atrial fibrillation; AVR = aortic valve replacement; CABG = coronary artery bypass graft; CV = cardiovascular; CVA = cerebrovascular accident (stroke); IABP = intra-aortic balloon pump; MI = myocardial infarction; MVR = mitral valve replacement; MVRRepair = mitral valve repair; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; VF = ventricular fibrillation; VT = ventricular tachycardia.

endocarditis) in the overall valve plus CABG population, it was included for consistency with the isolated valve model. Active endocarditis was present in 2.6% of patients undergoing mitral replacement plus CABG. (3) Mitral stenosis was rarely present among isolated CABG patients (0.35%). However, it was not uncommon (4.9%) among patients undergoing valve plus CABG surgery and was included as a candidate variable. It was present in 17.3% of mitral replacements and 5.0% of mitral repairs.

An indicator for valve procedure (AVR, MVR, MVRRepair) was included in the combined valve plus CABG model, as previously noted.

### Missing Data

Missing data are uncommon in the STS NCD, with a frequency of less than 1% missing for most variables. Model variables with more than 1% missing were ejection fraction

Table 2. Frequency of Endpoints in Overall Study Population 2002 to 2006

	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Overall (AVR + CABG, MVR + CABG, MVRRepair + CABG)									
N	101,661	101,661	99,218	101,661	101,661	101,661	101,661	101,661	101,661
Events	6,919	2,935	9,097	21,561	684	12,117	30,580	15,594	22,534
%	6.8	2.9	9.0	21.2	0.7	11.9	30.1	15.3	22.2
AVR + CABG									
N	66,074	66,074	64,710	66,074	66,074	66,074	66,074	66,074	66,074
Events	3,718	1,751	5,032	11,608	394	7,090	17,343	8,412	16,961
%	5.6	2.7	7.6	17.6	0.6	10.7	26.3	12.7	25.7
MVR + CABG									
N	13,663	13,663	13,181	13,663	13,663	13,663	13,663	13,663	13,663
Events	1,590	499	1,829	4,469	114	2,274	5,897	3,277	1,512
%	11.6	3.7	13.6	32.7	0.8	16.6	43.2	24.0	11.1
MVRRepair + CABG									
N	21,924	21,924	21,327	21,924	21,924	21,924	21,924	21,924	21,924
Events	1,611	685	2,236	5,484	176	2,753	7,340	3,905	4,061
%	7.4	3.1	10.3	25.0	0.8	12.6	33.5	17.8	18.5

AVR = aortic valve replacement; CABG = coronary artery bypass graft surgery; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); DS WI = deep sternal wound infection; Mort = mortality; MVR = mitral valve replacement; MVRRepair = mitral valve repair; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.



Table 3. List of Candidate Variables and Their Coding for STS Valve Plus CABG Models

Candidate Variables	Coding
<b>Continuous variables</b>	
Age <sup>a</sup>	Linear spline truncated from below at 50 with knot at 75.
Ejection fraction	Linear; values > 50 mapped to 50
Body surface area <sup>a</sup>	Quadratic polynomial modeled separately for males and females. <i>Note: BSA &lt; 1.4 and &gt; 2.6 were mapped to those values, respectively.</i>
Creatinine	Linear (only for patients not on dialysis). <i>Note: Creatinine &lt; 0.5 and &gt; 5.0 mapped to those values, respectively.</i>
Time trend <sup>a</sup>	Ordinal categorical variable with separate category for each 6-month harvest interval. Modeled as linear across the categories.
<b>Binary variables</b>	
Active infectious endocarditis	Yes/no
Dialysis	Yes/no
Preoperative atrial fibrillation	Yes/no
Shock	Yes/no
Female <sup>a</sup>	Yes/no
Hypertension	Yes/no
Immunosuppressive treatment	Yes/no
Preop IABP or inotropes	Yes/no
Peripheral vascular disease	Yes/no
Unstable angina (no MI < 7 days)	Yes/no
Left main disease	Yes/no
Aortic stenosis	Yes/no
Mitral stenosis	Yes/no
Aortic insufficiency	Defined as at least moderate (yes/no)
Mitral insufficiency	Defined as at least moderate (yes/no)
Tricuspid insufficiency	Defined as at least moderate (yes/no)
<b>Categorical variables</b>	
Surgery type	3 groups: AVR + CABG, MVR + CABG, MVRRepair + CABG
Chronic lung disease	Modeled as linear across categories (none, mild, moderate, severe)
CVD/CVA	3 groups: no CVD, CVD no CVA, CVD + CVA
Diabetes mellitus	3 groups: insulin diabetes, noninsulin diabetes, other or no diabetes
No. diseased coronary vessels	3 groups: < 2-vessel disease; 2-vessel disease; 3-vessel disease. Modeled as linear across the categories
MI	3 groups: < 24 hours, 1–21 days, > 21 days or no MI. <i>Note: groups 1 and 2 were subsequently collapsed for some models.</i>
Race	3 groups: black, Hispanic, other including Caucasian
Status	4 groups: elective, urgent, emergent no resuscitation, salvage or emergent with resuscitation
Previous cardiovascular operations	3 groups: 0 previous, 1 previous, ≥ 2 previous
CHF and NYHA class	3 groups: no CHF, CHF not NYHA IV, CHF and NYHA IV
<b>Interaction terms</b>	
Age by reoperation <sup>a</sup>	
Age by emergent status <sup>a</sup>	
Surgery type by each of the following:	Age, diabetes, dialysis, creatinine, reoperation, endocarditis, emergent status, CLD, CHF, EF, sex, shock, IABP/inotropes, mitral insufficiency, aortic insufficiency, mitral stenosis, aortic stenosis.

<sup>a</sup> These variables were forced into each model.

AVR = aortic valve replacement; CHF = congestive heart failure; CLD = chronic lung disease; CVA = cerebrovascular accident (stroke); CVD = cardiovascular disease; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; MVR = mitral valve replacement; MVRRepair = mitral valve repair; NYHA = New York Heart Association.

(5.9%), New York Heart Association functional class (3.8%), tricuspid insufficiency (2.6%), aortic insufficiency (2.1%), mitral insufficiency (1.5%), and creatinine/dialysis (1.2%).

To make full use of the available data, binary risk factors were modeled as yes versus no or missing. Thus, missing

values were analyzed as if the endpoint did not occur. Missing data on categorical variables were imputed to the lowest risk value, which, in most instances, was the mode. Missing data on continuous variables were imputed to the conditional median. For ejection fraction, we conditioned

on congestive heart failure and sex. For body surface area, we conditioned on sex. For serum creatinine, we conditioned on renal failure.

Although multiple imputation is generally preferred on statistical grounds [21], we chose single imputation for this analysis based largely on practical considerations, including computational intensity. Furthermore, the fraction of missing data was small, and single and multiple imputation would give similar results. Finally, multiple imputation is primarily used for calculating appropriate standard error estimates, but an adjustment to the standard errors would not impact our study results or the published risk algorithms. In a separate sensitivity analysis, we compared predicted risk estimates from our final models to risk estimates that were derived from analogous models using multiple instead of single imputation. For each endpoint, the relative difference in predicted risk was less than 6% (eg, an absolute difference of 5.0% versus 5.3%) for all patients in the development and validation samples, and it was less than 2% (eg, an absolute difference of 5.0% versus 5.1%) for 99% of patients. A summary of these analyses including regression coefficients and covariance matrices is available at [www.sts.org/riskmodels](http://www.sts.org/riskmodels).

### Final Variable Selection Procedure

Variables were initially selected using an automated stepwise model selection algorithm. The stepwise procedure began with a model that included all of the candidate variables except for interaction terms. Age, body surface area, and month of surgery were forced into each model. As in the isolated CABG and isolated valve models described in Parts 1 and 2 of this series, month of surgery was used only to adjust for time trends in the frequency of adverse outcomes over the 5-year study period. We adjusted for this to reduce potential confounding by time trends when estimating regression coefficients for the variables that are of primary interest (ie, patient preoperative risk factors—see example in Part 1). Surgery date was categorized into 6-month intervals and modeled as a linear trend across the ordinal categories. Surgery date is not included in the final risk prediction algorithm, and a patient's predicted risk does not depend on it. The published intercept parameter has been adjusted to incorporate the time trend, and this adjusted intercept reflects the baseline risk for a reference period of July to December 2006.

Other variables were selected in a stepwise fashion using a significance criterion of 0.05 for entry and removal. Ordinal categorical variables were initially coded such that removing an indicator variable caused a category to be combined with the lowest risk category (the reference group). In the case of myocardial infarction (MI), there were two outcomes (permanent stroke, prolonged length of stay) in which "MI 1 to 21 days" was retained but "MI less than 24 hours" was removed. For these two cases, the two MI categories were replaced by the single category "MI 21 days or less." The stepwise procedure was performed separately for each endpoint. Multiple interaction terms consisting of predictor variable and surgery type were also evaluated, and two additional interaction terms (age by reoperation

and age by emergent status) were forced into the models (see Tables 3 and 5).

The results of this initial selection process were then reviewed by surgeon members of the QMTF for face validity and consistency with previous STS or other valve models: (1) preoperative atrial fibrillation was forced into the model for permanent stroke; (2) an indicator variable for dialysis was forced into any model that included creatinine (this did not apply to the renal failure model, as patients with preoperative dialysis were excluded); (3) sex was forced into all models; and (4) each variable that interacted with surgery group was also included as a main effect.

After validating the model in the 40% validation sample, the development and validation samples were then rejoined, and the final model coefficients were estimated using the overall 100% combined sample. The final logistic regression model was estimated using generalized estimating equations with empirical (sandwich) standard error estimates to account for clustering of patients within institutions [22]. An independence working correlation matrix was used to apply the generalized estimating equations. With this approach, the estimated regression coefficients were identical to those obtained using ordinary logistic regression, but the standard errors were adjusted to account for the clustered data structure.

### Results

#### *Risk Factors, Outcomes, and Predictor Variables*

Table 1 presents the distribution of risk factors and endpoints in the overall 2002 to 2006 study population. Because there are three valve plus CABG categories, space limitations prevent display of the bivariate relationships for each predictor variable, endpoint, and valve plus CABG group. These are available upon request from STS.

Table 2 summarizes the overall frequency of adverse outcomes as well as the outcomes for the three major valve groups. Table 3 lists the candidate predictor variables and their coding schemes.

#### *Assessment of Model Fit and Discrimination*

The Hosmer-Lemeshow test was not employed to assess overall calibration. Large sample sizes make a significant *p* value almost inevitable, as all risk models are only approximations of reality [23]. Rather, we assessed calibration graphically by plotting observed versus predicted event rates within deciles of predicted risk in the development and validation samples (Fig 1). These plots were constructed for the overall sample and for subgroups based on surgery type (AVR plus CABG, MVR plus CABG, MVR-repair plus CABG); age (< 60, 60 to 79, ≥ 80 years); sex (male, female); diabetes mellitus (yes/no); status (elective, non-elective); and ejection fraction (≤ 40, > 40). Because of space constraints, only the overall sample results in the validation sample are presented. Additional results are available at [www.sts.org/riskmodels](http://www.sts.org/riskmodels).

In general, the models were well calibrated in the validation sample. The average absolute difference between observed versus predicted event rates across the decile categories ranged from 0.1% for deep sternal wound infec-

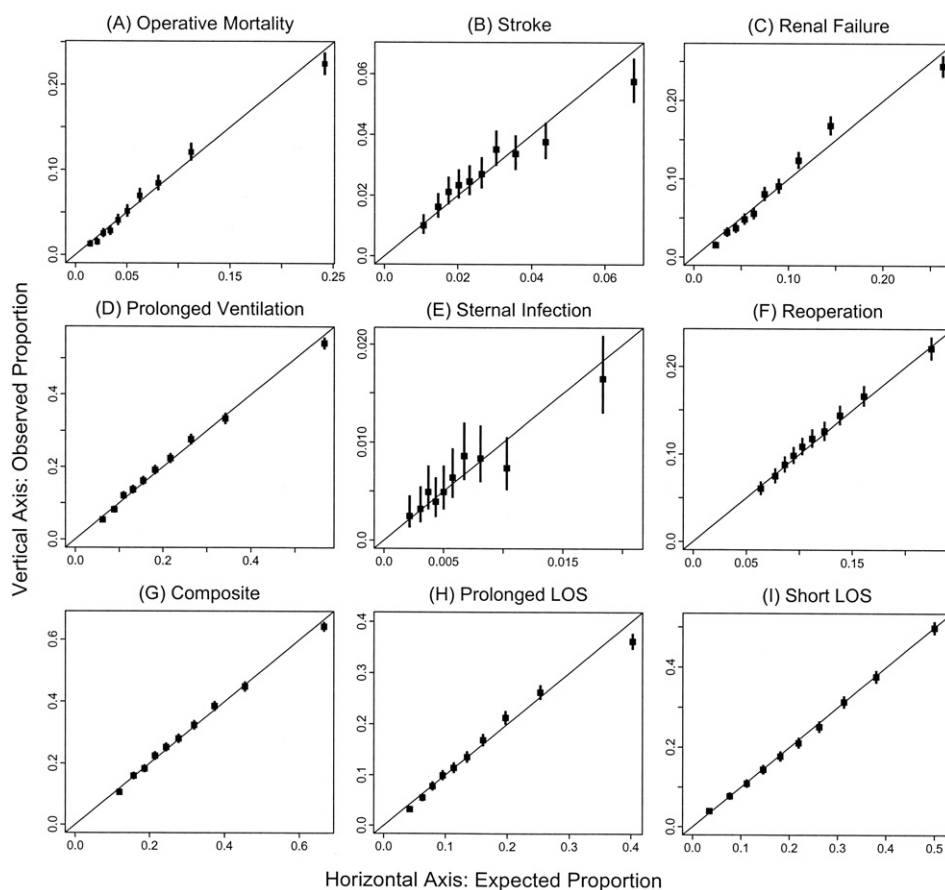


Fig 1. Plots of observed (O) versus expected (E) in validation sample

tion to 0.96% for prolonged length of stay. There was a very slight tendency for the models to overpredict risk in the highest decile. Although perfect prediction would be ideal, a slight overprediction implies that the model will give adequate credit to surgeons who take on patients with several model risk factors.

Discrimination was assessed by determining the c-

statistic, also known as the area under the receiver operating characteristic (ROC) curve. Table 4 presents the discrimination of the various models. In the validation sample, the c-index of the overall valve plus CABG operative mortality model was 0.750, and the c-indices of the morbidity models ranged from 0.617 for reoperation to 0.724 for renal failure and short length of stay.

Table 4. Discrimination of Models (C-Index) in Development and Validation Samples

	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
<b>Overall</b>									
Development sample	0.754	0.656	0.729	0.730	0.670	0.623	0.704	0.719	0.726
Validation sample	0.750	0.622	0.724	0.720	0.646	0.617	0.698	0.710	0.724
<b>AVR + CABG</b>									
Development sample	0.737	0.648	0.720	0.706	0.639	0.607	0.678	0.705	0.700
Validation sample	0.736	0.609	0.718	0.697	0.657	0.604	0.673	0.699	0.698
<b>MVR + CABG</b>									
Development sample	0.764	0.665	0.712	0.746	0.713	0.608	0.725	0.694	0.726
Validation sample	0.739	0.611	0.701	0.733	0.580	0.599	0.714	0.680	0.733
<b>MVRepair + CABG</b>									
Development sample	0.746	0.650	0.727	0.725	0.692	0.624	0.707	0.712	0.738
Validation sample	0.755	0.652	0.715	0.716	0.644	0.623	0.705	0.702	0.733

AVR = aortic valve replacement; CABG = coronary artery bypass graft; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); DSWI = deep sternal wound infection; Mort = mortality; MVR = mitral valve replacement; MVRepair = mitral valve repair; PLOS = prolonged length of stay; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

Table 5. Estimated Odds Ratios for CABG Mortality, Morbidity, and Length of Stay Models

A. Odds ratios for variables that do not interact with surgery group									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Preoperative AFib	1.20 (1.12, 1.29)	1.05 (0.94, 1.17)	1.18 (1.11, 1.26)	1.13 (1.07, 1.19)	NA	1.16 (1.10, 1.22)	1.15 (1.10, 1.20)	1.22 (1.15, 1.28)	0.71 (0.67, 0.75)
BSA 1.6 versus 2.0 among females	1.29 (1.19, 1.39)	1.34 (1.18, 1.52)	0.87 (0.81, 0.94)	1.08 (1.02, 1.14)	0.51 (0.39, 0.67)	1.13 (1.07, 1.23)	1.12 (1.07, 1.18)	0.97 (0.92, 1.03)	1.03 (0.96, 1.10)
BSA 1.6 versus 2.0 among males	1.58 (1.41, 1.77)	1.38 (1.17, 1.64)	1.18 (1.07, 1.31)	1.31 (1.21, 1.41)	0.71 (0.49, 1.03)	1.18 (1.12, 1.34)	1.32 (1.24, 1.41)	1.40 (1.29, 1.52)	0.81 (0.75, 0.88)
BSA 1.8 versus 2.0 among females	1.05 (1.00, 1.10)	1.16 (1.06, 1.26)	0.89 (0.85, 0.93)	0.99 (0.95, 1.02)	0.69 (0.61, 0.77)	1.03 (0.98, 1.06)	1.01 (0.98, 1.04)	0.94 (0.90, 0.97)	1.08 (1.04, 1.12)
BSA 1.8 versus 2.0 among males	1.15 (1.10, 1.20)	1.13 (1.07, 1.20)	1.01 (0.97, 1.05)	1.06 (1.03, 1.09)	0.83 (0.72, 0.95)	1.06 (1.04, 1.11)	1.07 (1.05, 1.10)	1.09 (1.06, 1.12)	0.96 (0.94, 0.99)
BSA 2.2 versus 2.0 among females	1.12 (1.02, 1.22)	0.87 (0.74, 1.02)	1.25 (1.15, 1.35)	1.13 (1.06, 1.20)	1.57 (1.32, 1.89)	1.04 (1.00, 1.17)	1.10 (1.04, 1.17)	1.19 (1.11, 1.27)	0.82 (0.76, 0.89)
BSA 2.2 versus 2.0 among males	1.04 (1.00, 1.09)	0.95 (0.90, 1.01)	1.15 (1.11, 1.18)	1.09 (1.06, 1.11)	1.25 (1.14, 1.37)	1.00 (0.95, 1.01)	1.07 (1.04, 1.09)	1.09 (1.06, 1.12)	0.91 (0.89, 0.93)
CVD with CVA	1.22 (1.11, 1.33)	1.72 (1.52, 1.95)	1.12 (1.04, 1.22)	1.27 (1.19, 1.34)	1.22 (0.95, 1.56)	1.12 (1.04, 1.20)	1.26 (1.20, 1.33)	1.26 (1.18, 1.35)	0.75 (0.70, 0.81)
CVD without CVA	NA	1.28 (1.13, 1.45)	1.14 (1.06, 1.23)	1.10 (1.04, 1.16)	NA	NA	1.11 (1.05, 1.17)	1.11 (1.05, 1.18)	0.85 (0.78, 0.92)
Diabetes, insulin	1.31 (1.20, 1.42)	1.16 (1.03, 1.30)	1.62 (1.52, 1.74)	1.32 (1.25, 1.40)	1.98 (1.59, 2.46)	NA	1.34 (1.28, 1.41)	1.49 (1.40, 1.58)	0.67 (0.62, 0.72)
Diabetes, noninsulin	1.12 (1.05, 1.19)	1.16 (1.06, 1.26)	1.28 (1.21, 1.35)	1.11 (1.07, 1.15)	1.30 (1.10, 1.54)	NA	1.12 (1.08, 1.16)	1.17 (1.12, 1.22)	0.84 (0.81, 0.88)
No. diseased coronary vessels (2 versus 1 or 3 versus 2)	1.15 (1.11, 1.19)	1.20 (1.14, 1.26)	1.17 (1.14, 1.21)	1.19 (1.16, 1.22)	1.28 (1.15, 1.42)	1.09 (1.06, 1.11)	1.16 (1.14, 1.18)	1.13 (1.10, 1.16)	0.82 (0.81, 0.84)
Hypertension	NA	1.19 (1.08, 1.31)	1.25 (1.18, 1.33)	1.10 (1.05, 1.15)	1.33 (1.09, 1.63)	NA	1.12 (1.08, 1.16)	1.08 (1.03, 1.13)	0.92 (0.88, 0.96)
Immunosuppressive treatment	1.35 (1.17, 1.54)	NA	1.30 (1.15, 1.47)	1.28 (1.17, 1.40)	NA	1.27 (1.14, 1.42)	1.26 (1.16, 1.37)	1.22 (1.11, 1.34)	0.75 (0.67, 0.84)
Left main disease	1.12 (1.05, 1.20)	NA	NA	1.06 (1.02, 1.11)	NA	NA	NA	NA	NA
Mitral insufficiency, moderate/severe	NA	NA	NA	NA	NA	NA	1.07 (1.01, 1.12)	NA	NA
Tricuspid insufficiency, moderate/severe	1.27 (1.15, 1.41)	NA	1.25 (1.13, 1.38)	1.15 (1.06, 1.24)	NA	NA	1.14 (1.07, 1.22)	NA	0.79 (0.69, 0.92)
Peripheral vascular disease	1.29 (1.21, 1.37)	1.15 (1.04, 1.27)	1.16 (1.10, 1.23)	1.18 (1.12, 1.24)	NA	1.15 (1.09, 1.22)	1.20 (1.15, 1.25)	1.16 (1.11, 1.22)	NA
Mitral stenosis	1.10 (0.99, 1.24)	NA	NA	NA	NA	NA	NA	1.09 (1.00, 1.18)	NA
MI 1–21 days	1.19 (1.10, 1.28)	NA	1.18 (1.10, 1.26)	1.28 (1.21, 1.35)	NA	NA	1.22 (1.16, 1.28)	NA	NA
MI ≤ 21 days <sup>a</sup>	NA	1.22 (1.11, 1.34)	NA	NA	NA	NA	NA	1.16 (1.10, 1.22)	NA
MI < 24 hrs	1.65 (1.42, 1.91)	NA	1.30 (1.10, 1.54)	1.41 (1.23, 1.62)	NA	1.15 (1.00, 1.32)	1.49 (1.30, 1.70)	NA	NA
Time trend per 6-month harvest interval	0.98 (0.96, 0.99)	0.98 (0.97, 1.00)	1.01 (1.00, 1.02)	1.01 (1.00, 1.02)	0.96 (0.93, 0.99)	0.99 (0.98, 1.00)	1.00 (0.99, 1.01)	1.01 (1.00, 1.02)	1.00 (0.99, 1.01)
Race black	NA	NA	1.15 (1.03, 1.30)	1.31 (1.19, 1.44)	NA	1.19 (1.06, 1.33)	1.21 (1.11, 1.32)	1.31 (1.19, 1.44)	0.65 (0.58, 0.72)
Race Hispanic	NA	NA	1.20 (1.03, 1.40)	1.17 (1.03, 1.32)	NA	1.08 (0.94, 1.24)	1.15 (1.03, 1.28)	1.13 (0.98, 1.30)	0.85 (0.71, 1.02)
Status, urgent versus elective	1.25 (1.17, 1.34)	NA	1.18 (1.10, 1.26)	1.26 (1.19, 1.33)	NA	1.14 (1.07, 1.21)	1.19 (1.14, 1.25)	1.28 (1.22, 1.35)	0.77 (0.72, 0.81)
Unstable angina	1.11 (1.03, 1.21)	0.89 (0.80, 1.00)	1.12 (1.05, 1.20)	1.06 (0.99, 1.13)	NA	NA	NA	NA	NA

Table 5. Continued

B. Odds ratios for AVR plus CABG									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50 (no reop, elective)	1.29 (1.20, 1.39)	1.28 (1.19, 1.38)	1.39 (1.32, 1.45)	1.23 (1.20, 1.27)	1.06 (0.92, 1.21)	1.19 (1.15, 1.23)	1.20 (1.16, 1.24)	1.37 (1.32, 1.42)	0.74 (0.72, 0.77)
Age 70 versus 50 (no reop, elective)	1.67 (1.45, 1.92)	1.64 (1.42, 1.91)	1.92 (1.75, 2.11)	1.52 (1.43, 1.62)	1.11 (0.85, 1.46)	1.41 (1.31, 1.51)	1.44 (1.36, 1.54)	1.86 (1.73, 2.01)	0.55 (0.52, 0.59)
Age 80 versus 50 (no reop, elective)	2.47 (2.08, 2.94)	2.03 (1.71, 2.42)	2.76 (2.47, 3.08)	1.96 (1.82, 2.11)	1.12 (0.82, 1.53)	1.67 (1.54, 1.82)	1.86 (1.73, 2.01)	2.67 (2.46, 2.91)	0.33 (0.30, 0.36)
CHF, not NYHA IV	1.24 (1.14, 1.34)	0.98 (0.88, 1.09)	1.19 (1.11, 1.28)	1.22 (1.16, 1.29)	NA	NA	1.14 (1.08, 1.19)	1.30 (1.23, 1.38)	0.84 (0.79, 0.89)
CHF, NYHA IV	1.48 (1.34, 1.64)	1.15 (1.00, 1.32)	1.35 (1.24, 1.48)	1.47 (1.36, 1.59)	NA	1.16 (1.08, 1.24)	1.36 (1.27, 1.45)	1.49 (1.39, 1.60)	0.73 (0.66, 0.82)
Creatinine per 1 unit	1.57 (1.49, 1.65)	1.27 (1.18, 1.36)	2.26 (2.13, 2.40)	1.46 (1.41, 1.52)	NA	1.28 (1.23, 1.34)	1.67 (1.60, 1.74)	1.51 (1.45, 1.58)	0.62 (0.58, 0.67)
Dialysis vs no dialysis and creatinine = 1.0	3.20 (2.84, 3.61)	1.42 (1.17, 1.73)	NA	2.27 (2.06, 2.51)	NA	1.65 (1.41, 1.92)	2.09 (1.91, 2.30)	2.42 (2.19, 2.67)	0.30 (0.25, 0.37)
EF per 10-unit decrease	1.10 (1.06, 1.15)	NA	1.06 (1.03, 1.08)	1.12 (1.10, 1.14)	NA	1.08 (1.05, 1.10)	1.11 (1.09, 1.13)	1.10 (1.08, 1.13)	0.87 (0.84, 0.89)
Preoperative IABP/inotropes	1.43 (1.30, 1.58)	NA	1.27 (1.15, 1.39)	2.18 (2.01, 2.36)	NA	1.16 (1.06, 1.27)	1.76 (1.63, 1.90)	1.41 (1.25, 1.58)	0.56 (0.48, 0.65)
Shock	1.68 (1.45, 1.94)	1.19 (0.94, 1.50)	1.17 (0.92, 1.50)	1.93 (1.72, 2.16)	NA	1.24 (1.09, 1.41)	1.79 (1.50, 2.15)	1.45 (1.29, 1.63)	NA
Female versus male (at BSA = 1.8)	1.36 (1.26, 1.47)	1.19 (1.07, 1.32)	1.18 (1.10, 1.26)	1.52 (1.44, 1.61)	1.11 (0.88, 1.40)	0.92 (0.87, 0.97)	1.20 (1.15, 1.26)	1.31 (1.24, 1.38)	0.61 (0.57, 0.64)
Active infectious endocarditis	2.04 (1.66, 2.50)	1.83 (1.37, 2.46)	1.52 (1.21, 1.91)	1.96 (1.69, 2.27)	NA	1.56 (1.28, 1.91)	2.11 (1.83, 2.44)	1.81 (1.41, 2.32)	0.28 (0.20, 0.38)
CLD (moderate vs mild or severe vs moderate)	1.19 (1.16, 1.23)	NA	1.12 (1.09, 1.15)	1.26 (1.22, 1.30)	1.32 (1.22, 1.42)	1.10 (1.07, 1.13)	1.18 (1.15, 1.21)	1.26 (1.22, 1.30)	0.83 (0.80, 0.85)
Reop, 1 previous operation <sup>b</sup>	2.20 (1.81, 2.67)	NA	1.29 (1.08, 1.55)	1.83 (1.58, 2.11)	NA	1.39 (1.16, 1.67)	1.50 (1.32, 1.69)	1.55 (1.33, 1.81)	0.67 (0.58, 0.77)
Reop, ≥ 2 previous operations <sup>b</sup>	2.46 (1.87, 3.24)	NA	1.47 (1.15, 1.89)	2.19 (1.80, 2.65)	NA	1.48 (1.15, 1.92)	1.77 (1.51, 2.06)	1.65 (1.34, 2.03)	0.53 (0.43, 0.65)
Status emergent, no resuscitation <sup>b</sup>	2.14 (1.62, 2.81)	2.21 (1.45, 3.37)	1.77 (1.31, 2.37)	2.71 (2.14, 3.44)	NA	1.41 (1.16, 1.70)	2.17 (1.74, 2.72)	2.72 (2.19, 3.38)	0.33 (0.22, 0.50)
Status emergent, with resuscitation or salvage <sup>b</sup>	4.56 (3.31, 6.29)	2.60 (1.53, 4.43)	1.86 (1.30, 2.65)	2.12 (1.54, 2.92)	NA	NA	3.34 (2.43, 4.61)	1.76 (1.31, 2.37)	0.18 (0.09, 0.34)



Table 5. Continued

C. Odds ratios for MVR plus CABG									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50 (no reop, elective)	1.51 (1.39, 1.64)	1.28 (1.19, 1.38)	1.39 (1.32, 1.45)	1.23 (1.20, 1.27)	1.06 (0.92, 1.21)	1.19 (1.15, 1.23)	1.27 (1.21, 1.32)	1.37 (1.32, 1.42)	0.68 (0.64, 0.72)
Age 70 versus 50 (no reop, elective)	2.28 (1.94, 2.68)	1.64 (1.42, 1.91)	1.92 (1.75, 2.11)	1.52 (1.43, 1.62)	1.11 (0.85, 1.46)	1.41 (1.31, 1.51)	1.60 (1.47, 1.75)	1.86 (1.73, 2.01)	0.46 (0.41, 0.52)
Age 80 versus 50 (no reop, elective)	3.95 (3.17, 4.93)	2.03 (1.71, 2.42)	2.76 (2.47, 3.08)	1.96 (1.82, 2.11)	1.12 (0.82, 1.53)	1.67 (1.54, 1.82)	2.18 (1.92, 2.48)	2.67 (2.46, 2.91)	0.25 (0.21, 0.30)
CHF, not NYHA IV	0.91 (0.80, 1.03)	0.80 (0.64, 0.99)	0.92 (0.82, 1.03)	1.02 (0.93, 1.11)	NA	NA	0.94 (0.87, 1.02)	1.03 (0.94, 1.12)	0.84 (0.79, 0.89)
CHF, NYHA IV	1.09 (0.95, 1.24)	0.93 (0.75, 1.17)	1.04 (0.92, 1.19)	1.22 (1.10, 1.35)	NA	1.16 (1.08, 1.24)	1.13 (1.03, 1.23)	1.17 (1.06, 1.30)	0.73 (0.66, 0.82)
Creatinine per 1 unit	1.57 (1.49, 1.65)	1.27 (1.18, 1.36)	1.82 (1.66, 2.01)	1.46 (1.41, 1.52)	NA	1.28 (1.23, 1.34)	1.67 (1.60, 1.74)	1.51 (1.45, 1.58)	0.66 (0.57, 0.78)
Dialysis vs no dialysis and creatinine = 1.0	3.20 (2.84, 3.61)	1.42 (1.17, 1.73)	NA	2.27 (2.06, 2.51)	NA	1.21 (0.95, 1.55)	2.09 (1.91, 2.30)	2.42 (2.19, 2.67)	0.30 (0.18, 0.48)
EF per 10-unit decrease	1.23 (1.16, 1.30)	NA	1.06 (1.03, 1.08)	1.12 (1.10, 1.14)	NA	1.08 (1.05, 1.10)	1.11 (1.09, 1.13)	1.10 (1.08, 1.13)	0.89 (0.82, 0.95)
Preoperative IABP/inotropes	1.43 (1.30, 1.58)	NA	1.27 (1.15, 1.39)	2.18 (2.01, 2.36)	NA	1.16 (1.06, 1.27)	1.76 (1.63, 1.90)	1.29 (1.14, 1.46)	0.51 (0.39, 0.65)
Shock	1.68 (1.45, 1.94)	1.19 (0.94, 1.50)	1.21 (0.97, 1.50)	1.93 (1.72, 2.16)	NA	1.24 (1.09, 1.41)	2.76 (2.22, 3.42)	1.45 (1.29, 1.63)	NA
Female versus male (at BSA = 1.8)	1.36 (1.26, 1.47)	1.19 (1.07, 1.32)	1.18 (1.10, 1.26)	1.17 (1.08, 1.28)	1.11 (0.88, 1.40)	0.92 (0.87, 0.97)	1.20 (1.15, 1.26)	1.31 (1.24, 1.38)	0.66 (0.59, 0.74)
Active infectious endocarditis	2.04 (1.66, 2.50)	1.83 (1.37, 2.46)	1.52 (1.21, 1.91)	1.96 (1.69, 2.27)	NA	1.56 (1.28, 1.91)	2.11 (1.83, 2.44)	2.08 (1.62, 2.67)	0.28 (0.20, 0.38)
CLD (moderate vs mild or severe vs moderate)	1.19 (1.16, 1.23)	NA	1.12 (1.09, 1.15)	1.18 (1.12, 1.24)	1.32 (1.22, 1.42)	1.10 (1.07, 1.13)	1.18 (1.15, 1.21)	1.20 (1.14, 1.26)	0.83 (0.80, 0.85)
Reop, 1 previous operation <sup>b</sup>	2.20 (1.81, 2.67)	NA	1.29 (1.08, 1.55)	1.38 (1.19, 1.61)	NA	1.15 (0.95, 1.38)	1.50 (1.32, 1.69)	1.30 (1.10, 1.53)	0.81 (0.66, 0.99)
Reop, ≥ 2 previous operations <sup>b</sup>	2.46 (1.87, 3.24)	NA	1.47 (1.15, 1.89)	1.66 (1.35, 2.03)	NA	1.22 (0.95, 1.56)	1.77 (1.51, 2.06)	1.38 (1.12, 1.71)	0.64 (0.50, 0.82)
Status emergent, no resuscitation <sup>b</sup>	2.14 (1.62, 2.81)	2.21 (1.45, 3.37)	1.77 (1.31, 2.37)	2.71 (2.14, 3.44)	NA	1.41 (1.16, 1.70)	2.17 (1.74, 2.72)	2.72 (2.19, 3.38)	0.26 (0.16, 0.43)
Status emergent, with resuscitation or salvage <sup>b</sup>	4.56 (3.31, 6.29)	2.60 (1.53, 4.43)	1.86 (1.30, 2.65)	2.12 (1.54, 2.92)	NA	NA	3.34 (2.43, 4.61)	1.76 (1.31, 2.37)	0.14 (0.07, 0.27)

Table 5. Continued

D. Odds ratios for MVRepair plus CABG									
Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Age 60 versus 50 (no reop, elective)	1.46 (1.36, 1.57)	1.28 (1.19, 1.38)	1.39 (1.32, 1.45)	1.23 (1.20, 1.27)	1.06 (0.92, 1.21)	1.19 (1.15, 1.23)	1.28 (1.23, 1.33)	1.37 (1.32, 1.42)	0.66 (0.63, 0.69)
Age 70 versus 50 (no reop, elective)	2.14 (1.86, 2.46)	1.64 (1.42, 1.91)	1.92 (1.75, 2.11)	1.52 (1.43, 1.62)	1.11 (0.85, 1.46)	1.41 (1.31, 1.51)	1.63 (1.51, 1.76)	1.86 (1.73, 2.01)	0.44 (0.40, 0.48)
Age 80 versus 50 (no reop, elective)	3.60 (2.97, 4.33)	2.03 (1.71, 2.42)	2.76 (2.47, 3.08)	1.96 (1.82, 2.11)	1.12 (0.82, 1.53)	1.67 (1.54, 1.82)	2.23 (2.00, 2.49)	2.67 (2.46, 2.91)	0.23 (0.20, 0.27)
CHF, not NYHA IV	0.96 (0.85, 1.09)	1.05 (0.90, 1.23)	0.99 (0.88, 1.10)	1.10 (1.02, 1.19)	NA	NA	1.06 (0.99, 1.14)	1.17 (1.08, 1.26)	0.84 (0.79, 0.89)
CHF, NYHA IV	1.16 (1.02, 1.32)	1.23 (1.04, 1.46)	1.12 (0.99, 1.27)	1.32 (1.21, 1.44)	NA	1.16 (1.08, 1.24)	1.27 (1.17, 1.37)	1.33 (1.22, 1.45)	0.73 (0.66, 0.82)
Creatinine per 1 unit	1.57 (1.49, 1.65)	1.27 (1.18, 1.36)	1.87 (1.72, 2.04)	1.46 (1.41, 1.52)	NA	1.28 (1.23, 1.34)	1.67 (1.60, 1.74)	1.51 (1.45, 1.58)	0.59 (0.53, 0.67)
Dialysis vs no dialysis and creatinine = 1.0	3.20 (2.84, 3.61)	1.42 (1.17, 1.73)	NA	2.27 (2.06, 2.51)	NA	1.88 (1.52, 2.31)	2.09 (1.91, 2.30)	2.42 (2.19, 2.67)	0.35 (0.24, 0.49)
EF per 10-unit decrease	1.09 (1.04, 1.15)	NA	1.06 (1.03, 1.08)	1.12 (1.10, 1.14)	NA	1.08 (1.05, 1.10)	1.11 (1.09, 1.13)	1.10 (1.08, 1.13)	0.84 (0.81, 0.87)
Preoperative IABP/inotropes	1.43 (1.30, 1.58)	NA	1.27 (1.15, 1.39)	2.18 (2.01, 2.36)	NA	1.16 (1.06, 1.27)	1.76 (1.63, 1.90)	1.56 (1.40, 1.73)	0.52 (0.44, 0.62)
Shock	1.68 (1.45, 1.94)	1.19 (0.94, 1.50)	1.69 (1.41, 2.01)	1.93 (1.72, 2.16)	NA	1.24 (1.09, 1.41)	2.17 (1.81, 2.60)	1.45 (1.29, 1.63)	NA
Female vs male (at BSA = 1.8)	1.36 (1.26, 1.47)	1.19 (1.07, 1.32)	1.18 (1.10, 1.26)	1.25 (1.15, 1.36)	1.11 (0.88, 1.40)	0.92 (0.87, 0.97)	1.20 (1.15, 1.26)	1.31 (1.24, 1.38)	0.60 (0.55, 0.66)
Active infectious Endocarditis	2.04 (1.66, 2.50)	1.83 (1.37, 2.46)	1.52 (1.21, 1.91)	1.96 (1.69, 2.27)	NA	1.56 (1.28, 1.91)	2.11 (1.83, 2.44)	2.98 (1.86, 4.77)	0.28 (0.20, 0.38)
CLD (moderate vs mild or severe vs moderate)	1.19 (1.16, 1.23)	NA	1.12 (1.09, 1.15)	1.21 (1.16, 1.27)	1.32 (1.22, 1.42)	1.10 (1.07, 1.13)	1.18 (1.15, 1.21)	1.16 (1.10, 1.21)	0.83 (0.80, 0.85)
Reop, 1 previous operation <sup>b</sup>	2.20 (1.81, 2.67)	NA	1.29 (1.08, 1.55)	1.55 (1.32, 1.82)	NA	1.49 (1.23, 1.82)	1.50 (1.32, 1.69)	1.32 (1.10, 1.58)	0.80 (0.68, 0.95)
Reop, ≥ 2 previous operations <sup>b</sup>	2.46 (1.87, 3.24)	NA	1.47 (1.15, 1.89)	1.86 (1.53, 2.26)	NA	1.59 (1.20, 2.11)	1.77 (1.51, 2.06)	1.41 (1.11, 1.79)	0.63 (0.51, 0.79)
Status emergent, no resuscitation <sup>b</sup>	2.14 (1.62, 2.81)	2.21 (1.45, 3.37)	1.77 (1.31, 2.37)	2.71 (2.14, 3.44)	NA	1.41 (1.16, 1.70)	2.17 (1.74, 2.72)	2.72 (2.19, 3.38)	0.43 (0.29, 0.66)
Status emergent, with resuscitation or salvage <sup>b</sup>	4.56 (3.31, 6.29)	2.60 (1.53, 4.43)	1.86 (1.30, 2.65)	2.12 (1.54, 2.92)	NA	NA	3.34 (2.43, 4.61)	1.76 (1.31, 2.37)	0.23 (0.12, 0.44)

<sup>a</sup> For CVA and PLOS, MI coded ≤ 21 days; for all other endpoints, MI coded < 24 hrs or 1 to 21 days. <sup>b</sup> Variable interacts with age. Reported odds ratio represents effect of risk factor for patients aged 50 years old.

BSA = body surface area; CHF = congestive heart failure; CLD = chronic lung disease; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; Mort = mortality; NA = not applicable; NYHA = New York Heart Association; PLOS = prolonged length of stay; PVD = peripheral vascular disease; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.

## Odds Ratios

**Table 5** presents the odds ratios and 95% confidence intervals (CI) derived from these models. “Not applicable” indicates that those predictors were not included in a particular risk model.

Odds ratios that do not interact with surgery type are summarized in Part A of **Table 5**. Several variables interact with surgery type, and the odds ratios for these variables differ for some of the endpoints depending on the specific type of surgery, as summarized in **Tables 5B, C, and D** (AVR plus CABG, MVR plus CABG, MVRepair plus CABG). For example, in the model for prolonged length of stay, the odds ratio for active endocarditis is 1.81 (95% CI: 1.41 to 2.32) for AVR plus CABG; 2.08 (95% CI: 1.62 to 2.67) for MVR plus CABG; and 2.98 (95% CI: 1.86 to 4.77) for MVRepair plus CABG.

## Final Model Intercept and Coefficients

The algorithms for calculating predicted risk values, including the intercepts and regression coefficients, are presented in the Appendix.

## Limitations

The limitations of the STS valve plus CABG models are similar to those discussed in Part 1 of this series.

## Conclusion

A new STS model has been developed for valve surgery combined with CABG. This model includes specific indicator variables for each major type of valve plus CABG procedure (AVR plus CABG, MVR plus CABG, MVRepair plus CABG). Models have been developed for operative mortality, individual morbidity endpoints, a composite morbidity or mortality endpoint, and short and prolonged postoperative length of stay. Overall model performance is excellent.

## References

- Hannan EL, Kilburn H Jr, O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. *JAMA* 1990;264:2768–74.
- Edwards FH, Clark RE, Schwartz M. Coronary artery bypass grafting: the Society of Thoracic Surgeons National Database experience. *Ann Thorac Surg* 1994;57:12–9.
- O'Connor GT, Plume SK, Olmstead EM, et al. Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Northern New England Cardiovascular Disease Study Group. *Circulation* 1992;85:2110–8.
- Shahian DM, Blackstone EH, Edwards FH, et al. Cardiac surgery risk models: a position article. *Ann Thorac Surg* 2004;78:1868–77.
- Shroyer AL, Coombs LP, Peterson ED, et al. The Society of Thoracic Surgeons: 30-day operative mortality and morbidity risk models. *Ann Thorac Surg* 2003;75:1856–64.
- Edwards FH, Peterson ED, Coombs LP, et al. Prediction of operative mortality after valve replacement surgery. *J Am Coll Cardiol* 2001;37:885–92.
- Jamieson WR, Edwards FH, Schwartz M, Bero JW, Clark RE, Grover FL. Risk stratification for cardiac valve replacement. Na-

- tional Cardiac Surgery Database. Database Committee of The Society of Thoracic Surgeons. *Ann Thorac Surg* 1999;67:943–51.
- Hannan EL, Wu C, Bennett EV, et al. Risk index for predicting in-hospital mortality for cardiac valve surgery. *Ann Thorac Surg* 2007;83:921–9.
- Hannan EL, Racz MJ, Jones RH, et al. Predictors of mortality for patients undergoing cardiac valve replacements in New York State. *Ann Thorac Surg* 2000;70:1212–8.
- Nowicki ER, Birkmeyer NJ, Weintraub RW, et al. Multivariable prediction of in-hospital mortality associated with aortic and mitral valve surgery in Northern New England. *Ann Thorac Surg* 2004;77:1966–77.
- Nowicki ER. What is the future of mortality prediction models in heart valve surgery? *Ann Thorac Surg* 2005;80:396–8.
- Jin R, Grunkemeier GL, Starr A. Validation and refinement of mortality risk models for heart valve surgery. *Ann Thorac Surg* 2005;80:471–9.
- Ambler G, Omar RZ, Royston P, Kinsman R, Keogh BE, Taylor KM. Generic, simple risk stratification model for heart valve surgery. *Circulation* 2005;112:224–31.
- Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9–13.
- Gardner SC, Grunwald GK, Rumsfeld JS, et al. Comparison of short-term mortality risk factors for valve replacement versus coronary artery bypass graft surgery. *Ann Thorac Surg* 2004;77:549–56.
- Grover FL, Edwards FH. Similarity between the STS and New York State databases for valvular heart disease. *Ann Thorac Surg* 2000;70:1143–4.
- Rankin JS, Hammill BG, Ferguson TB Jr, et al. Determinants of operative mortality in valvular heart surgery. *J Thorac Cardiovasc Surg* 2006;131:547–57.
- Roques F, Nashef SA, Michel P. Risk factors for early mortality after valve surgery in Europe in the 1990s: lessons from the EuroSCORE pilot program. *J Heart Valve Dis* 2001;10:572–7.
- van Gameren M, Kappetein AP, Steyerberg EW, et al. Do we need separate risk stratification models for hospital mortality after heart valve surgery? *Ann Thorac Surg* 2008;85:921–30.
- Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691–2.
- Little RJA, Rubin DB. *Statistical analysis with missing data*. 2nd ed. Hoboken, NJ: Wiley-Interscience, 2002.
- Liang KY, Zeger SL. Longitudinal data-analysis using generalized linear-models. *Biometrika* 1986;73:13–22.
- Marin JP, Romano PS. Size matters to a model's fit. *Crit Care Med* 2007;35:2212–3.

## Appendix

### Regression Coefficients and Variable Definitions for STS 2008 Valve Plus CABG Models

For each endpoint, the formula for calculating a patient's predicted risk of the endpoint has the form:

$$\text{Predicted Risk} = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

where  $x_1, x_2, \dots, x_n$  denote patient preoperative risk factors (eg, quantitative variables such as age, and comorbidities coded as 1=present, 0=absent); and  $\beta_0, \beta_1, \dots, \beta_n$  denote regression coefficients (numerical constants). Regression coefficients for each endpoint are presented in **Appendix Table 1**. The variables  $x_1, x_2, \dots, x_n$  are the same for each endpoint and are defined in **Appendix Table 2**. The regression coefficient for the time trend is not presented. Instead, the intercept has been adjusted to incorporate the time trend. This adjusted intercept reflects the baseline risk for a reference period of July–December 2006.

Appendix Table 1. Regression Coefficients

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
Intercept	−5.24391	−5.14546	−5.32535	−3.63438	−6.50043	−3.16980	−2.99714	−4.15892	1.18582
Preoperative AFib	0.18430	0.04634	0.16567	0.12059	0.00000	0.14910	0.13766	0.19656	−0.34095
Age function 1	0.02560	0.02487	0.03268	0.02106	0.00545	0.01715	0.01838	0.03115	−0.02970
Age function 3	0.02758	−0.00709	0.00671	0.00791	−0.00985	−0.00021	0.01425	0.00985	−0.04542
Age by reop function	−0.00861	0.00458	0.00077	−0.00673	0.00314	−0.00399	−0.00202	−0.00678	0.00656
Age by status function	−0.00507	−0.01979	−0.00178	−0.00750	0.01627	−0.00029	0.00229	−0.02247	0.00692
Age by MVR function	0.01564	0.00000	0.00000	0.00000	0.00000	0.00000	0.00527	0.00000	−0.00866
Age by MVRepair function	0.01240	0.00000	0.00000	0.00000	0.00000	0.00000	0.00599	0.00000	−0.01159
BSA function 1	−1.14176	−0.81169	−0.41848	−0.66843	0.86401	−0.51266	−0.70411	−0.84204	0.51295
BSA function 2	2.25471	0.94689	1.84088	1.80467	0.42453	0.70024	1.70623	2.10402	−1.66758
CHF but not NYHA IV	0.21206	−0.01726	0.17460	0.20063	0.00000	0.00000	0.12880	0.26291	−0.17652
CHF and NYHA IV	0.39457	0.14109	0.30146	0.38383	0.00000	0.14499	0.30567	0.39791	−0.31077
CHF by MVR function	−0.31077	−0.20917	−0.25767	−0.18455	0.00000	0.00000	−0.18635	−0.23729	0.00000
CHF by MVRepair function	−0.24791	0.06897	−0.18667	−0.10484	0.00000	0.00000	−0.06920	−0.10954	0.00000
CLD function	0.17713	0.00000	0.11379	0.23345	0.27571	0.09280	0.16523	0.22999	−0.19234
CLD by MVR function	0.00000	0.00000	0.00000	−0.06780	0.00000	0.00000	0.00000	−0.04591	0.00000
CLD by MVRepair function	0.00000	0.00000	0.00000	−0.04014	0.00000	0.00000	0.00000	−0.08501	0.00000
Creatinine function 1	0.44794	0.23545	0.81612	0.38147	0.00000	0.24620	0.51256	0.41472	−0.47658
Creatinine by MVR function	0.00000	0.00000	−0.21574	0.00000	0.00000	0.00000	0.00000	0.00000	0.06652
Creatinine by MVRepair function	0.00000	0.00000	−0.18787	0.00000	0.00000	0.00000	0.00000	0.00000	−0.04407
CVD without prior CVA	0.00000	0.24847	0.13299	0.09769	0.00000	0.00000	0.10255	0.10601	−0.16643
CVD and prior CVA	0.19754	0.54344	0.11571	0.23581	0.19686	0.10974	0.23332	0.23319	−0.28560
Diabetes, noninsulin	0.11060	0.14576	0.24490	0.10365	0.26281	0.00000	0.11462	0.15846	−0.17020
Diabetes, insulin	0.26870	0.14582	0.48504	0.27893	0.68330	0.00000	0.29508	0.39583	−0.40448
Dialysis	1.61151	0.58833	0.00000	1.20290	0.61527	0.74332	1.25181	1.29747	−1.67728
Dialysis by MVR function	0.00000	0.00000	0.00000	0.00000	0.00000	−0.30339	0.00000	0.00000	0.04745
Dialysis by MVRepair function	0.00000	0.00000	0.00000	0.00000	0.00000	0.13058	0.00000	0.00000	0.09778
Ejection fraction function	0.00989	0.00000	0.00534	0.01113	0.00000	0.00703	0.01061	0.00995	−0.01440
EF by MVR function	0.01056	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00228
EF by MVRepair function	−0.00117	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.00309
Endocarditis, active	0.71327	0.60657	0.41797	0.67172	0.00000	0.44757	0.74858	0.59333	−1.27854
Endocarditis by MVR function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.13795	0.00000
Endocarditis by MVRepair function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.49934	0.00000
Female	0.30852	0.17170	0.16668	0.41874	0.10654	−0.08221	0.18594	0.26947	−0.50044
Female by MVR function	0.00000	0.00000	0.00000	−0.25972	0.00000	0.00000	0.00000	0.00000	0.08895
Female by MVRepair function	0.00000	0.00000	0.00000	−0.19373	0.00000	0.00000	0.00000	0.00000	−0.00229
Female by BSA function 1	0.51233	0.07575	0.76032	0.48032	0.80594	0.16701	0.41581	0.91055	−0.59086
Female by BSA function 2	−0.27980	−0.88628	−0.57622	−0.49740	0.58767	0.52524	−0.40427	−0.78096	0.15748
Hypertension	0.00000	0.17080	0.22638	0.09581	0.28851	0.00000	0.11445	0.07602	−0.08668

Appendix Table 1. Continued

Variable	Mort	CVA	RF	Vent	DSWI	Reop	Comp	PLOS	SLOS
IABP or inotropes	0.36025	0.00000	0.23674	0.77918	0.00000	0.15075	0.56477	0.34008	−0.58536
IABP by MVR function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.08732	−0.09462
IABP by MVRrepair function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.10281	−0.06743
Immunosuppressive treatment	0.29654	0.00000	0.26400	0.24814	0.00000	0.24041	0.23332	0.19750	−0.28819
Insufficiency, mitral	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06414	0.00000	0.00000
Insufficiency, tricuspid	0.24006	0.00000	0.22040	0.13606	0.00000	0.00000	0.13318	0.00000	−0.23141
Left main disease	0.11450	0.00000	0.00000	0.06181	0.00000	0.00000	0.00000	0.00000	0.00000
MI 1–21 days	0.17038	0.00000	0.16476	0.24560	0.00000	0.00000	0.19751	0.00000	0.00000
MI ≤ 21 days	0.00000	0.19671	0.00000	0.00000	0.00000	0.00000	0.00000	0.14599	0.00000
MI < 24 hours	0.49918	0.00000	0.26240	0.34321	0.00000	0.13716	0.39731	0.00000	0.00000
MVR	0.14888	0.32659	0.90926	0.76504	0.28437	0.41642	0.41322	0.73530	−0.82339
MVRrepair	−0.07374	0.06933	0.51275	0.28204	0.19499	0.07390	−0.03949	0.30384	−0.03552
No. diseased coronary vessel function	0.13746	0.18243	0.15791	0.17277	0.24582	0.08187	0.14767	0.12474	−0.19250
Peripheral vascular disease	0.25173	0.13776	0.14995	0.16591	0.00000	0.14312	0.18062	0.14863	0.00000
Race black	0.00000	0.00000	0.14301	0.26900	0.00000	0.17364	0.19182	0.26856	−0.43385
Race Hispanic	0.00000	0.00000	0.18384	0.15363	0.00000	0.08065	0.13561	0.12286	−0.15901
Reop, 1 previous operation	0.78624	0.00000	0.25782	0.60179	0.00000	0.33209	0.40293	0.43757	−0.39723
Reop, ≥ 2 previous operations	0.90015	0.00000	0.38499	0.78263	0.00000	0.39502	0.56875	0.50334	−0.63237
Reop by MVR function	0.00000	0.00000	0.00000	−0.27846	0.00000	−0.19608	0.00000	−0.17836	0.18262
Reop by MVRrepair function	0.00000	0.00000	0.00000	−0.16306	0.00000	0.06985	0.00000	−0.16007	0.17613
Shock	0.51917	0.17321	0.15810	0.65653	0.00000	0.21271	0.58409	0.36987	0.00000
Shock by MVR function	0.00000	0.00000	0.02883	0.00000	0.00000	0.00000	0.43045	0.00000	0.00000
Shock by MVRrepair function	0.00000	0.00000	0.36429	0.00000	0.00000	0.00000	0.19084	0.00000	0.00000
Status urgent	0.22591	0.00000	0.16451	0.22905	0.00000	0.12800	0.17511	0.24758	−0.26626
Status emergent	0.75852	0.79460	0.56854	0.99818	0.00000	0.34063	0.77631	1.00162	−1.09633
Status salvage	1.51811	0.95665	0.61798	0.75178	0.00000	0.00000	1.20732	0.56482	−1.72252
Status by MVR function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.25083
Status by MVRrepair function	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.25943
Stenosis, mitral	0.09879	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.08469	0.00000
Unstable angina	0.10722	−0.11292	0.11597	0.05762	0.00000	0.00000	0.00000	0.00000	0.00000

Afib = atrial fibrillation; BSA = body surface area; CHF = congestive heart failure; CLD = chronic lung disease; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; MVR = mitral valve replacement; MVRrepair = mitral valve repair; NYHA = New York Heart Association; PLOS = prolonged length of stay; PVD = peripheral vascular disease; Reop = reoperation; RF = renal failure; SLOS = short length of stay; Vent = prolonged ventilation.



Appendix Table 2. Definition of Variables Appearing in STS 2008 Valve Plus CABG Models

Variable	Definition
Intercept	= 1 for all patients
Preoperative AFib	= 1 if patient has history of preoperative atrial fibrillation, = 0 otherwise
Age function 1	= max (age – 50, 0)
Age function 3	= max (age – 75, 0)
Age by reop function	= Age function 1 if surgery is a reoperation, = 0 otherwise
Age by status function	= Age function 1 if status is emergent or salvage, = 0 otherwise
Age by MVR function	= Age function 1 if operation is MVR, = 0 otherwise
Age by MVRepair function	= Age function 1 if operation is MVRepair, = 0 otherwise
BSA function 1	= max (1.4, min [2.6, BSA]) – 1.8
BSA function 2	= (BSA function 1) <sup>2</sup>
CHF but not NYHA IV	= 1 if patient has CHF and is not NYHA class IV, = 0 otherwise
CHF and NYHA IV	= 1 if patient has CHF and is NYHA class IV, = 0 otherwise
CHF by MVR function	= 1 if patient has CHF and operation is MVR, = 0 otherwise
CHF by MVRepair function	= 1 if patient has CHF and operation is MVRepair, = 0 otherwise
CLD function	= 0 if no CLD, = 1 if mild CLD, = 2 if moderate CLD, = 3 if severe CLD
CLD by MVR function	= CLD function if operation is MVR, = 0 otherwise
CLD by MVRepair function	= CLD function if operation is MVRepair, = 0 otherwise
Creatinine function 1	= max (0.5, min [creatinine, 5.0]) if patient is not on dialysis, = 0 otherwise
Creatinine by MVR function	= Creatinine function 1 if valve operation is MVR, = 0 otherwise
Creatinine by MVRepair function	= Creatinine function 1 if valve operation is MVRepair, = 0 otherwise
CVD without prior CVA	= 1 if patient has history of CVD and no prior CVA, = 0 otherwise
CVD and prior CVA	= 1 if patient has history of CVD and a prior CVA, = 0 otherwise
Diabetes, noninsulin	= 1 if patient has diabetes not treated with insulin, = 0 otherwise
Diabetes, insulin	= 1 if patient has diabetes treated with insulin, = 0 otherwise
Dialysis	= 1 if patient requires dialysis preoperatively, = 0 otherwise
Dialysis by MVR function	= 1 if patient has history of dialysis and operation is MVR, = 0 otherwise
Dialysis by MVRepair function	= 1 if patient has history of dialysis and operation is MVRepair, = 0 otherwise
Ejection fraction function	= max (50 – ejection fraction, 0)
EF by MVR function	= Ejection fraction function if valve operation is MVR, = 0 otherwise
EF by MVRepair function	= Ejection fraction function if valve operation is MVRepair, = 0 otherwise
Endocarditis, active	= 1 if patient has active endocarditis, = 0 otherwise
Endocarditis by MVR function	= 1 if patient has active endocarditis and valve operation is MVR, = 0 otherwise
Endocarditis by MVRepair function	= 1 if patient has active endocarditis and valve operation is MVRepair, = 0 otherwise
Female	= 1 if patient is female, = 0 otherwise
Female by MVR function	= 1 if female and operation is MVR, = 0 otherwise
Female by MVRepair function	= 1 if female and operation is MVRepair, = 0 otherwise
Female by BSA function 1	= BSA function 1 if female, = 0 otherwise
Female by BSA function 2	= BSA function 2 if female, = 0 otherwise
Hypertension	= 1 if patient has hypertension, = 0 otherwise
IABP or inotropes	= 1 if patient requires IABP or inotropes preoperatively, = 0 otherwise
IABP by MVR function	= 1 if patient requires preop IABP or inotropes and operation is MVR, = 0 otherwise
IABP by MVRepair function	= 1 if patient requires preop IABP or inotropes and operation is MVRepair, = 0 otherwise
Immunosuppressive treatment	= 1 if patient has received immunosuppressive therapy within 30 days, = 0 otherwise
Insufficiency, mitral	= 1 if patient has at least moderate mitral insufficiency, = 0 otherwise
Insufficiency, tricuspid	= 1 if patient has at least moderate tricuspid insufficiency, = 0 otherwise
Left main disease	= 1 if patient has left main disease, = 0 otherwise
MI 1–21 days	= 1 if history of MI 1 to 21 days prior to surgery, = 0 otherwise
MI ≤ 21 days <sup>a</sup>	= 1 if patient has history of MI within 21 days prior to surgery, = 0 otherwise (for CVA and PLOS; coded as < 24 hours and 1–21 days for others)
MI < 24 hours	= 1 if history of MI < 24 hours prior to surgery, = 0 otherwise
MVR	= 1 if valve operation is mitral valve replacement, = 0 otherwise
MVRepair	= 1 if valve operation is mitral valve repair, = 0 otherwise
No. diseased coronary vessel function	= 2 if triple-vessel disease, = 1 if double-vessel disease, = 0 otherwise

Appendix Table 2. Continued

Variable	Definition
Peripheral vascular disease	= 1 if patient has peripheral vascular disease, = 0 otherwise
Race black	= 1 if patient is black, = 0 otherwise
Race Hispanic	= 1 if patient is nonblack Hispanic, = 0 otherwise
Reop, 1 previous operation	= 1 if patient has had exactly 1 previous CV surgery, = 0 otherwise
Reop, $\geq 2$ previous operations	= 1 if patient has had 2 or more previous CV surgeries, = 0 otherwise
Reop by MVR function	= 1 if surgery is a reoperation and operation is MVR, = 0 otherwise
Reop by MVRRepair function	= 1 if surgery is a reoperation and operation is MVRRepair, = 0 otherwise
Shock	= 1 if patient was in shock at time of procedure, = 0 otherwise
Shock by MVR function	= 1 if shock and operation is MVR, = 0 otherwise
Shock by MVRRepair function	= 1 if shock and operation is MVRRepair, = 0 otherwise
Status urgent	= 1 if status is urgent, = 0 otherwise
Status emergent	= 1 if status is emergent (but not resuscitation), = 0 otherwise
Status salvage	= 1 if status is salvage (or emergent plus resuscitation), = 0 otherwise
Status by MVR function	= 1 if status is emergent or salvage and operation is MVR, = 0 otherwise
Status by MVRRepair function	= 1 if status is emergent or salvage and operation is MVRRepair, = 0 otherwise
Stenosis, mitral	= 1 if patient has mitral stenosis, = 0 otherwise
Unstable angina	= 1 if patient has unstable angina and no MI within 7 days of surgery, = 0 otherwise

<sup>a</sup> MI coded  $\leq 21$  days for CVA and PLOS endpoints; for all other endpoints, coded as  $< 24$  hours and 1 to 21 days.

Note: See [www.sts.org](http://www.sts.org) for exact definitions of terms used above.

BSA = body surface area; CABG = coronary artery bypass graft surgery; CHF = congestive heart failure; CLD = chronic lung disease; Comp = composite adverse event (any); CVA = cerebrovascular accident (stroke); CVD = cerebrovascular disease; DSWI = deep sternal wound infection; EF = ejection fraction; IABP = intra-aortic balloon pump; MI = myocardial infarction; Mort = mortality; MVR = mitral valve replacement; MVRRepair = mitral valve repair; NYHA = New York Heart Association; PLOS = prolonged length of stay; PVD = peripheral vascular disease; Reop = reoperation; RF = renal failure; SLOS = short length of stay; STS = The Society of Thoracic Surgeons; Vent = prolonged ventilation.