METHODOLOGICAL ISSUES IN THE DEVELOPMENT AND APPLICATION OF PREDICTIVE RULES TO EVALUATE OUTCOMES OF CORONARY ARTERY BYPASS GRAFT SURGERY

by

Joan Ivanov, RN, MSc

A thesis submitted in conformity with the requirements

for the degree Doctor of Philosophy.

Graduate Department of the Institute of Medical Sciences

University of Toronto

Methodological Issues in the Development and Application of Predictive Rules to Evaluate Outcomes of Coronary Artery Bypass Graft Surgery.

Doctor of Philosophy, 2001

Joan Ivanov

Institute of Medical Sciences, University of Toronto

ABSTRACT

Drawing on data from a well established clinical registry for tracking preoperative characteristics and postoperative outcomes of coronary artery bypass surgery (CABG), this thesis explores several methodological and practical issues arising from the derivation, validation and application of clinical prediction rules.

Among the uses of prediction rules for operative mortality following cardiac surgery is risk-adjusted outcomes profiling. We examined whether clinicians and managers should use an existing index without modification, recalibrate that index for their populations, or derive a new model with additional risk factors. This study illustrates that poorly calibrated risk algorithms can bias the calculation of risk-adjusted outcomes.

We next demonstrated how a clinical prediction rule could be used to assist in the assessment of temporal trends in patient severity and operative mortality. In addition to providing a long-term perspective on outcome trends, this study also showed how a clinical prediction rule can be combined with contingency table analysis to document temporal shifts in risk profiles and outcomes.

A further application of clinical prediction rules is decision-support for pre-

operative patient counselling. The findings of this study highlight both the rationale for using predictive rules as a guide to decision-making, as well as the continuing challenges in persuading clinicians to use such rules.

Two recurrent issues in the use of clinical prediction rules are the short time horizon of many studies and the costs of collecting data for longer-term follow-up studies. This study illustrates the feasibility of linking clinical and administrative data in the development of more sophisticated clinical prediction rules. The findings highlight the need to evaluate long-term, as well as short-term outcomes to evaluate the benefit of surgical revascularization and the utility of adding some measurement of outcomes other than mortality.

In conclusion, drawing on data for CABG surgery, this thesis addresses several methodological and practical issues surrounding the development and application of predictive rules. We believe that these findings are generalizable to a wide range of medical diagnoses and surgical procedures where prediction rules can be used to enhance our insights into prognosis, as well as the effectiveness and efficiency of clinical interventions.

ACKNOWLEDGMENTS

I am unable to fully convey how profoundly grateful I am to Dr. C. David Naylor, my mentor and supervisor, for helping me achieve something I had never dreamed possible. We should be able to access another language at times like these when words like "grateful" or "indebted" seem so woefully inadequate. Dr. Naylor's brilliant insights, clearly focused guidance, unbelievable energy, and gentle encouragement made this a truly remarkable, productive and completely fulfilling experience. I will always be enormously proud to be counted among his many fortunate graduate students, and I hope that in the years to come, I will justify his commitment to me.

I am also indebted to my Program Advisory Committee, Dr. Jack V. Tu and Dr. Stephen E. Fremes, for their generosity and valuable insights.

I would like to extend my deepest gratitude to my many colleagues and friends in the Division of Cardiovascular Surgery at the Toronto General Hospital and the Institute for Clincial Evaluative Sciences who helped me throughout this programme. There are far too many to list but I would like to take this opportunity to mention a few: Mona Shaw (administrative assistant to Dr. Naylor at ICES), Kate Andrew (administrative assistant to Dr. Naylor at the Faculty of Medicine, University of Toronto), Josie Chapman-Smith and Hazel Pollard in the Graduate office at the Institute of Medical Sciences, and the database managers upon whom I relied for the most fundamental elements of this work - Susan Collins (Cardiovascular Surgery, Toronto General Hospital), Jeri Sever (Cardiovascular Surgery, Sunnybrook and Women's College Health Science Centre) and Claus Wall (ICES). I would also like to thank Kathy Sykora, Keyi Wu, and Peter Austin (ICES and the Clinical Epidemiology Unit at SWCHSC) for their valuable statistical help.

I am especially grateful to Dr. Tirone E. David, Chief of Cardiovascular Surgery at the Toronto General Hospital for his unfailing, generous support and for his many kindnesses toward me.

I lovingly thank my family and friends who patiently, and with good humour, endured my frequent distractions, and who were constant sources of encouragement and praise.

Lastly, I would like to thank the Heart and Stroke Foundation of Canada for Research Trainee funding in support of this work over the past four years.

This has been a truly enriching experience for me. And it serves to support the notion that if you are inspired by a great teacher and are fortunate enough to collaborate with a remarkable group of talented people, then even undreamed possibilities are within your grasp.

"...That which we are, we are -
Made weak by time and fate,
But strong in will,
To strive, to seek, to find,
And not to yield...."

from Ulysses by Lord Tennyson

This thesis is dedicated to the loving memory of my parents,
Isabel Fanny and Milorod (Mike) Ivanov.
"Dear hearts, wish you were here...."

TABLE OF CONTENTS

CHAPTER ONE

Introduction

CHAPTER TWO

Methodological Issues

CHAPTER THREE

Knowledge to Date

CHAPTER FOUR

Evaluation of three modelling strategies in the use of predictive risk indices for assessing mortality after coronary artery bypass graft surgery

CHAPTER FIVE

Application of a predictive rule used for risk stratification: Evaluation of fifteen-year trends in risk severity and operative mortality in elderly patients undergoing coronary bypass surgery

CHAPTER SIX

Predictive accuracy study: Comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery

CHAPTER SEVEN

The development and application of a predictive rule to evaluate the long-term outcomes of coronary artery bypass surgery

CHAPTER EIGHT

Summary

Chapter One

Introduction

1.1 Introduction

Medical decision making affects the health of the nation and strongly influences the allocation and consumption of health care resources. Therapeutic decision making requires an appreciation of the prognosis of disease and risk stratification, e.g., patients with the same disease can have different outcomes. The development and application of predictive rules have become increasingly important for use in clinical and health service research and as tools to help guide clinical decision making.[1] Examples of clinical prediction rules include the Apgar score for newborns,[2,3] the Glasgow Coma Scale[4] which evaluates level of consciousness,[4] the Injury Severity Score which assesses trauma patients,[5,6] the Ottawa Ankle Rule[7] which guides physicians' decisions to order an X-ray for injured ankles, the Acute Physiology and Chronic Health Evaluation score (APACHE) which evaluates patients' probability of outcomes in the intensive care unit,[8,9] and the Urgency Rating Scale developed for the Cardiac Care Network of Ontario to determine the suggested maximum time for patients waiting in the queue for cardiac surgery.[10-13] Clinical prediction rules estimate the probability of outcomes by combining information contained in a number of patient-specific characteristics in a way which best evaluates the patient's risk of either disease or adverse outcome. Predictive rules can provide key contextual data for guiding medical decision making, evaluating technologies, estimating clinical risk and benefit for an individual patient, and for the design of randomized controlled trials, e.g., risk group stratification or sample size calculations. More recently, predictive rules have been used to compare risk-adjusted outcomes across providers.

The advent of large computerized clinical and administrative databases has

allowed for the evaluation of numerous risk factors and their impact on outcomes, and the construction of statistically derived risk algorithms.[14,15] Spiegelhalter stated that the goal of statistically derived predictive rules is to provide an explicit probabilistic prediction to aid in a formal decision process.[16] However, bias can arise from multiple sources during the development and application of predictive rules (e.g., "up-coding", outcome ascertainment and statistical over- or under-fitting). A more detailed discussion of methodological issues is presented in Chapter Two.

The development and application of predictive rules to evaluate the outcomes of cardiac surgery are relatively new. The most common use of predictive rules within the contextual framework of cardiac surgery has been to provide a method of risk adjustment for the comparison of hospital outcomes and quality of care across providers -- surgeons or institutions. The use of predictive rules in cardiac surgery was born out a need to evaluate how structures and process effect outcomes as well as from institutional and surgical concerns regarding the publication of crude, unadjusted operative mortality rates in the United States. Since the late 1980's, several groups in the United States, Canada and elsewhere have published extensively on this subject.[17-71] A brief review will be presented in Chapter Three of the major contributors to the development of predictive rules used to evaluate outcomes of cardiac surgery.

Drawing on data from a well-established clinical registry of patients undergoing coronary artery bypass surgery, this thesis explores several methodological and practical issues arising from the derivation, validation and application of clinical prediction rules for: (1) provider-specific comparisons of risk-adjusted operative mortality (Chapter Four: *Ivanov J, Tu JV, Naylor CD: Ready-made, recalibrated, remodelled? Issues in the use of*

risk indices for assessing mortality after coronary artery bypass graft surgery. Circulation 1999;99:2098-2104); (2) risk group stratification (Chapter Five: *Ivanov J, Weisel RD, David TE, Naylor CD: Fifteen-year trends in risk severity and operative mortality in elderly patients undergoing coronary artery bypass graft surgery. Circulation 1998;97:673-680*); and (3) patient counselling (Chapter Six: *Ivanov J, Borger MA, David TE, Cohen G, Walton N: Predictive Accuracy Study: Comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery. Ann Thorac Surg 2000;70:162-168*). In Chapter Seven, lessons learned in the previous three chapters will be employed to develop a new predictive rule for the evaluation of long-term outcomes following coronary artery bypass surgery. Each of the key study chapters, four to seven, is presented as an independent study, containing an abstract, introduction, methods, results and discussion sections. The contributions made in this thesis and suggestions for future research will be discussed in the final summary chapter (Chapter Eight).

1.2 Chapter Four: Evaluation of Three Modelling Strategies in the Use of Predictive Risk Indices for Assessing Mortality After Coronary Artery Bypass Graft Surgery

A perplexing issue facing administrators or clinicians who wish to use a predictive rule is whether to use an existing, external risk index or to develop an internal model. If an external model is used, should the regression coefficients be recalibrated. In Chapter Four, we evaluated three modelling strategies: using an external predictive rule, recalibrating the external model in the dataset at hand, or developing a new, internal

model. We used standard statistical methodology to evaluate the three strategies and their impact on ranking, by risk-adjusted operative mortality, for fourteen surgeons. Additionally we introduced a novel strategy to compare calibration curves between the models: whereas others have evaluated calibration curves qualitatively or only reported the coefficient of determination (R^2) for observed versus expected probabilities of operative mortality, we performed an additional evaluation and reported the slope and intercept of the linear regression to quantify the degree of over- or under-estimation of each model. We also introduced the use of analysis of covariance as a method of comparing competing models in the same dataset.[72]

1.3 Chapter Five: Application of a Predictive Rule Used for Risk Stratification - Evaluation of Fifteen-year Trends in Risk Severity and Operative Mortality in Elderly Patients Undergoing Coronary Bypass Surgery

Within the context of cardiac surgery, the use of predictive rules for risk stratification has been employed more commonly as a method of comparing observed to expected probabilities, either during the model validation step or for provider-specific comparisons.[17,18,41,54] In Chapter Five we demonstrated the application of a predictive rule to stratify patients into relative risk groups for research purposes. We developed a predictive rule for operative mortality in approximately 19,000 patients who underwent coronary artery bypass surgery over a fifteen year period. We then applied cutpoints of the model-derived risk scores to construct relative risk groups (low, medium, high). This chapter clarifies the issue of how risk groups, constructed from a predictive rule, can be used for temporal trend evaluation of risk factors and outcomes in combination with

simple contingency table analysis.[73] Additionally, we developed a new rule for prediction of operative mortality in contemporary, elderly patients who undergo isolated coronary bypass surgery.

1.4 Chapter Six: Predictive Accuracy Study: Comparison of a Statistical Model to Clinicians' Probability Estimates of Adverse Outcomes Following Coronary Bypass Surgery

In Chapter Six we wished to evaluate the application of a predictive rule in clinical practice as decision-support for pre-operative patient counselling. Surgeons, residents and a nurse clinician were randomized to receive a predictive rule for either the first or second set of 50 patient vignettes and asked to estimate the probability of each patient's risk of operative mortality and prolonged intensive care unit stay after reviewing the vignettes. Clinicians' estimates were compared to probability estimates calculated from the predictive rule; additionally comparisons were made between the clinicians and the statistical model for discrimination and precision.[74]

1.5 Chapter Seven: The Development and Application of a Predictive Rule to Evaluate the Long-term Outcomes of Coronary Artery Bypass Surgery

Issues identified in the previous three chapters which were important to the development and application of a clinical predictive rule for hospital outcomes were addressed in the development of a novel predictive rule to evaluate late outcomes following cardiac surgery. Data linkage between a large clinical database and

administrative data allowed for identification of the multivariable, independent predictors of late survival and re-admission to hospital for cardiac events. Additionally, risk ratios from the Cox analysis of survival were used to determine risk scores for each patient. Patients were stratified into relative risk groups and risk-adjusted survival was compared between surgeons.

This method of linking an institutional, clinical database to administrative data for the purpose of follow-up is unique in the cardiac surgery literature. A large Canadian study by Ghali and colleagues[75] used Canadian Institute of Health Information data to determine risk factors and outcomes in over 50,000 patients undergoing coronary artery bypass surgery. However, their study was limited by the determination of comorbidity from ICD-9 codes and the lack of cardiac-specific risk factors such as left ventricular ejection fraction and extent of coronary artery disease. Two American studies which linked clinical data to administrative data were limited by the sole evaluation of mortality in Medicare patients only,[45] or in a relatively small cohort of patients following CABG.[76] Thus we present a unique Canadian study which evaluates not only survival in a large population of CABG patients but also their freedom from re-admissions for cardiac events. Also, a predictive rule was developed to predict late survival following CABG.

Chapter Two

Methodological Issues

Several authors have described methodological standards for the development and application of predictive rules.[7,14-16,77-102] Some of the major issues surrounding model development, application and sources of potential bias will be reviewed briefly in this chapter.

2.1 Prognostic and Outcome Variables

Dawson[103] suggested that each variable contained in a prognostic scale can be treated as a "test" or "measurement" and as such, we must be aware of the potential biases and sources of imprecision that can arise from multiple sources of measurement. Definitions of outcomes and prognostic variables should be precise, reliable, reproducible and free of ascertainment bias.[14,82,90] Wasson[14] suggested that rules predicting objective biological outcomes rather than sociological or behavioural outcomes are likely to be more robust when applied externally in different patient settings. The variables used to build the model should be sensible, comprehensible and have a biological association with the outcome. Feinstein referred to "content validity" as the judgemental appraisal of the underlying components of an index.[104] Therefore, subject knowledge should guide the selection of candidate variables submitted to multivariable analysis. The degree of bias present in modelling is directly proportional to the number of variables entered into the model.[89] For example, a model which is internally valid but which includes idiosyncratic variables would be biased in favour of the original dataset from which it was derived. The early deletion of unimportant or unreliable data will result in models with less over-fitting and increased generalizability. Inaccuracies in model development can arise from prognostic variables which violate

assumptions of linearity with the outcome variable, the omission of important predictors from the model, and a high frequency of missing data or improper imputation methods.[77] Strict attention to the validity of data used to develop a predictive rule will ensure minimal bias at this crucial, fundamental step in model building.

2.2 Thesis-Specific Issues Regarding the Data

Data Source and Handling

The Division of Cardiovascular Surgery at the Toronto General Hospital has maintained a database of all patients undergoing cardiac surgery since January 1st, 1982. Information was collected *prospectively* on every patient at the time of their operation and entered into a dBASEIV database by a trained, experienced data abstractor. Consistent monitoring is undertaken to ensure accuracy, especially for variables such as the Canadian Cardiovascular Society angina class and descriptors of acuteness which are susceptible to subjectivity and variability in coding. The surgical checklist (Appendix 1), attached to the front of every chart, is filled out by the surgeon at the time of surgery, prior to any postoperative events. A database manager is responsible for collecting these checklists from the chart after the patient is discharged, verifying risk factor information by reviewing the chart and filling out the information regarding postoperative morbidity or mortality. The database manager is highly trained and skilled at detecting miscoding of variables or "up-coding". Periodic data validation studies at the Toronto General Hospital conducted by the Cardiac Care Network of Ontario have demonstrated an error rate which is consistently less than 2% for all variables except for operative mortality, where the error rate is less than 1/100th of one percent and represents keystroke errors

rather than ascertainment errors.

The surgeons at the Toronto General Hospital are familiar and comfortable with the layout and format of the data checklist. Therefore, we can argue that this check list has reasonable *face validity*.[90] Variables included on the form are a result of years of discussion and modification within the cardiac surgery group and therefore are considered *sensible* and comprehensive measurements of patient characteristics and outcomes.

Data Entry and Coding

There are no coding transformations required prior to data entry. The data entry program has been developed to provide range and logic checks for each variable entered. To minimize the uncertainty associated with blank data fields, missing data are coded as "-9". Typically, the missing data rate in the Toronto General Hospital cardiac database is less than 1%.

Variable or coding transformations are done by the statistician during the analysis: for example, arbitrary cutpoints in age could be used to define "elderly" as above or below 70 years of age. Two or more variables could be grouped to create specific cohorts of patients, e.g., patients who have had a myocardial infarction within the month prior to surgery and who were still experiencing post-infarctional angina. Variables such as "timing of surgery" could be collapsed from four ordinal variables to a dichotomous variable (elective, non-elective). The validity of these transformations would be checked by evaluating their association with the outcome variable during the initial steps of statistical analysis.

Candidate Variables

Appendix 2 includes the code book for all variables contained in the Toronto General Hospital's Division of Cardiovascular Surgery Clinical Database. For the purpose of this thesis, only those variables which are relevant to coronary artery bypass graft surgery are used. Variables specific to valvular, congenital or other non-coronary surgery were excluded.

Briefly, information was collected which measured:

(1) demographics (e.g., age, sex, height, weight, previous cardiac surgery etc.);

(2) cardiac pathology (e.g., extent and specifics of cardiac disease, namely which coronary artery vessels are significantly stenotic, other cardiac pathology, left ventricular ejection fraction);

(3) symptom status (e.g., severity and pattern of angina or shortness of breath);

(4) functional status (e.g., New York Heart Association classification)

(5) other comorbidity (e.g., diabetes, renal failure, hypertension, peripheral vascular disease, chronic obstructive lung disease etc.);

(6) operative mortality;

(7) length of stay.

Variables which measure intra-operative or postoperative events were excluded from the development of the predictive rules in this thesis since their purpose(s) was "prognosis" prior to surgery and/or quality of care comparisons. Each item measures only one element of information.

Data Limitations

We were constrained by the characteristics of this existing database from which we developed and evaluated the predictive risk instruments in this thesis. This is a typical problem faced by those who wish to construct predictive rules based on data which has already been collected. There may be some important items which have been omitted or inadequately measured. For example, in the TGH database "diabetes" was coded as either present or absent. There may have been a difference between the type of diabetes (non-insulin versus insulin dependent) and the risk of operative mortality but we were unable to assess this relationship. However, the deficiencies in some items would most likely have contributed to "noise" and biased results toward the null hypothesis.

In addition, some variables which were recorded in a "Notes" field in the database, may be clinically very important to outcome but because of their low prevalence in the population of interest, fail to meet the statistical requirements for inclusion in a model (e.g., severe vasculopathy, allergy to heparin, etc.).

2.3 Statistical Modelling

The choice of which statistical model to use to construct a predictive rule depends on certain assumptions:

1) The study subjects are random and represent independent observations.

2) The distribution of the response (outcome) variable has certain properties:

    a) Logistic regression has no distributional assumptions: the predictor variable X is linearly related to log odds of the outcome variable (log (P/1-P)), where P=predicted probability of the

outcome;

b) The Cox model assumes proportional hazards over time. In Cox models for survival (S) to time=t, log(-log(S(t))) and log $\lambda(t)$ are linearly related to X, where log λ= the underlying hazard function;

3) Fully parametric tests (i.e., Weibull or log-logistic) do have distributional assumptions for specific shapes;

4) Predictors act in an additive fashion unless interaction terms are included. Interactions must be explored, especially those associated with age and risk factors, temporal variables, quality and quantity of symptoms. Interactions should be pre-specified to reduce the number of parameters submitted to the model. A pooled test of all interactions is useful because if none are significant, it may not be necessary to test any further.[77,81]

Logistic regression models generate probabilities (P) of an outcome from the formula:

$$P = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2\ldots.+\beta_{n}X_n}}{1 + e^{\beta_0+\beta_1X_1+\beta_2X_2\ldots.+\beta_{n}X_n}}$$

where β_0 = the constant and β_i = the regression coefficient for each level of a risk factor (X) in the model which characterizes the patient.

Stepwise selection adds variables to the model until the residual chi square is not significant.[92] The statistician must pay attention to interaction, collinearity, and

influential observations.[86] The choice of which variables to submit for consideration to the model should not be based solely on bivariate analysis.[94] All explanatory variables with a bivariate P value <0.25, as well as those found commonly in other major risk indices but failing to meet the critical alpha level, can be submitted to logistic regression analyses.[89,105]

Mathematical techniques to construct predictive rules include equations where likelihood can be assessed by summing weights assigned to predictors to form a score.[83] Risk weights for each prognostic variable can be constructed from rounded regression coefficients[90] or odds ratios.[39,106] Coste and colleagues[90] verified that the loss due to rounding was minimal and that the resulting simplified scoring system was attractive to clinicians because it did not require complex computation. The usefulness of a predictive rule in a clinical setting may depend on its simplicity and ease of application.[107]

Model validity is evaluated by assessing both predictive accuracy and calibration.

2.4 Predictive Accuracy

Discrimination measures a predictor's ability to separate patients with different outcomes. If a model has poor discrimination, then no amount of adjustment or calibration can correct the model. If discrimination is good, re-calibration will not sacrifice discrimination.[77,78]

For logistic or Cox models, discrimination can be quantified by the C index for ordinal assessment which is equivalent to the area under the Receiver-Operator Characteristic curve (ROC) for binomial outcomes. The area under the ROC curve is related to rank correlation between predicted and observed outcomes. The areas under

ROC curves are widely applicable and easily understood by clinicians.[77] The ROC curve area reflects the percent of concordant pairs, i.e., the member of the pair who has the disease or outcome is the one with the higher disease/outcome probability.[16,77,90,92,93,100,108,109] The area under the ROC curve assesses resolution, not calibration, and is not affected by the model's constant.[83] The area under the ROC curve can be used to compare predictive ability of two models,[90,108,109] however ROC curves are not sensitive for detecting small differences in discrimination between two models.[77]

2.5 Calibration

Hosmer-Lemeshow Goodness-of-Fit

For probabilistic prediction to be useful for medical decision making, the model must be reliable or calibrated.[16] Calibration refers to the *extent* of bias. Calibration for categorical assessments can be evaluated by the Hosmer-Lemeshow goodness-of-fit statistic (HL).[77,85,92,96,97,105] HL is derived from calculating the Pearson χ^2 for a 2 x g table of observed and expected frequencies. This statistic has an asymptotic χ^2 (g-2) distribution.[90] The null hypothesis for the HL statistic is that the model fits the data, therefore P values less than 0.05 indicate a significantly *imprecise* model.

Calibration Curves

The precision of a predictive model can also be assessed by constructing calibration curves of observed versus predicted probabilities.[17,37,89,90,92] Subgroups are formed based on composite risk scores, relative risk groups (e.g., low, medium, high risk), cutpoints of the predicted probability of outcome or clinical definitions. The mean predicted probability of outcome for each subgroup is plotted against the mean observed

outcome. The degree of departure from the 45° line, representing perfect calibration, is an assessment of bias, or over- or under-fitting. The R^2 from linear regression analysis estimates the strength of the linear relation between predicted and observed probabilities, or the amount of variation that is explained by the model. It is a method which can directly assess the amount of over- or under-fitting.[90]

Brier Scores

The Brier score measures both accuracy and calibration at the individual patient level. The Brier score (BS) is a quadratic penalty score with values between 0 and 1; therefore, the lower the Brier score, the more accurate the judgement. The formula is as follows:

$$BS = (P - d)^2$$

where P=predicted probability of outcome, d (event) 0=event did not occur, 1=event occurred. The Brier statistic as a measure is a combination of two sorts of error - calibration and discrimination. A statistic can be derived from the Brier score to test for the calibration of the model. Under the null hypothesis of perfect calibration, this statistic follows a standard normal distribution, and P-values can be derived from this distribution. [15,16,93,110-114]

Over- under-fitting/power

Goodness of fit does not imply good predictive ability. The fit of a model can be improved by increasing the number of variables in the model, but this strategy puts the model at risk of over-fitting and therefore reduces the model's performance in subsequent studies.[95] Diamond refers to over-fitting as a problem with calibration.[83] Inaccuracies due to over-fitting occur from fitting idiosyncrasies in small datasets

resulting in too many parameters for too few outcomes.[77,115] Models which contain prognostic variables associated with very broad confidence intervals about their regression estimates are red flags for statistical over-fitting.[86]

Under-fitting can occur as a result of lower power, therefore some important variables may be omitted because there are insufficient outcomes to fully model all the important predictors.[86]

Parsimony refers to fitting the smallest number of variables in a model which best describes the outcome. However, Spiegelhalter[16] stated that if the purpose of the model is prediction, then fitting a parsimonious model should not be the specific goal of analysis. Issues of unbiasedness and identification of a small number of explanatory variables are irrelevant. Spiegelhalter believed that the aim of developing a predictive model is to allow for uncertainty about the estimated parameters, leading to less extreme predictions or biased estimates and prediction. He concluded that variable selection should be governed more by the convenience and costs of measurement than by attempts to fit a parsimonious model.

Statistical errors can be avoided by paying attention to power.[82] Hsieh and colleagues have studied the number of outcome events required per predictor variable in logistic regression and concluded that 10 events per outcome minimized bias. This convention has been supported by others.[77,87,95,115] Peduzzi and colleagues performed a simulation study for the number of outcome events required per prognostic variable and demonstrated that as the event/variable ratio decreased, bias of regression coefficients increased.[87] If the dataset is small and there are numerous variables which could be independently associated with the outcome, data reduction techniques which do not use

the outcome variable may be employed to reduce the number of parameters submitted to the model. Data reduction techniques include principle components clustering, deriving clinical summary scores,[77] or combining risk factors which determine a specific subgroup membership using propensity scores.[116]

However, even if predictive accuracy and calibration are optimal, a model developed in a training set from empirical variable selection can be so highly adapted to the derivation data that the utility of the model in external datasets is compromised.[16]

2.6 Validation

The purpose of validating a model is to provide nearly unbiased estimates of predictive accuracy which are of relatively low variance.[77] The best methods for obtaining nearly unbiased internal assessments are data-splitting, cross-validation, or bootstrapping.

Data-splitting: A model is developed in a proportion of the dataset and tested for accuracy and calibration in the remaining data. With this method, the indices of accuracy may vary greatly with different splits. It is not an appropriate method if the model is then re-formed in the entire dataset for external use.[77]

Cross-validation is repeated data-splitting. The benefit is that the training samples can be large (e.g., total dataset minus 50 patients). This method reduces variability by not relying on a single split. However, others have shown that this method is relatively inefficient due to the high variation of accuracy estimates when the process is repeated.[77]

The *jackknife* method is performed by excluding one patient and then rederiving

the rule and applying it to the one excluded patient, repeated many times. The frequency that the excluded patient is misclassified is reported.

Bootstrapping provides nearly unbiased estimates of predictive accuracy that are of relatively low variance. *Bootstrap* methods measure two error rates. One is obtained by applying the original rule to a randomly chosen (with replacement) population. The second is obtained by using the new population to calculate a second rule and then applying it directly to that population, many times. The difference between the two error rates is averaged over all populations. The size of this average is an estimate of the stability of the rule as it might be applied externally. This is a good method when data are too precious to waste using a split-sample method.[14,77]

Both jackknife and bootstrap methods use the variability of the original dataset to simulate performance of the predictive rule in an external population. These methods will not eliminate bias in patient selection or data collection. The most stringent validation method is to apply the predictive rule in an external dataset and then evaluate predictive accuracy and precision.[14,80,81,86,95] The most common reason for poor external validation is statistical over-fitting.

2.7 External Application

To re-iterate: predictive rules developed in one population of patients often perform poorly when applied to external datasets.[14,80,81,90,93,95,96] The most common problem associated with poor fitting of an external rule results from statistically over-fitting the original model.[58,83,86,90,95] Even if an external model was well validated internally, secular changes in treatment or outcomes can reduce the applicability of the

model. The site and case-mix of the population from whence the rule was developed must be carefully defined. Selection bias or difference in case-mix may distort the original rule such that it is not representative of the population for which it is being applied. Additionally, predictive models may perform less well in external datasets because of differences in surveillance strategies and differences in the definitions of predictors or outcomes.[14,90,93,96] If an external model has good discrimination but poor precision, Teres[96] suggested recalibration as a method to improve predictive accuracy.

2.8 Summary

Major methodological issues surrounding the development and application of predictive rules include: 1) prognostic variable selection, 2) statistical over- or under-fitting, 3) model discrimination and calibration, 4) model selection (e.g., external or internal), and 5) model validation. These issues will be evaluated in this applied thesis within the context of coronary artery bypass surgery.

In Chapter Four we evaluate a method to assess calibration in a single model using linear regression analysis and a novel method of comparing two or more competing models using analysis of covariance. We demonstrate the advantages of recalibrating an existing, external index. Additionally, we demonstrate the improvement in both discrimination and precision achieved by remodelling an external rule.

In Chapter Five we demonstrate how a predictive rule can be used to establish a temporal benchmark. Relative risk groups constructed from a predictive rule can be used to evaluate temporal trends in risk severity and outcomes using simple, contingency table analysis.

Chapter Six represents a unique study in application. A predictive rule was offered to clinicians as a tool to guide their estimated probabilities of operative mortality and prolonged length of ICU stay. In this study we used a method of prevalence adjustment of a deliberately skewed patient sample to compare discrimination between a statistical model's and clinicians' probability estimates of outcomes. This study highlights the difficulties that may be encountered when trying to introduce a predictive rule into clinical practice.

The major methodological lessons learned in the previous chapters were incorporated in the development of a novel predictive rule to evaluate long-term survival following coronary artery bypass surgery. In Chapter Seven we present a unique Canadian study in which we linked a clinical database to administrative databases to follow patients for long-term survival and freedom from re-admissions to hospital for cardiac events.

Chapter Three

Knowledge to Date

A brief review of the experiences in Toronto, Ontario and the United States, and the forces which gave rise to the proliferation of predictive risk indexes in the cardiac surgery literature will be presented in this chapter. The major contributors to the evolution of cardiac surgery risk algorithms will also be reviewed.

3.1 Toronto Experience

In 1981, a project was initiated at the Toronto General Hospital to create a clinical data registry for all adult patients undergoing cardiac surgery. The two other University of Toronto affiliated hospitals which had adult cardiac surgery programs (Toronto Western Hospital and St. Michael's Hospital) also expressed an interest in participating in the database. The purpose of this database was to prospectively collect patient-specific, procedure-specific and outcome information on every adult patient undergoing any cardiac surgery in the city of Toronto. The information was to be used for research into the impact of risk factors on outcomes, the evaluation of temporal trends in risk factors and outcomes, clinical research projects, and monitoring quality of care.

It took one year to establish the list and coding of the variables to be collected, develop the computerized format, organize the infrastructure for data collection, and learn the necessary computer and statistical skills required to input and access the data. The original dataset contained information on cardiac-specific risk factors but there was no information collected on comorbidity, e.g., diabetes, peripheral vascular disease, hypertension, renal failure, etc.

The initial format involved coding the information on 80 column computer cards and then passing those cards through a card reader to the IBM mainframe at the

University of Toronto. The University mainframe was accessed by telephone modem from a remote dumb terminal. The statistical program, SAS (SAS Institute, Cary, NC) was used to read the data and generate reports.

In 1987, the data platform was changed to a personal computer-based dBASEIV format. At this time additional information which measured comorbidity (e.g., diabetes, hypertension, peripheral vascular disease, renal failure etc.) was added to the variables collected. Also, at this time, St. Michael's Hospital withdrew from the citywide database. Data collected from the previous five years were down loaded from the University mainframe and appended into the dBASE files. Data were entered and managed in dBASE and then uploaded to the University mainframe for analysis. This structure remained in place until 1990 when all data were managed and analyzed on a personal computer. In 1990, the Division of Cardiovascular Surgery at Sunnybrook Health Sciences Center used the dBASEIV structure to create their own database.

Since the inception of the cardiac surgery database, each participating hospital has received a confidential annual quality assurance report which was internally distributed. Additionally, this database has provided a rich source of information for observational research.[72,74,95,117-136]

3.2 Ontario Experience

In 1991, the Cardiac Care Network of Ontario was inaugurated under the direction of a committee of providers with input and financial support from the Ministry of Health. All patients who had a cardiac catheterization in the province were registered in the database. Surgical priority was recommended by a scoring algorithm based on

weights assigned to variables measuring symptom status, response to medical treatment and coronary anatomy.[10-13] The risk weights for recommended surgical priority were determined by a consensus panel of providers. Each panelist reviewed 438 case scenarios and was asked to independently rank each patient for the maximum acceptable waiting period from the time of coronary angiography to surgery. Seven levels of recommended wait times ranging from "emergency" to "marked delay" were constructed. This urgency rating scale has been used to triage every patient undergoing coronary angiography in the Province of Ontario since April 1991 in an effort to manage patients waiting in queue for coronary bypass surgery.

Morgan and colleagues[137] examined deaths in the queue waiting for cardiac surgery in ~29,000 patients from 1991 to 1995. The death rate for those waiting for isolated coronary artery bypass surgery was 0.48%. Naylor et al[138]found that registered patients who survived six months following an acute myocardial infarction and who were waiting for CABG were at a much higher risk of death than an age-gender matched population, however, they were at similar risk as those living with coronary artery disease. Morgan et al suggested that additional reductions in the already low rate of deaths in the queue could be achieved by even shorter waiting times, better compliance with existing guidelines, and guideline revisions to upgrade patients with left ventricular dysfunction. Reports from the Cardiac Care Network via the Institute for Clinical Evaluative Sciences to the Ministry of Health have directly influenced additional funding for cardiac care in the Province of Ontario and prompted increases in the maximum allowed case load per institution for coronary bypass surgery. These reports highlight the vital importance of an inclusive registry to health service management.

The Cardiac Care Network has regularly produced confidential, risk-adjusted, coronary artery bypass graft surgery outcome profiles by centre from 1993 to 1998. In 1999, the Institute for Clinical Evaluative Sciences published *Cardiovascular Health and Services in Ontario: An ICES Atlas*. This was the first public report of institution-specific, risk-adjusted operative mortality rates following coronary artery bypass surgery in the Province of Ontario.[139]

The Cardiac Care Network of Ontario has also provided valuable data for the development of a predictive rule. Tu and colleagues[39] published a multicenter predictive risk index to evaluate a combined outcome of operative mortality and prolonged length of intensive care unit stay in 1995 based on data collected by the Cardiac Care Network. This model was derived and internally validated in a heterogeneous population of coronary artery bypass graft surgery and valvular heart surgery. Six variables were modelled: age, gender, LV grade, previous coronary bypass surgery, surgical priority, and valve surgery. Although statistically robust, the model was criticized for not completely characterizing patients. Additionally, the mixed nature of the model resulted in the dilution of the impact of some of the risk factors for the purpose of prediction of operative mortality in coronary artery bypass surgery patients only (e.g., previous surgery, poor ventricular function and female gender). The external validity of this rule is evaluated in Chapter Four along with strategies designed to improve its performance in a population of patients undergoing isolated coronary artery bypass surgery.

3.3 American Experience

Cardiac surgery became the first medical discipline to be subjected to public

scrutiny. The results of open heart surgery in Medicare patients were first made public in 1986 and 1987.[17,18] In 1986 there were 320,000 coronary artery bypass procedures performed in the USA costing $7.5 billion for 1.5% of all health care expenditures.[22] Prompted by an effort to achieve cost containment in cardiac surgery, the Health Care Financing Administration disclosed survival rates following cardiac surgery with the clear intention equating operative mortality with quality of care.[32,35,56] This was later dubbed "score card surgery" by the press.

In 1988, the *New York Times* filed a request under the Freedom of Information Act to obtain information from Medicare for the operative mortality (OM) results following coronary bypass surgery. In 1991, *Newsweek* sued the New York State Department of Health for their database of information not only regarding coronary bypass surgery but also surgeon-specific outcomes. The Supreme Court of New York agreed that the public had a right to know to enable them to make informed decisions.

The State of Pennsylvania was also sued under the Freedom of Information act to release provider-specific data on outcomes of cardiac surgery. In 1989, a law was passed in Pennsylvania requiring hospitals to report mortality. In 1990, the Pennsylvania Health Care Cost Containment Council (Harrisburg, PA) published a "*Consumers Guide to Coronary Artery Bypass Graft Surgery*". This report did not include surgeon-specific results, but the surgeons were ranked in order of their risk-adjusted outcomes. The risk algorithm used to calculate risk-adjusted outcomes was proprietary and access to the formula was denied. The risk-adjustment system was criticised for being poorly designed, in that there was no distinction made between preoperative conditions and postoperative complications.[140]

These first reports from New York and then Pennsylvania by the Health Care Financing Administration contained either raw operative mortality rates or risk-adjusted rates derived from questionable data or methods, and therefore did not fairly reflect quality of care.[17,18] The result was a flurry of public outcry from surgeons complaining about unfair comparisons, and a concerted effort by several groups to develop valid risk-adjustment algorithms which would "level the playing field".

The Pennsylvania report resulted in widespread gaming. A survey found that 90% of surgeons made no use of the report and thought the report was mis-leading. The report had little influence on referral patterns but may have introduced a barrier to care for high risk patients.[140] Carey and colleagues[140] stated that there is evidence that public releases in New York and Pennsylvania may have had a negative impact on the provision of care to the most severely ill patients. Green et al[141] found that there was an increase in prevalence in five risk factors in the New York State reporting system after the release of the first public "score card". Variations of some risk factors were greater each year than could be expected, therefore Green concluded that there was again concern that improved results reflected "gaming" of comorbid variables rather than real improvements in technical quality of care. He suggested that perhaps the fault should not lie with the report cards per se but rather the method used for reporting results. Both Cary and Green felt that disclosures must be accompanied by non-punitive, voluntary quality improvement to promote an integrated, rational and evidence-based approach to the care of patients with coronary artery disease. In response to criticism regarding out-migration of high risk patients, Hannan and colleagues[26] suggested that improvements in quality of care in New York State were not related to changes in case-mix but rather an exodus of

low-volume surgeons with high risk-adjusted operative mortality rates and better performance of new surgeons.[26]

In the state of Minnesota in 1992, major payers in the state asked for risk-adjusted outcomes so they could select "centers of excellence". In 1993, surgeons organized to collect, analyze and report reliable, statistically valid data using the Society of Thoracic Surgeons voluntary database. Confidential reports and ongoing quality improvements have resulted in a decrease in mortality in both confidential and public reporting systems.[140]

An example of confidential, voluntary quality improvement can be found in the experience of the Northern New England group. The Northern New England group, comprised of five hospitals in Maine, Vermont and New Hampshire, introduced a quality assurance program in 1987. In 1990, risk-adjusted outcomes were distributed internally, preserving the anonymity of institutions and surgeons. Teams from each hospital visited each site to observe processes of care in 1990-1991. Needed changes to processes of care were identified and as a result, operative mortality decreased 24% in 1992-1993. This collaboration was felt to be a key concept for quality improvement. The authors concluded that "cross-fertilization" (one institution visiting another) resulted in a synthesis of best practice.[140] The resulting improvements in the process and structures of care were voluntary and not punitive.[18] Similar results were found in Ontario where annual risk-adjusted outcomes were also declining. The Ontario results were confidentially distributed to each centre but were not disseminated publicly.[142] Tu and colleagues[142] suggested that alternative approaches do exist for achieving improvements in mortality rates following CABG other than public disclosure.

The positive outcome of the initial public disclosures of institution and surgeon-specific outcomes was typified by a proposal by Hammermeister and others[50,52,57] for a new paradigm which would focus on monitoring quality of care through risk-adjusted outcomes. Hammermeister proposed a participatory continuous improvement model as a synthesis of three concepts: 1) continuous quality improvement, 2) intellectually and altruistically motivated self-examination and self-improvement, and 3) a modern medical information system. Hammermeister believed that participatory continuous improvement models could provide the framework to positively influence practice patterns which would then result in improved access, quality and cost-effectiveness of care.[50]

Many predictive models have been constructed for cardiac surgery patients to calculate expected operative mortality or morbidity for the purpose of comparing risk-adjusted outcomes across providers. Methodological sophistication has evolved over the past decade. The application of results from these models has either been external and public, or internal and confidential. Both applications have been focused on improving the quality of care provided to patients undergoing cardiac surgery. However, whether observed improvements in hospital outcomes can be directly attributed to provider profiling remains controversial. We present the major American contributors to the development of predictive risk algorithms below.

3.4 The Parsonnet Model

Parsonnet and colleagues[17] developed one of the earliest predictive rules to evaluate operative mortality following cardiac surgery.

Parsonnet devised a method to stratify patients by their predicted risk of operative

mortality by developing a simple scoring system from data which was readily available from a uniform reporting system in New Jersey. Parsonnet initially used a mixed model of coronary artery bypass ± valve surgery. Sixteen risk factors were evaluated by logistic regression analysis to predict operative mortality in a test set of 3500 patients and validated in a set of 1332 patients. However, the predictive risk model presented in their Table 2 contained 18 variables: gender, morbid obesity, diabetes, ejection fraction, age, reoperation, preoperative intra-aortic balloon pump, left ventricular aneurysm, emergency surgery after failed PTCA, renal dialysis, "catastrophic" states, other "rare" circumstances, mitral valve surgery, increased pulmonary artery pressures, aortic valve surgery, high aortic valve gradient, and CABG combined with valve surgery. Parsonnet stated that risk weights for each variable were derived from the logistic regression odds ratios and relative risk groups determined by cutpoints of expected operative mortality: good 0-4%, fair 5-9%, poor 10-14%, high 15-19% and extreme ≥20%.

An additive risk score was assigned to each patient in the test set by non-study personnel. The score was also externally validated in four other state hospitals. The 95% confidence intervals for observed mortality were compared to predicted mortality for each relative risk group. Linear regression analysis was used to compare the observed to the expected probabilities of mortality for the risk scores. The coefficient of determination (R^2) was calculated for the calibration curves in both the derivation and validation sets, however, the intercepts and slopes were not reported.

The authors excluded some risk factors from the scoring system because they did not satisfy the criteria of being easily quantifiable and readily available. For example, chronic obstructive lung disease (COPD) may have been an important predictor variable

but it was not available unless a pulmonary function test was done. Other variables were too subjective or not universally available, such as, the number of aortocoronary bypass grafts performed, whether or not an internal thoracic artery was used as a conduit, or operative priority. The inclusion of idiosyncratic variables or those which measured process of care may have threatened the external validity of this *prognostic* index.

The authors stated that they used the odds ratios to form the risk scores but an examination of their Tables 1 and 2 results in some confusion. Some variables which comprise the additive risk score were not listed among the variables submitted to the logistic regression. The odds ratio for diabetes was 1.58 (exponentiated regression coefficient of 0.456), but the risk score for diabetes was 3. The regression coefficient for morbid obesity was -0.271 (odds ratio 0.76), suggesting that lower weight patients were at higher risk of operative mortality. However, the risk score given to morbid obesity was 3. The variable, acute catastrophic states, measured several domains and was defined as acute structural valve failure, cardiogenic shock or acute renal failure. The odds ratio associated with this variable was 4.3. However, the risk score (10-50) assigned to "acute catastrophic states" seemed quite arbitrary. These issues may have contributed to the poor performance of the Parsonnet model in external databases.[40,72,143] Despite this, the Parsonnet rule has been used in several studies as a metric for risk-adjustment or assessment.[144-146]

In 1996, Parsonnet and colleagues[19] re-tooled their predictive rule for isolated CABG patients by eliminating optional fields and reweighting the variables. Patterned after the Society of Thoracic Surgery model, the number of variables increased from the original 18 to 20. They reported an improved fit between predicted and observed

probabilities for the highest risk patients.

3.5 New York State Model

Hannan and colleagues have written extensively on databases and processes used to develop and apply predictive rules to the outcomes of cardiac surgery in New York State.[20-28]

In response to the Health Care Financing Administrations' 1987 report, the Subcommittee on Statistics of Cardiac Surgery Advisory Committee in New York State developed a patient-specific report card in 1988 to replace the aggregate hospital forms previously submitted to the Department of Health. These reports were filled out on every patient undergoing cardiac surgery in 30 hospitals in New York State. The new system was named the Cardiac Surgery Reporting System (CSRS). The purpose of these reports was to: 1) evaluate the appropriateness of intervention with regard to long-term benefit, 2) identify risk factors associated with hospital mortality, 3) to calculate provider-specific, risk-adjusted rates, and 4) evaluate quality of care with regards to short term risks.[20,21] In the first six months of 1989, 7,596 patients were registered in the database.

Variables included in the report were those that were considered easily obtainable; they included: age, gender, ejection fraction, previous myocardial infarction, number of open heart operations in previous admissions, diabetes requiring medication, dialysis dependence, disasters (acute structural defect, renal failure, cardiogenic shock, gunshot), unstable angina, intractable congestive heart failure, left main trunk narrowed more than 90%, and type of operation performed. Similar to the Parsonnet model, this model also contained a variable, "disasters" which contained more than one element of

information. The internal consistency of the data was checked. A PC based computer system was developed which contained internal error and logic checks.

In the initial report of their proposed predictive rule to risk-adjust operative mortality,[21] logistic regression analysis was used with backward, stepwise variable selection. Age was collapsed into three categories and there were no significant first order interactions. The authors used a split-sample method to derive and test the model. However, they did not report 95% confidence intervals or standard errors around the probability estimates.

Low and high outlier hospitals were identified by the standardized mortality ratio (observed ÷ expected probability of outcome). The authors found more variation in mortality rates than could be expected by chance alone.

This was the first study in which detailed clinical information from an entire state was used to identify risk factors and institutions with significantly high risk-adjusted hospital mortality for all cardiac surgery patients.

In 1991, Hannan and colleagues[22] evaluated the relationship between the volume of cases and risk-adjusted operative mortality in the 1989 CSRS database of 12,448 CABG patients. Previous studies using claims or administrative data had failed to show a relationship between low volume and poor outcomes. However, the databases used in the previous studies had lacked information on important cardiac risk factors. Hannan et al[22] found that low volume surgeons were a major contributor to outlier status in four out of five hospitals. High volume surgeons had an operative mortality rate of 2.7% whereas low volume surgeons had an operative mortality rate of 4.3%.

The authors compared the Cardiac Surgery Reporting System data which

contained clinical information to the State-wide Planning and Research Cooperative System's (SPARCS) administrative data and evaluated the calculation of risk-adjusted operative mortality from each system.[23] Three important risk factors, not available in the administrative dataset (ejection fraction, reoperation and left main disease), accounted for much of the difference in predictive power between the two systems. The authors concluded that SPARCS would improve if secondary diagnoses such as diabetes, urgency, left main disease etc. were included in the database. Hannan et al concluded that clinical data coding was unrestrained by ICD-9 codes and had the ability to distinguish between comorbidity and postoperative complications. The kappa statistic in this study for agreement between the clinical and administrative datasets for operative mortality was 0.97.

Hannan wished to evaluate changes in risk-adjusted operative mortality during the first four years of the CSRS.[24] Surgeons and hospitals were divided into three groups based upon their performance in 1989. All groups and providers demonstrated a decrease in risk-adjusted operative mortality between 1989 and 1992, with the greatest decrease in the highest provider group. Expected operative mortality increased in all three groups over this four year time period, from an increase of 51% in the lowest group to 88% in the highest group. In fact, in the first year after the original report, expected operative mortality increased 16% in group 1, 28% in group 2, and 38% in group 3. These results suggested either a significant increase in case-mix severity in one year or possible "gaming" or "up-coding" of risk factors.

Some of the improvements may have been a result of overall quality improvement. In the highest risk group, risk-adjusted operative mortality fell from

11.8% in 1989 to 4.8% in 1992, a decrease of 59% suggesting that not all the improvements in outcome were related to gaming. However, without any internal quality control of the data, the results remain suspect. An imprecise intercept derived from the 1989 model and applied to the 1992 data may have also accounted for the increase in expected operative mortality. The prevalence of risk factors in the 1992 dataset were not presented in this paper so case-mix comparisons could not be made with the original 1989 dataset from which the model was derived.

Hannan et al continued to evaluate the improvements in outcomes in New York State.[25] The 1989-1992 CSRS dataset of 57,187 patients was used to calculate risk-adjusted outcomes from the formula: observed/expected operative mortality times the overall average operative mortality. The model had a ROC = 0.787 and HL P value of 0.16. Adjusted odds ratios and their 95% confidence intervals were presented. Hannan identified problems with audits in one out of the 10 hospitals evaluated. The authors also identified that there may have been some potential out-migration to centers in other states.

Quality improvements were identified in the process of care for emergency patients. Referral patterns were changed for high risk patients who were directed to centers with lower risk-adjusted rates. Some surgeons were denied practice and the heads of some divisions were either changed or programs reorganized. These factors may have contributed to an overall improvement in the quality of care.[25]

New York state annually reports institution-specific and surgeon-specific outcomes of cardiac surgery.[55] In response to criticisms that quality improvements in New York state resulted from a change in case-mix, Hannan and colleagues performed an

additional evaluation of the relationship between operative mortality and surgical volume.[26] They concluded that improvements were not related to changes in case-mix but rather an exodus of low volume surgeons with high risk-adjusted operative mortality, combined with better performance of new surgeons.

In 1997, Hannan et al[28] concluded that there was no bias against operating on high risk patients in New York State. High risk patients comprised 7.3% of the total population and had an expected operative mortality greater than 7.5%. However, their data showed that there was a 73% increase in the prevalence of high risk patients in only two years. Gaming or "up-coding" in the CSRS remain suspect.

3.6 Society of Thoracic Surgeons Model

Clark and colleagues[35] described the development of the first national database for cardiothoracic surgeons in the United States. The framework of this database was designed for smaller, community cardiothoracic surgeons rather than large tertiary centers. The impetus for the creation of the Society of Thoracic Surgeons (STS) database began with the Health Care Financing Administration's (HCFA) release of raw mortality rates for CABG and the response of the public to the misinterpretation of the raw rates. There was no risk stratification in the HCFA reports. In 1986, the Standards and Ethics Committee to the Council of the STS proposed the development of a nationwide, voluntary database, stating that "risk stratification is the essence of responsible cardiac surgery". The STS solicited an outside contract with Summit Medical Systems Inc in 1990 who offered a four day educational course for data quality. The Society of Thoracic Surgeons provided a template for voluntary data gathering and continues to maintain

provider confidentiality.[140] In June, 1990 the first mailing was sent to members. By February 1991, 330 members had contributed 70,000 patients. By January 1993, there was an 86% increase in the size of the database.[35]

Edwards, Clark and Schwartz[32] developed the first statistical model to predict operative mortality on behalf of the STS in approximately 80,900 patients undergoing CABG between 1980 and 1990. Some of the data pre-dated the STS database therefore ~2000 records were excluded due to invalid data. Additionally, data which pre-dated the STS database were subjected to varying methods of scrutiny and variable definition. Because of significant temporal trends in CABG outcomes and the relatively smaller amount of data prior to 1984, the authors agreed to develop the predictive model in the 1984 to 1990 data only. Data were split into equal derivation/validation sets.

Logistic regression was used to identify the independent predictors of operative mortality. Bayes Theorem was used to calculate the conditional probabilities of individual patient outcomes. Bayes modelling was chosen because the authors felt it was extremely flexible in accommodating temporal changes in the patient population. As the population changes over time, the conditional probability matrix will change to reflect the influence of current patient characteristics. As a result of using Bayes, continuous variables had to be collapsed into dichotomous variables, e.g., age<50 years (Yes/No), age 50-70 years (Yes/No), and age>70 years (Yes/No). There was no description of the validity of the chosen cutpoints, i.e., sensitivity or specificity.

Variables included in the Bayesian predictive model were: age, female gender, morbid obesity, current smoking>100 pack years, diabetes, renal failure, hypertension, cerebrovascular disease, COPD, valvular heart disease, previous cardiac surgery, number

of disease coronary arteries, left main stenosis, ejection fraction, ventricular aneurysm, previous myocardial infarction, angina stability, PTCA emergency, cardiogenic shock, IV nitroglycerine use, and IV inotropic support. There were a total of 22 prognostic elements included which increased to 33 input variables to the Bayes model.

Patients in the validation set were ordered by their predicted probability of Operative mortality as calculated by Bayes Theorem. Five risk groups were constructed to contain approximately the same number of deaths in each group. The authors compared the number of predicted to the number of observed deaths within each risk strata but did not perform any formal statistical testing of differences. Additionally, there were no confidence intervals reported for the evaluation of temporal trends and no model diagnostics (ROC, HL) reported for model validation.

In 1995, Shroyer and colleagues[115] updated the STS predictive model in over 143,700 isolated CABG patients from 374 practices in the USA. Approximately 5000 records were eliminated because of poor data quality, resulting in a final dataset of 138,762 patients.

The authors evaluated competing strategies for model building: allowing a large number of variables to compete in the model versus parsimony with a few clinically important variables. They also employed some data reduction via data review to omit those variables shown only to affect long-term outcomes but not operative mortality. Statistical methodology was improved in this report. It included checking for assumptions of each predictor variable with the outcome; collinearity and scaling checked; and some subjective variables (e.g., NYHA) were recoded (e.g., rest symptoms versus no rest symptoms). They used clinical or statistical imputation for missing

variables and included a table of the variable and the rule used for imputation.

Logistic regression analysis was used as the sole method for the development of a predictive rule in isolated CABG patients. The final model contained 33 variables. Shroyer et al reported the adjusted odds ratios but no confidence intervals or standard errors. The C statistic was good for both the test set (0.80) and the validation set (0.79) but the HL goodness-of-fit P value was <0.001 for both sets.

The patient sample was divided into deciles based on their observed operative mortality. The model significantly over-estimated risk of operative mortality in the higher risk subsets. This is a classic example of a model that discriminates well but is poorly calibrated, resulting in a lack of fit, especially at higher risk scores.

3.7 Pennsylvania State

Griffith et al[56] compared the state government-imposed Health Care Cost Containment Council model with the STS model. The Council contracted MedisGroup of Mediqual Systems Inc of Westborough, Massachusetts for use of their "Severity of Illness System (now known as Atlas Outcomes). This system unfortunately did not include some cardiac-specific risk factors important for predicting outcomes. The 1990 Pennsylvania report on surgeon-specific outcomes following CABG was based primarily on the MedisGroup program. The University of Pittsburgh sued the Council and in 1991 moved to the STS model. Hattler[54] constructed risk groups using the STS program cutpoints of expected operative mortality. The results demonstrated a poor fit between expected and observed probabilities, especially for higher risk groups.

Griffith stated that public reporting may have contributed to an improvement in

risk-adjusted outcomes as hospitals and administrators addressed processes of care along with poor performance. All risk groups improved with no obvious case shift. He felt that the STS database was flawed because it was voluntary and "gaming" was a risk. Data continue to be suspect because there is no internal audit.

The author challenged cardiovascular surgeons at the 115th Annual Meeting of the American Surgical Association (Chicago, Ill) to learn the language and methods of these statistical models.[56]

3.8 Veterans Administration Models

The Department of Veterans Affairs (VA) Continuous Improvement in Cardiac Surgery Study was initiated in 1987 to develop risk-adjustment models.[53] The database contained 12,712 patients from 43 VA hospitals. A one page information form with 34 variables was filled out on each patient. Operative mortality was defined but not the explanatory variables. A risk group was defined by the presence or absence of a risk factor. Logistic regression models were developed for each risk subgroup.

Data reduction by a series of univariate and multivariable analyses was undertaken to define a minimum of patient-related risk factors containing the maximum predictive power for operative mortality. Variables were collected from non-invasive testing (i.e., history, physical, lab) which reflected the "chronic" state of health. Non-invasive testing was used to evaluate the "acute" state of health preoperatively and invasive testing for cardiac disease severity. The authors used the non-invasive model only to avoid losing 14% of their data. The data were reduced to 14 variables and separate logistic regression models were run to evaluate odds ratios with and without the

variable present. The purpose of the report by Griffith et al[53] was to identify important variables. There was no attempt at modelling and there was no comparison of predicted versus observed outcomes.

Hammermeister and colleagues[52] reported the results of their most recent year's data from the VA. In 1994, approximately 7000 patients underwent coronary artery bypass surgery. Patient-level risk factors and outcomes were collected prospectively and subjected to logistic regression analysis to calculate the predicted probability of operative mortality. The authors noted a 14% reduction in observed to expected mortality ratios over a 4.5 year period and concluded that a large-scale, low-cost program of continuous quality improvement using risk-adjusted outcomes is feasible. They further concluded that the program of monitoring risk-adjusted outcomes was associated with a decrease in operative mortality.

3.9 Northern New England Model

The Northern New England Group is comprised of five centers in Maine, New Hampshire and Vermont. O'Connor and colleagues' [38] early report identified hospital-specific and surgeon-specific differences in risk-adjusted operative mortality. A simple ten variable model was developed to predict operative mortality in CABG patients. Variables included age, gender, BSA, Charlson comorbidity index, redo surgery, ejection fraction, LV end diastolic pressure, number of diseased vessels, left main disease, and surgical priority. The model could be criticized for including both gender and BSA as these variables often demonstrate high collinearity. In a subsequent paper, O'Connor et al[37] demonstrated a correlation R^2 of 0.99 between observed and predicted mortality in

an independent test set, however, there was no formal testing of the slope or intercept of the relationship.

O'Connor and colleagues[147] commented that the on-going evaluation of quality improvement depends on a reliable and trusted data-gathering infrastructure combined with technical quantitative expertise. It should combine feedback of outcomes data with training in continuous quality improvement techniques. The preservation of anonymity, site visits and an ongoing collaboration has resulted in quality improvement in Northern New England resulting in a 24% decrease in risk-adjusted operative mortality.

3.10 Cleveland Clinic Models

Higgins and colleagues[41] were concerned about the lack of standardization for risk-adjustment and a need for a predictive model which will predict both mortality and morbidity. They developed a predictive severity score in a test set of 5051 patients and validated it in 4069 patients undergoing CABG. Morbidity and criteria for diagnosis were explicitly defined. The patient sample was well documented. Logistic regression analysis was used to determine the independent predictors of outcome. The risk weights, 1-6, for each prognostic variable were assigned based on univariate results, odds ratios, degree of significance in the logistic regression analysis and clinical judgements. Hosmer-Lemeshow goodness-of-fit statistics and the area under the Receiver-Operator Characteristic curve were used to evaluate model validity. The resulting model contained thirteen predictor variables: emergency priority, increased creatinine, severe LV dysfunction, redo operation, operative mitral insufficiency, age, previous vascular surgery, chronic obstructive lung disease, anemia, operative aortic stenosis, weight <65

kg, diabetes and cerebrovascular disease. Sensitivity and specificity analysis were used to determine cutpoints of the severity score which would identify patients at high risk of operative mortality (score>6) or morbidity (score>4). Discrimination was good for both outcomes; however, precision was poorer for the morbidity model.

The authors concluded that the severity system was easy to use because the items were routinely available without need for special tests, in contrast to the Parsonnet model where pulmonary function tests were required for a diagnosis of COPD. The model was most useful for predicting good outcomes in low risk patients. The model did not identify those at risk for poor outcomes and does not in general, identify those at high risk.

In a later study employing a sequential risk analysis, Higgins et al[42] demonstrated that the fit between predicted and observed outcomes could be improved by including processes of care in the model. This method has limited use for a purely prognostic risk index but allows for revision of prognosis and directed ICU care; as observed by Christakis and colleagues in a substudy of a randomized controlled trial.[148] Higgins et al also concluded that cutpoints of the severity score used to classify patients into risk subgroups was useful for group analysis but had limited use for predicting outcome in individual patients.

3.11 Comparing Predictive Models

Daly and colleagues' [69] comparison of five models showed that the New York and Parsonnet models included variables which were related to processes of care, which would improve their predictive ability but which would not be appropriate for a purely

preoperative prognostic risk index.

Orr et al[40] evaluated four external models in a sample of 866 patients; the Parsonnet, Northern New England, Cleveland Clinic and New York models. Although the models were generally accurate with ROC curves ranging from 0.7 to 0.74, only two models were reasonably precise (Northern New England and Cleveland Clinic). Orr felt that the problem was related to statistical over-fitting to the study population. The inclusion of processes of care may have contributed to the lack of precision of the Parsonnet and New York models.

The 1995 report by Jones et al on behalf of the Working Panel Group for Cooperative CABG Database Project[70]compared seven external risk models used to predict operative mortality following CABG. The models compared were from the Society of Thoracic Surgeons, New York, Veterans Affairs, Duke University, Northern New England, Minnesota, and New Jersey. The purpose of this evaluation was to determine how many variables were needed to produce valid predictive operative mortality probabilities.

Seven core variables were identified by the panel; age, gender, previous surgery, LV ejection fraction, left main disease, the number of diseased coronary arteries and surgical priority. Important Level 1 and Level 2 variables were identified by the percent of the total model chi square contributed by each predictor. Level 1 variables included height, weight, PTCA on current admission, recent MI, angina, serious ventricular dysrhythmia, congestive heart failure, mitral regurgitation, diabetes, cerebrovascular disease, peripheral vascular disease, COPD, and high creatinine.

The C index did not improve significantly when level 1 variables were added to

the seven core variables. A similar study was performed by Tu and colleagues[67] who concluded that there was relatively little benefit in model discrimination attained by adding additional variables into a model which contained the most important half dozen predictors. Jones and colleagues called for standardization of databases and predictive rules used in cardiac surgery.

Weightman et al[71] compared four external risk scores in a database of 927 CABG patients. The models compared included those by Parsonnet, Higgins, Trembly and Tu models. Weightman calculated the sensitivity and specificity of each model. The authors concluded that the Parsonnet model had moderate complexity and significant observer discretion for definition of risk factors. By contrast, Tu's model was easy to calculate from commonly available variables. They recommended the use of the Tu model. Weightman and colleagues noted that all the models lacked the ability to predict risk in individual patients; however, their database was perhaps too small to support rigorous comparisons.

3.12 Summary

As this overview of methodological issues and knowledge to date suggests, finding the right model for local use in provider profiling or patient counselling is not straightforward. Local providers recurrently face the challenge of deciding whether to use a well-validated "off-the-shelf" index that may not be ideally suited to the local setting, or to reshape such a model for local use, or to derive a new internal model exclusively for local use. This issue is explored further in Chapter Four.

Predictive models developed for use in coronary artery bypass surgery may vary

greatly in size, from a relatively small model of five prognostic variables such as the CCN model, to a much larger 33 variable model such as that developed from the STS database. The balance between parsimony and more complete patient characterization remains an unresolved issue in the literature. In Chapter Four we evaluate the advantages of modestly increasing the size of a model from five, to nine, easily obtainable prognostic variables.

Model discrimination is inarguably the most important criterion for a valid model. However, model calibration is equally important, especially if the purpose of the model is to compare risk-adjusted outcomes across providers or to counsel patients regarding the risks and benefits of a procedure. Previously, some authors have calculated the coefficient of determination (R^2) between predicted and observed outcomes. However, they have not reported either the slope or intercept of the relationship. One could argue that a linear relationship whose intercept is significantly different from zero indicates a model which either over- or under-estimates the observed values. In Chapter Four we present a novel method for evaluating model calibration as well as for comparing two or more competing models. We also demonstrate the effect of model imprecision on the ranking of surgeons by their risk-adjusted operative mortality.

Many reports of predictive rules designed to identify risk factors for operative mortality following CABG have been limited either by not providing a long-term perspective on outcome trends or by not incorporating risk-adjustment algorithms that take into account the temporal shifts in risk profiles among patients receiving CABG. In Chapter Five we demonstrate that improvements in risk-adjusted outcomes were associated with an overall temporal trend towards improved outcomes which pre-dated

the public disclosure of risk-adjusted outcomes in the United States. In Chapter Five we also illustrate a simple method for temporal trend analysis of risk severity and outcomes using relative risk groups constructed from a predictive rule followed by contingency table analysis.

Although the focus of predictive algorithms in cardiac surgery has been directed more recently towards provider-profiling, the most generic use of clinical predictive rules has been to aid medical decision making and patient counselling. No model can predict the exact individual who will experience an event, however, statistical models do have the advantage of being able to integrate a large amount of information and making reasonably accurate estimates of event rates in subgroups of patients. Some authors have advocated for the inclusion of explicit probabilities of outcomes in the informed consent. The issue of interest in Chapter Six is the update of a predictive rule into practice by clinicians. We evaluate whether or not a predictive rule would aid clinicians in their probability estimates of operative mortality and prolonged length of intensive care unit stay.

The development and application of predictive rules in cardiac surgery has been almost exclusively directed towards hospital outcomes. However, the net benefit of coronary artery bypass must be evaluated by examining long-term outcomes as well as the short-term risks of surgery. In Chapter Seven we examine long-term survival and freedom from re-admission to hospital for cardiac events in a unique study which links a large clinical database to administrative data. Specifically, we examine whether a predictive rule designed for operative mortality can be expanded to evaluate long-term survival and freedom from cardiac events requiring hospitalization. Additionally, we

develop a novel predictive rule to estimate survival probabilities over the first five years following surgery and explore a method by which this rule could be used to evaluate risk-adjusted, long-term survival.

Chapter Four

Evaluation of Three Modelling Strategies in the Use of Predictive Risk Indices for Assessing Mortality After Coronary Artery Bypass Graft Surgery *

* Adapted from:

Ivanov J, Tu JV, Naylor CD: Ready-made, recalibrated, remodelled? Issues in the use of risk indices for assessing mortality after coronary artery bypass graft surgery. *Circulation* 1999;99:2098-2104

4.1 ABSTRACT

Background: Risk indices for operative mortality following cardiac surgery are used for comparative profiling of surgeons or centres. We examined whether clinicians and managers should use an existing index without modification, recalibrate that index for their populations, or derive a new model altogether.

Methods: Drawing on 7,491 consecutive patients who underwent isolated coronary artery bypass graft surgery (CABG) at two Toronto teaching hospitals between 1993 and 1996, we compared three strategies: 1) using a "ready-made", external model that had originally been derived and validated in our jurisdiction; 2) recalibrating the "ready-made" model to better fit the population; and 3) deriving a new, internal model with additional, or different risk factors. The three strategies were compared for: statistical accuracy, i.e. area under a Receiver-Operator Characteristic curve (ROC); statistical precision, measured as Hosmer-Lemeshow (HL) goodness-of-fit; and actual impact on both risk-adjusted operative mortalities (RAOM) and performance rankings for 14 surgeons. The newly derived index was externally validated in the 1998 set of 1793 CABG patients at the Toronto General Hospital.

Results: The new model was slightly more accurate than the existing index (ROC 0.78 versus 0.76 , $P<0.05$), albeit not different than the recalibrated model (ROC = 0.77). HL was significant ($P<0.001$) for the original model indicating a poor fit between predicted and observed results whereas the recalibrated and remodelled indices demonstrated excellent calibration (HL = 0.248, 0.383 respectively). The statistical imprecision of the "ready-made" model resulted in a significant underestimation of RAOM (1.6 ±0.2%) compared to the recalibrated and remodelled indices (2.5±0.2%, P = 0.048). Relative

ranking for 7 surgeons with the lowest RAOM was identical across all three models. However, remodelling resulted in re-ranking of surgeons with higher RAOM and a loss of correlation between the models for rank of surgeon (Spearman r_s=0.857, P=0.07). External validation of the remodelled index resulted in a ROC of 0.802 and HL P value of 0.42 suggesting continued good accuracy and calibration.

Conclusions: Poorly calibrated risk algorithms biased the calculation of risk-adjusted operative mortality. Any existing index used for risk assessment in cardiac surgery should be episodically recalibrated or compared to a new model to ensure that its performance remains optimal.

4.2 INTRODUCTION

Along with the proliferation of public "report cards" on cardiac surgery in the United States, researchers have published many predictive rules or risk-adjustment algorithms for mortality, morbidity and length of hospital stay after surgery.[17,21,22,32,37-39,41,53,65,92,95,115,149,150] These risk indices aim at identifying and weighting the patient characteristics that affect the probability of specific adverse outcomes. Indices are then used retrospectively to adjust for case-mix differences among surgeons and centres when performance profiles are compiled. They are also used prospectively for patient counselling and for the identification of high-risk patient subgroups for special care or research.

Patient populations — and the risk factors associated with adverse outcomes — do change over time and also differ between centres. Thus, risk models derived and validated in one locale usually perform less well when applied in other settings or even to more contemporary patients in the same setting.[17,40,41,92,149] Clinicians and managers considering the use of a risk index accordingly have three basic options:

1. They can use an existing, external index, knowing that the identified risk factors or at least the weights assigned to them may not be ideal for their patient populations ('ready-made').
2. They can accept the risk factors in a published index, but adjust the precision of the index to their own patient population by 'recalibrating' the weights assigned to a published model.
3. They can derive an internal index from their own data ('remodel').

In this chapter, we explore the implications of these options for assessment of operative mortality using a detailed dataset with information on consecutive patients undergoing isolated coronary artery bypass graft surgery [CABG] at two Toronto teaching hospitals. Specifically, we have compared the analytical strategies of 'recalibrating' and 'remodelling' against a 'ready-made' rule developed by the authors for the Ontario Cardiac Care Network (CCN).[39]

4.3 METHODS

4.3.1 Data Sources

We examined clinical risk factors and operative mortality (OM) for all 7491 patients undergoing isolated CABG under the care of fourteen surgeons at The Toronto Hospital (n = 5343) and Sunnybrook Health Science Centre (n = 2148) between April 1, 1993 and December 31, 1996. Details of this database have been previously published.[128] Data from earlier years were deliberately excluded as they were included in the multi-centre dataset used to derive and validate the CCN index. Twenty patients were missing one or more data elements used in the analyses.

4.3.2 Analysis

General issues

Data were collected and managed in dBASEIV datasets. The SAS 6.12 for Windows[151] and BMDP/DYN LR[152] programs were used for statistical analysis. We were not seeking to recalibrate or rederive an index for external and general application. Thus, we forewent split-sample methods (i.e. separate derivation and validation steps)

with both the 'recalibrating' and 'remodelling' strategies described below.

Generating the models

'Ready-made': We used a six-variable risk index developed for the Ontario Cardiac Care Network (CCN), drawing on all patients undergoing cardiac surgery in the province between April 1991 to March 1993. Since CABG served as the reference level for type of surgery in the original model, one variable – type of surgery – was set to its null value, leaving only five variables and their associated risk scores. The original regression coefficients for each variable were used to calculate patient-specific predicted probability (P) of OM from the formula:

$$P_{OM} = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2\ldots.+\beta_{n}X_n}}{1 + e^{\beta_0+\beta_1X_1+\beta_2X_2\ldots.+\beta_{n}X_n}}$$

where β_0 = the constant and β = the regression coefficient for each level of a risk factor in the model that characterizes the patient.

'Recalibrated': The five explanatory variables from the CCN index were included in logistic regression analyses of the 1993-1996 dataset to re-estimate mortality-specific regression coefficients and related risk scores. The predicted probability of OM as well as the total risk score for each patient was calculated as for the CCN model.[39,106]

'Remodelled': The University of Toronto cardiac surgery registry covers a wide variety of potential risk factors.[128] All explanatory variables with a univariate P value <0.25, as well as those found commonly in other major risk indices but failing to meet

the critical alpha level, were submitted to logistic regression analyses using forward selection combined with backwards elimination.[89,105] The best logistic regression model was determined by two diagnostic criteria: the Hosmer-Lemeshow goodness-of-fit statistic (HL)[105] and the area under the Receiver-Operator Characteristic curve (ROC).[92,108,109] As in the CCN index, odds ratios were rounded to the nearest integer, and an additive risk index created.[17,39] Because of small numbers, risk scores greater than 18 were collapsed, (i.e. score $\geq$18 for OM).

Statistical comparisons of index performance

Statistical accuracy, or model discrimination, was assessed with the area under the ROC curve [108,109] for each model, with comparisons between models as described by Hanley and McNeil.[109]

Statistical precision, or model calibration, was evaluated by the Hosmer-Lemeshow goodness-of-fit statistic.[105] We also plotted the mean predicted probability for OM against observed OM for each total risk score,[89,92] and carried out a weighted linear regression to evaluate whether the relationship was over- or under-estimated.[17,37,90] A slope of 1 and intercept of 0 would indicate a perfect fit of predicted to observed outcomes.[89] Differences in slopes and intercepts between the three regressions were evaluated by analysis of covariance with pairwise comparisons as appropriate.

Accuracy and precision at the individual patient level were evaluated by the calculation of Brier Scores. The Brier score (BS) is a quadratic penalty score with values between 0 and 1; therefore, the lower the Brier score, the more accurate the judgement.

The formula is as follows:

$$BS = (P - d)^2$$

where P=predicted probability of outcome, d (event) 0=event did not occur, 1=event occurred. The precision of a set of probabilities is evaluated by calculating the associated Z statistic. Z statistics >|1.96| indicate a set of judgements which is significantly imprecise.[15,16,93,110-114]

Clinically-salient comparisons of index performance

Expected mortality for each surgeon for each model was calculated as the weighted mean predicted probability of OM based on the observed risk factors in his/her case load. Risk-adjusted operative mortality (RAOM) was calculated by dividing the observed mortality by the expected mortality and then multiplying that ratio by the overall mortality rate in the study population (0.0226). This result can be interpreted as the mortality rate a surgeon would have if his/her case-mix was similar to the average case-mix in the study.[67]

Within each model, the difference between the mean observed mortality minus the mean expected mortality was evaluated by paired *t* test for the null hypothesis (H_0) that the difference equalled zero.[92] The differences in RAOM across 14 surgeons and three models were evaluated by analysis of variance.

For each model, the surgeons were ranked from 1 (lowest RAOM) to 14 (highest RAOM) based on how they ranked in the original CCN model. We examined qualitatively whether the ranking of surgeons changed and also calculated Spearman rank correlation coefficients (R_s) across models.

External Validation

The regression coefficients and intercept from the newly derived index were applied to the 1998 dataset of 1793 patients undergoing isolated CABG at the Toronto General Hospital. Model validity was evaluated by the area under the ROC curve and the HL goodness-of-fit statistic as well as regression analysis of the mean weighted predicted versus observed values for each risk score. This exercise was repeated for the original CCN index.

Testing an external model

For complementary insights, we used a model developed and validated by Parsonnet and colleagues[17]in a database of over 4800 patients undergoing cardiac surgery in New Jersey to calculate predicted probability of OM for each patient. The variables included in the Parsonnet model and the regression coefficients are as follows: age (0.054), aortic valve disease (0.235), bypass only (-0.588), bypass plus other procedures (0.647), elevated cholesterol (0.083), diabetes (0.456), "catastrophic" states (1.455), family history (-0.065), female gender (0.509), hypertension (0.263), left ventricular aneurysm (-0.533), left ventricular ejection fraction (0.271), mitral valve disease (0.835), obesity (-0.271), preoperative intra-aortic balloon pump (1.473), reoperation (0.893), smoking (0.089); the constant was -7.032. Further definitions of these variables are found in the report.[17] We evaluated ROC and the HL statistic.

4.4 RESULTS

4.4.1 Risk factors and risk groups for the three models

There were 169 operative deaths (2.26%) in the 1993-1996 patients. The prevalence of risk factors and their univariate association with operative mortality (OM) are shown in Table 4.1.

The independent predictors of operative mortality for the new internal model were: LV grade, age group, previous bypass surgery, the timing of surgery, sex, triple vessel disease, left main coronary artery disease, peripheral vascular disease, recent myocardial infarction, acute coronary insufficiency, and a history of hypertension. Given its recurrence as a risk factor in other published indices, we forced preoperative renal insufficiency into the model but its inclusion unfavourably affected both model discrimination and precision, possibly because of its low prevalence in our database or collinearity with other important predictors.

Table 4.2 contains the original odds ratios for OM and risk weights for the "ready-made" CCN model as well as the odds ratios, their 95% confidence intervals and risk weights from the recalibrated and remodelled indices. Comparing the CCN and remodelled indices, all five original risk factors do recur, but there are differences in weights, most notably for repeat operation and grade IV ventricular function.

Table 4.3 shows the number of patients defined by each risk score for each model and the observed operative mortality for that model's score. The addition of four explanatory variables to the remodelled index redefined "risk" in 87% of patients who bore at least one of the four conditions and resulted in a lower prevalence of both patients

and OM at the lower risk scores.

4.4.2 Statistical parameters of index performance

As shown in Table 4.4, the longer, remodelled index showed a small increment in accuracy over the original CCN index (ROC 0.78 vs. 0.76, P <0.05), with the recalibrated index between them (ROC = 0.77). The 'ready-made' CCN model showed significantly poor fit between predicted and observed results (P <0.001) at both the group (HL) and individual patient levels (BS), while the other models had acceptable calibration.

Another method of assessing model fit is shown in Figure 4.1, depicting the mean predicted probability of operative mortality versus observed operative mortality for each cumulative risk score. The original CCN model over-estimated predicted OM (top panel). In contrast, for the recalibrated and new models, slopes were closer to 1, and intercepts were not significantly different from zero. Analysis of covariance confirmed that there was a significant difference in intercepts between models. Pairwise comparisons showed that, as with the ROC curve area, the significant difference arose only in comparing the original CCN model and the newly remodelled index (P=0.044).

4.4.3 Operative Mortality: Observed, Expected, and Adjusted

We compared observed and expected mortality for each model. The observed minus expected (O-E) mortality rates for each surgeon were compared and found to be significantly different from zero for the CCN model (-1.04 ± 0.4%, P=0.011) but not with recalibration (0.17 ± 0.3%. P=0.62) or remodelling (0.21 ± 0.3%, P=0.55). Inter-model

comparisons by analysis of variance confirmed that the CCN model had a significantly greater disparity between observed and predicted results as compared to both the recalibrated (P=0.017 versus CCN) and remodelled indices (P=0.014 versus CCN).

As shown in Figure 4.2, this higher *expected* OM in the "ready-made," CCN model resulted in an under-estimation of risk-adjusted operative mortality (P=0.047 as compared to the remodelled index). RAOM was similar to unadjusted OM for the recalibrated and remodelled indices.

4.4.4 Surgeon-specific rankings

Table 4.5 depicts the relative ranking of surgeons for each of the three models from lowest RAOM (rank=1) to highest (rank=14). Recalibration resulted only in surgeons #8 and #9 exchanging positions. Remodelling, however, resulted in surgeon #13 shifting up by three positions and surgeons #10 and #11 each moving down two ranks. Despite these latter shifts, the overall Spearman correlation coefficient (R_s) showed a significant association between ranks for the CCN and new models ($R_s = 0.982$, P=0.012) because the positions of the first seven surgeons were stable across models. However, examining the seven higher ranking surgeons revealed a diminished correlation ($r_s = 0.857$, P=0.07).

4.4.5 External Validation

The external validation exercise conducted in the 1998 dataset of 1793 patients undergoing isolated CABG at the Toronto General Hospital (OM=1.3%) resulted in a

ROC of 0.802 and HL P value of 0.42 for the remodelled index. The regression analysis results comparing predicted to observed mortality for the remodelled index as applied to the 1998 data were as follows: intercept = 0.006 (P=0.045), slope = 1.09 (P<0.001), and $R^2 = 0.93$ (P<0.001).

The original CCN index also demonstrated good discrimination (ROC = 0.794), however, it remained significantly imprecise (HL P value <0.001). Regression analysis results for the CCN index as applied to the 1998 data were as follows: intercept = 0.057 (P=0.04), slope = 0.439 (P=0.016), and $R^2 = 0.54$ (P=0.016).

4.4.6 External model performance

The Parsonnet model performed poorly in our 1993-1996 dataset of isolated bypass surgery. The area under the ROC curve was only 0.45 and the HL was <0.001.

4.5 DISCUSSION

We have compared three possible strategies for assessing risk-adjusted outcomes of cardiac surgery --- 'off-the-shelf' use of a simple, multipurpose risk index, recalibrating that published index to ensure a better fit to the available data, and deriving a new, internal model with additional risk factors. The comparisons were effected in a clinical database of all isolated coronary artery bypass surgery patients undergoing operation between April 1 1993 and December 31, 1996 at two large teaching hospitals in Toronto, Canada. The newly derived model was validated in an external dataset according to methods suggested by Harrell[77] and others,[14,86,90] and demonstrated

continued good accuracy and precision. The external model developed by Parsonnet and colleagues[17] performed very poorly in our 1993-1996 dataset. The original CCN model continued to have good discrimination in the 1998 dataset but remained significantly imprecise.

Our rationale was to offer some guidance to providers who must respond to the burgeoning literature on risk indices. We accordingly discuss the implications of our findings under three subheadings below.

4.5.1 Implications for benchmarking improvements over time

The original CCN model was derived and validated for earlier years. As such, it tended to overestimate the chances of post-operative death for high risk patients, with the result that risk-adjusted outcomes improved. Experience with the Society of Thoracic Surgeons [32,35] risk adjustment algorithm has been similar. Their original 23-variable model was derived using Bayesian methods from data on tens of thousands of subjects and scores of centres. Nonetheless, in recent years the model has predicted a rising probability of operative mortality owing to an increasing prevalence of high-risk patients, even as observed operative mortality has decreased.

If the goal of an outcomes analysis is to determine trends in mortality over time, then arguably a risk model derived and validated from an earlier period can be used, because it anchors practice historically and controls for the evolution of case-mix. One limitation is that improved reportage of risk factors ("upcoding") in contemporary groups of patients may lead to a spurious impression that risk-adjusted outcomes are improving.

For example, critics have charged that this phenomenon, rather than the impact of public report cards, explains the improved outcomes of cardiac surgery in New York State.[141]

4.5.2 Implications for contemporaneous quality management or patient counselling

Setting aside temporal benchmarking, the usual goals of an outcomes analysis are contemporary quality management or risk prediction by, respectively, comparing the risk-adjusted outcomes of surgeons or centres, or identifying patients in high-risk subgroups. For this purpose, our findings suggest that practitioners and managers should consider recalibrating an existing index or developing a new model with the data at hand.

Recall first that our "ready-made" CCN index was originally derived and validated in Ontario, using data from nine centres, including the two hospitals that contributed subsequent patients to the current study. Thus, it is perhaps not surprising that the CCN index still showed good discrimination in this study, with a ROC area of 0.76. However, the CCN index showed poor calibration associated with over-estimation of expected operative mortality — a feature of model performance that is undesirable for both prospective risk prediction and post-hoc risk adjustment. Poor calibration presumably occurred not only because of temporal shifts in case-mix, but also because the CCN index was developed in a dataset that combined ischemic and valvular heart disease, and we were applying it to an isolated CABG series. Variables such as redo surgery and poor ventricular function would have different associated risks for isolated CABG versus a mixed population of valve and bypass surgery. Parsonnet,[17] for

example, observed a deterioration in model performance when his predictive rule, developed in a dataset that combined CABG with valve surgery, was subsequently tested for CABG and valve procedures separately. Additionally, the Parsonnet model contained variables such as "catastrophic states" which measured several domains such as acute structural valve failure, cardiogenic shock or acute renal failure. The inclusion of variables which occur with significant infrequency combined with a weighting strategy which was somewhat arbitrary seriously threatened the validity of the Parsonnet model in our database. The CCN index was also derived to cover both OM and length-of-stay outcomes, and the recalibration here was outcome-specific.

Recalibration was therefore a promising strategy in this context. More generally, it allows a group of practitioners to remain efficient in data collection, restricting their efforts to careful documentation of a limited number of pre-specified variables. By reweighting these variables and fine-tuning the risk index, analysts may sometimes mitigate shifts in case-mix and outcomes that occur either over time or as the index is applied to centres other than those from whence it was derived.

Indeed, recalibration did lead to some improvements in model performance in this test case. Whereas the original CCN index demonstrated significantly poor fit with data from this new series of almost 7500 patients, the recalibrated index fit the data well and we avoided over-estimating operative risk. The recalibrated model also showed similar discrimination to a new and more complex model, and yielded similar relative ranks for most surgeons. However, for the higher RAOM surgeons, the new model did lead to some alterations in surgeon outcome rankings — an observation that underscores the

potential practical importance of even small marginal improvements in model accuracy from a statistical standpoint.

In sum, for clinicians and managers who have developed their own index in the past, or found an index that shows acceptable performance in their patient populations, episodic recalibration of that index may suffice. However, in those instances where there are profound differences in case-mix or event rates, it will be prudent to derive a new model with the data at hand.

4.5.3 How many risk factors are enough?

Recently the Society of Thoracic Surgeons published its updated risk model for 1995,[115] developed from a database of over 138,000 patients operated on at 374 hospitals throughout the United States and Canada. The model shows excellent accuracy, but now requires 33 predictor variables. Apart from increased costs of data collection, and increased risks of data 'gaming' or random errors, the large numbers of explanatory variables also increase the chances of statistical over-fitting and model instability when applied to specific centres.

In contrast, the original CCN model was designed to be parsimonious and robust for multi-centre comparisons.[39] The new model adds only four variables, bringing the total to nine for isolated CABG. These factors are similar to those reported previously by our group[128] and others,[17,21,22,32,37-39,41,53,65,95,149,150] and include those highlighted in recent guidelines from the Working Group Panel on the Co-operative CABG Database Project.[70] Despite minor differences in surgeon rankings, this new model had similar

performance characteristics to the recalibrated CCN model with only five variables. This latter result is consistent with our earlier findings on the limited marginal improvements in model performance with increasing numbers of predictor variables.[67] Accurate and complete data collection on a constrained set of important variables appears to be a prudent strategy.

Summary

Our findings illustrate that temporal and inter-centre differences in case-mix make it difficult to achieve optimal predictive performance with "ready-made" risk indexes. This observation argues against the proliferation of published risk indices in the clinical literature that either affirm well-known prognostic factors or add new variables with minimal marginal impact. We have also demonstrated that recalibration of existing indices may sometimes be sufficient to ensure adequate risk prediction even when models are parsimonious. As a precaution, however, we suggest that centres collect data fastidiously on a modest-sized set of key variables such as those suggested by the Working Group Panel, [70] continue to evaluate the model's validity, and undertake intermittent remodelling to ensure that emerging risk factors are not inadvertently overlooked.

Table 4.1 **Univariate Results for the Prevalence and Operative Mortality of Risk Factors**

Risk Factor	Prevalence N (%)	OM N (%)	χ^2 P value	Unadjusted Odds Ratio
Sex: Male	5932 (79)	113 (1.9)		
Female	1559 (21)	56 (3.6)	<0.001	1.89
Age: <65	4066 (54)	56 (1.4)		
65-74	2653 (35)	73 (2.8)		1.99
≥75	772 (10)	40 (5.2)	<0.001	3.75
LV Grade: 1	2428 (32)	22 (0.9)		
2	3177 (42)	57 (1.8)		2.00
3	1666 (22)	63 (3.8)		4.22
4	220 (3)	27 (12.3)	<0.001	13.3
Timing: Elective	4153 (55)	65 (1.6)		
Semi-Urgent	3035 (42)	83 (2.7)		1.69
Emergency	218 (3)	21 (9.6)	<0.001	6.00
CABG Redo: No	7059 (94)	139 (2.0)		
Yes	432 (5.8)	30 (6.9)	<0.001	3.45
Left Main Disease: No	6296 (84)	127 (2.0)		
Yes	1195 (16)	42 (3.5)	0.001	1.75
Triple Vessel Disease: No	2149 (29)	30 (1.4)		
Yes	5342 (71)	139 (2.6)	0.001	1.86
Peripheral Vascular Disease: No	6398 (85)	121 (1.9)		
Yes	1089 (15)	47 (4.3)	<0.001	2.26
Hypertension: No	3733 (50)	63 (1.7)		
Yes	3741 (50)	104 (2.8)	0.001	1.65
Recent MI: No	5501 (74)	101 (1.8)		
Yes	1977 (26)	67 (3.4)	<0.001	1.89
Previous Stroke/TIA: No	6909 (92)	151 (2.2)		
Yes	579 (7.7)	18 (3.1)	0.151	1.41
Acute Coronary Insufficiency: No	4884 (65)	89 (1.8)		
Yes	2606 (35)	80 (3.1)	0.001	1.72

Risk Factor	Prevalence N (%)	OM N (%)	χ^2 P value	Unadjusted Odds Ratio
Diabetes: No	5662 (76)	117 (2.1)		
Yes	1824 (24)	51 (2.8)	0.067	1.33
Renal Failure: No	7227 (96)	149 (2.1)		
Yes	260 (3.5)	19 (7.3)	<0.000	3.48

Legend for Table 4.1 OM=operative mortality, MI=myocardial infarction, TIA=transient ischemic attack.

Table 4.2 **Logistic Regression-based Odds Ratios and Risk Weights for Operative Mortality**

Variable	CCN Model[39]	Risk Wgt*	Recalibrated OR (95% CI)	Risk Wgt	Remodelled OR (95% CI)	Risk Wgt
LV Grade: 2	1.29	1	1.84 (1.1, 3.0)	2	1.81 (1.1, 3.0)	2
3	1.95	2	3.75 (2.3, 6.2)	4	3.63 (2.2, 6.0)	4
4	3.64	3	13.4 (7.3, 25)	13	11.6 (6.2, 22)	12
Age: 65-74	1.88	2	1.82 (1.3, 2.6)	2	1.67 (1.2, 2.4)	2
≥75	2.59	3	3.36 (2.2, 5.2)	3	2.94 (1.9, 4.6)	3
CABG Redo	3.22	2	3.16 (2.1, 4.8)	3	3.30 (2.1, 5.1)	3
Timing: Semi-Urg	1.45	1	1.28 (0.9, 1.8)	1	1.20 (0.8, 1.7)	1
Emergency	5.70	4	3.68 (2.1, 6.5)	4	3.45 (1.9, 6.2)	3
Female	1.68	1	1.92 (1.4, 2.7)	2	1.82 (1.3, 2.6)	2
TVD	x		x		1.58 (1.0, 2.4)	2
Hypertension	x		x		1.45 (1.0, 2.0)	1
Left Main Disease	x		x		1.40 (1.0, 2.0)	1
PVD	x		x		1.39 (1.0, 2.0)	1
Constant	0.0100		0.00395		0.00221	
HL	0.45		0.354		0.599	
ROC	0.75		0.755		0.778	

Legend for Table 4.2: Risk wgt = risk weight, OR = odds ratio, 95% CI = 95% confidence interval, HL = Hosmer-Lemeshow goodness-of-fit statistic, ROC = area under the Receiver-Operator Characteristic curve, x = variable was not submitted to the model

The referent values are (in order) LV Grade 1, age 65 years, primary CABG, elective timing, male sex, single or double vessel disease, no history of hypertension, no left main

disease, no peripheral vascular disease.

* The risk weights for the "ready-made" CCN model were derived from averaging odds ratios across three separate outcomes, operative mortality, prolonged ICU length of stay and prolonged hospital length of stay. Odds ratios are presented here just for the operative mortality model.

Table 4.3 **Operative Mortality at Each Risk Score**

Risk Score	CCN Model N (%)	Recalibrated N (%)	Remodelled N (%)
0	1/703 (0.14)	1/703 (0.14)	0/150 (0)
1	10/1312 (0.76)	1/340 (0.29)	0/161 (0)
2	12/1533 (0.78)	12/1342 (0.89)	1/465 (0.22)
3	31/1379 (2.25)	5/897 (0.56)	2/622 (0.32)
4	23/1134 (2.03)	19/1195 (1.59)	6/848 (0.71)
5	28/804 (3.48)	22/996 (2.21)	6/926 (0.65)
6	32/383 (8.36)	19/618 (3.07)	13/922 (1.41)
7	13/140 (9.29)	21/587 (3.58)	17/916 (1.86)
8	6/50 (12.0)	13/303 (4.29)	20/746 (2.68)
9	8/29 (27.6)	13/132 (9.85)	16/536 (2.99)
10	5/17 (29.4)	6/100 (6.00)	15/418 (3.59)
11	-	2/29 (6.90)	16/268 (5.97)
12	-	2/12 (16.67)	12/137 (8.76)
13	-	6/67 (10.45)	7/92 (7.61)
14	-	4/36 (11.11)	5//52 (9.62)
15	-	8/71 (8.45)	4/51 (7.84)
16	-	15/56 (28.6)	3/39 (7.69)
17	-	-	7/35 (20.0)
≥18	-	-	17/82 (20.7)

Legend for Table 4.3: The addition of four variables to the remodelled index and the re-weighting of key prognostic factors resulted in fewer patients at lower risk scores in the remodelled index and substantially lower observed OM as compared to the other two models.

Table 4.4 **Evaluation of Accuracy and Precision of Predicted versus Observed OM for Each of the Three Indices**

	CCN Model	Recalibrated	Remodelled
ROC	0.760 ± 0.022	0.770 ● 0.021	0.780 ± 0.021*
HL	<0.001	0.248	0.383
BS (x 100)	2.14 ± 0.18	2.10 ± 0.15	2.08 ± 0.15
P value	<0.001	0.47	0.44

Legend for Table 4.4: ROC = area under the receiver-operator characteristic curve ± the standard error. * Different from the CCN model, P=0.045

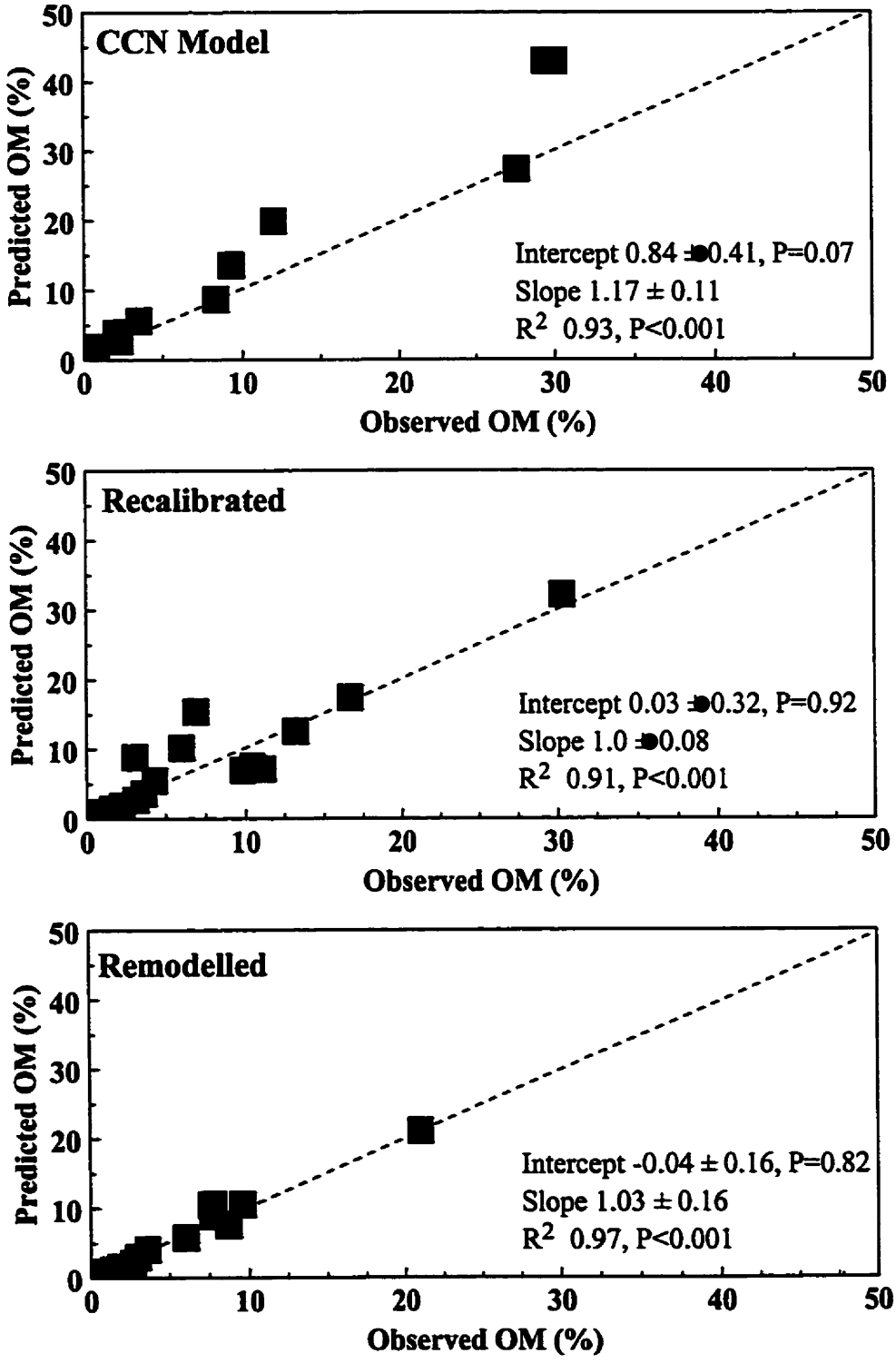
HL = Hosmer-Lemeshow goodness-of-fit P value, BS = Brier Score (± standard error); a P value <0.05 indicates poor calibration.

Table 4.5 **Risk-adjusted Operative Mortality Ranking for Each Surgeon for Each Model**

CCN Model	Recalibrated	Remodelled
Surgeon rank	Surgeon rank	Surgeon rank
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	9	9
9	8	8
10	10	13
11	11	12
12	12	10
13	13	11
14	14	14

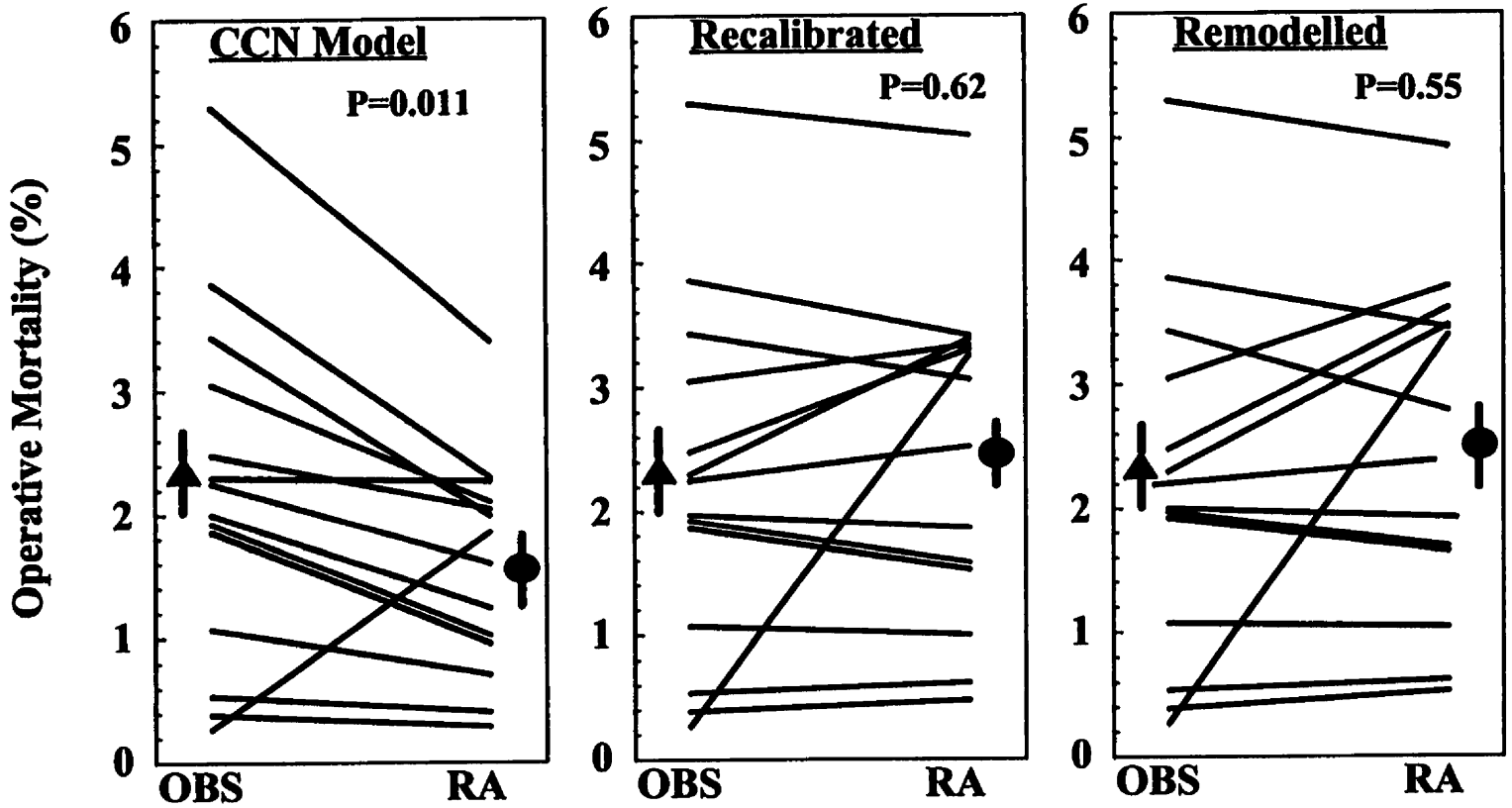
Legend for Table 4.5: The number for each surgeon was assigned according to his/her ranking position for risk-adjusted operative mortality (RAOM) rate as calculated by the original CCN model. The modelling strategy did not affect the relative ranking for the seven surgeons with the lowest RAOM (Spearman correlation coefficient = 1.0, $P<0.0001$). The type of model did however affect the ranking of surgeons with higher RAOM (Spearman correlation coefficient = 0.857, P=0.069). Surgeon #13 moved up 3 places; surgeons #10 and #11 shifted down 2 places with remodelling.

Figure 4.1



Legend 4.1 The top panel depicts the results for the mean predicted probability of operative mortality (OM) as calculated by the original CCN model's regression coefficients versus observed mortality at each cumulative risk score (closed squares). The parameter estimates for the weighted linear regression analysis are given (intercept and slope) as well as the R square. In the middle panel, regression coefficients and risk weights were re-estimated for those five variables in the CCN model. The bottom panel depicts results for the more complex, nine variable model. The dashed diagonal line represents a perfect fit between predicted and observed values. The CCN model was different from the Remodelled index by ANCOVA, P=0.044.

Figure 4.2



Legend 4.2 The difference between observed (OBS) operative mortality (OM) and risk-adjusted (RA) OM is depicted for each surgeon (solid lines). The triangles represent the mean observed operative mortality. The circles represent the risk-adjusted OM. The P values refer to the difference between the observed and risk-adjusted results. RAOM was different between the three models by ANOVA (P=0.047).

Chapter Five

Application of a Predictive Rule Used for Risk Stratification: Evaluation of Fifteen-year Trends in Risk Severity and Operative Mortality in Elderly Patients Undergoing Coronary Bypass Surgery *

* Adapted from:

Ivanov J, Weisel RD, David TE, Naylor CD: Fifteen-year trends in risk severity and operative mortality in elderly patients undergoing coronary artery bypass graft surgery. Circulation 1998;97:673-680

5.1 ABSTRACT

Background: Trends in risk-severity and operative mortality (OM) were examined in 3330 consecutive patients aged 70 years and older who underwent isolated coronary artery bypass graft surgery (CABG) between 1982 and 1996.

Methods and Results: The proportion of elderly patients rose significantly over time ($P<0.001$). Crude OM among the elderly was 7.2% in 1982-86, fell to 4.4% in 1987-1991, but did not improve thereafter. Logistic regression analysis of OM was used to construct relative risk groups (low, medium, high). The prevalence of high-risk elderly patients rose significantly over time ($P = 0.001$) from 16.2% in 1982-86 to 19.5% in 1987-91 and 26.9% in 1992-96. OM in high-risk patients fell significantly ($P = 0.044$) from 17.2% in 1982-86 to 9.1% in 1987-1991, and was 8.9% in 1992-1996.

Contemporary independent predictors of OM among elderly patients were: poor ventricular function (LV Grade 2-3, odds ratio [OR] 2.6, 95% confidence interval [CI] 1.3-5.2; LV Grade 4, OR 10.7, CI 4.4-26); previous CABG (OR 3.7, CI 2.0-7.0); female sex (OR 1.8, CI 1.1-2.8); peripheral vascular disease (OR 1.8, CI 1.1-2.8); and diabetes (OR 1.7, CI 1.1-2.7). Previous angioplasty was protective (OR 0.3, CI 0.1-0.9).

Conclusions: Previous reports of hospital outcomes in the elderly have been limited either by not providing a long-term perspective on outcome trends or by not incorporating risk-adjustment algorithms that take into account the temporal shifts in risk profiles among patients receiving CABG. This study highlights the methodological advantages of using a predictive rule to establish a temporal benchmark and stratifying patients into relative risk groups, and then using simple contingency table analysis to evaluate temporal trends in risk severity and outcomes. We observed that OM in elderly

patients has declined significantly in recent years despite an increase in the prevalence and severity of their risk factors. A careful weighing of risk, rather than advanced age alone, should determine who is offered surgical revascularization. In this regard, poor ventricular function and redo CABG surgery continue to have the greatest impact on OM in elderly patients.

5.2 INTRODUCTION

Within the context of cardiac surgery, the use of predictive rules for risk stratification has more commonly been employed as a method of comparing observed to expected probabilities, either during the model validation step or during provider-specific comparisons.[17,18,41,54] In this chapter, we demonstrate and evaluate the use of a predictive rule as a research tool used for risk stratification in elderly patients undergoing coronary artery bypass surgery (CABG).

The elderly comprise the fastest growing segment of North American society and the greatest increases in numbers are in the oldest group, i.e., persons ≥85 years.[153-156] Increasing numbers of elderly patients now undergo CABG.[155,157,158] Indeed, over the past 20 years, the definition of "elderly" in the literature on cardiac surgical outcomes has gradually increased from persons ≥65 to those ≥80.[118,136,157-172] This shift reflects reductions in the risks of surgery arising from improvements in technology, skills and patient selection.[153,156,157,160-162,165,173,174]

In Ontario, Canada's most populous province, the population-based utilization of CABG since 1985 has doubled among persons aged 65-74 years and population-based service rates more than tripled among persons aged 75 years and over.[175] Despite these trends, the 1993 CABG service rate in Ontario was about half that of New York State for patients aged 65-74 and only one-third the New York rate for those ≥75 years of age.[176] New York, in turn, is at the low end of the American range for population-based CABG rates.

Although there have been no randomized trials comparing the efficacy of surgery to medical therapy for ischemic heart disease in a targeted group of elderly patients, the

lower rates of service among elderly Ontarians do raise questions about whether some patients are being denied procedures that may improve their quality of life. Stason[177] has suggested that the elderly tend to value improved quality of life more highly than prolonged longevity. Furthermore, several studies have reported improved quality of life in elderly patients who underwent CABG.[164,167,178-180] However, if the vital risk of surgery is large, then the risk-benefit ratio may be tipped toward medical therapy for the elderly.

We have accordingly examined the trends in postoperative mortality among elderly Ontarians undergoing CABG at Canada's largest teaching hospital. We have used multivariable methods and a predictive rule to calculate risk-adjusted outcomes, thereby creating a "level playing field" for temporal comparisons of outcomes. More specifically, we report on: a) the temporal changes in the prevalence of elderly patients (≥70 years of age) undergoing CABG over a 15-year period in Toronto, b) the risk-adjusted temporal changes in clinical severity and in-hospital operative mortality among these subjects, and c) the contemporary predictors of postoperative mortality in a 1991-96 cohort of elderly patients undergoing CABG.

5.3 METHODS

5.3.1 Data Source

Clinical, operative and outcome data were collected prospectively in a computerized database for 19,009 consecutive patients undergoing isolated CABG between January 1, 1982 and December 31, 1996 at The Toronto Hospital (formerly the Toronto Western and Toronto General Hospitals prior to 1990). Patients undergoing

valve, congenital, aortic root, ventricular aneurysm repair, transplantation, ventricular mapping etc. were excluded from this study.

5.3.2 Outcome and Explanatory Variables

The outcome of interest for this study is operative mortality (OM), defined as any postoperative, in-hospital death.

Core baseline explanatory variables collected since the inception of the database in 1982 included: age, sex, LV Grade (1=EF>60%, 2=EF 40-60%, 3=EF 20-39%, 4=EF<20%), previous bypass surgery, urgency of surgery (elective, semi-urgent = surgery during the same admission as a cardiac catheterization or cardiac event, emergency = surgery within 12 hours of a cardiac catheterization or cardiac event), number of diseased coronary arteries, presence of a significant stenosis (>50% by visual evaluation of the cineangiogram) of the left main coronary artery, severity of angina, and New York Heart Association functional class.

In 1990, the database was expanded to more fully characterize our patients by adding information such as recent myocardial infarction, diabetes, peripheral vascular disease, previous angioplasty or stent, history of hypertension, renal failure (dialysis), preoperative stroke or transient ischemic attack, body size, and chronic obstructive lung disease, to name a few. Details of this database have been published elsewhere.[128]

5.3.3 Analysis

Data were collected and managed in dBASE IV datasets. The SAS for PC[151] and BMDP/DYN LR[152] programs were used for statistical analyses. Chi square or Fisher's

exact tests were used to evaluate categorical data univariately. Multivariable logistic regression methods were used to calculate risk-adjusted mortality and to calculate factor-adjusted odds ratios. Model discrimination was evaluated by the area under the Receiver-Operator Characteristic curve (ROC)[108,109] and calibration was assessed by the Hosmer-Lemeshow goodness-of-fit statistic.[105] For goodness-of-fit, the null hypothesis is that the model fits the data. Therefore, a non-significant P value is desired since a P value <0.05 would indicate a poor fit between predicted and observed results.

Evaluation of temporal trends

Rather than build a complex model to assess the temporal trends in incidence, risk-profiles, and outcomes of elderly versus non-elderly patients, we used a simpler approach based on risk stratification and contingency tables, as outlined below.

The data prior to 1990 lacked the full range of potential predictors of postoperative mortality. Thus, any temporal comparative analysis required a simple predictive rule that included only key predictors. To this end we started from the previously-validated predictive algorithm developed by Tu and colleagues[39] as a template. That algorithm was designed for both valve surgery and CABG, and therefore included type of surgery (isolated CABG, isolated valve, or valve + CABG) as a variable. Other validated models for predicting isolated CABG outcomes have included left mainstem disease as a prognostic variable; hence this item was added to the model instead of type of surgery.

To recalibrate the resulting six-variable algorithm for this one-centre temporal analysis, we performed a logistic regression analysis in the entire 1982-1996 cohort of patients. The resulting adjusted odds ratios for six explanatory variables: age (<65, 65-

74, ≥75), sex, previous CABG, urgency of surgery, LV dysfunction (LV Grade 2-3, LV Grade 4) and left main coronary artery disease, were rounded as in the algorithm of Tu et al. These rounded, adjusted odds ratios served as risk weights for each level of the predictor variables. A risk score for each patient was calculated by summing the risk weights for the variables that described the patient's baseline characteristics. Logical cutpoints of the observed mortality, determined by frequency analysis, for each risk score was used to construct relative risk groups (e.g., low, medium, high).

Next, data from 15 years were divided into three, five-year time cohorts based on date of operation: 1982-1986, 1987-1991, and 1992-1996. Patients were further divided into a younger cohort (<70 years) and an elderly cohort (≥70 years). This allowed us to use contingency table analysis to evaluate changes in the prevalence of elderly patients, their risk factors and operative mortality over time and between the three risk groups.

As a complementary method, logistic regression analysis for operative mortality was performed solely for the elderly cohort (1982-1996). This allowed us to generate elderly-specific odds ratios for the six explanatory variables, as well as the risk reduction in operative mortality associated with time after adjusting for those variables.

Contemporary Predictors of Operative Mortality

In the final step of the analysis, we focused on the cohort of elderly patients undergoing CABG between 1991 and 1996, a group that was better characterized with additional data as outlined above. This enabled us to determine the contemporary predictors of operative mortality as contrasted to the six core explanatory variables used for the temporal trend analysis. The following variables were tested by Chi-square analysis for their univariate association with operative mortality: diabetes, peripheral

vascular disease, history of hypertension, previous angioplasty/stent, renal failure, New York Heart Association class, recent preoperative myocardial infarction, preoperative stroke/transient ischemic attack, number of diseased vessels, severity of angina, and body size. We included all variables with a P value <0.20, as well as those found to be clinically important in other models[70] regardless of whether they met the critical alpha level for inclusion. These variables were submitted for consideration to a stepwise logistic regression analysis using forward selection combined with backward elimination. The best model was determined by two criteria: the area under the ROC curve, and the Hosmer-Lemeshow statistic. Because of the limited number of events, we did not undertake split-sample methods to validate the model. Such validation would in any case confirm the model's applicability in our setting, but not prove generalizability to other centres. Thus, we simply present this set of predictors for information, and as a hypothesis for consideration and validation by others.

5.4 RESULTS

5.4.1 Knowledge to Date

Table 5.1 lays out 18 papers retrieved from a Medline search that demonstrate the changing definition of "elderly" and also the range of operative mortality (5-20%) reported for isolated CABG over the past 20 years.

5.4.2 Generalizability

All 19,009 consecutive patients undergoing isolated coronary artery bypass surgery at The Toronto Hospital between January 1, 1982 and December 31, 1996 were

examined. As noted, the core variables for this study were: age, sex, LV Grade, previous CABG, urgency of surgery and left main disease. There were 346 patients with one or more of these data elements missing, making the database 98.2% complete for core information.

5.4.3 Increasing Prevalence of Elderly CABG Patients

There were 15,679 (OM=2.2%) patients under the age of 70 years who underwent CABG between 1982 and 1996. "Elderly" was defined as those patients ≥70 years of age at the time of operation (n=3,330, OM=4.95%). Figure 5.1 demonstrates the yearly increase in the prevalence of those patients aged 70-74 (top panel) and those ≥75 (bottom panel) over the 15 years of this study. Both groups increased significantly over time ($P<0.001$). The absolute numbers for patients under 70 remained fairly stable for each five-year time cohort (~5,300) but the number of elderly patients almost tripled from 593 in the 1982-86 cohort to 1,726 in the 1992-96 group.

5.4.4 Changing Risk Severity and Operative Mortality

There was a 34% overall relative risk reduction in the operative mortality rate (all patients) from 1982-86 (OM=3.52%) to the following time cohorts (OM=2.34% for both 1987-1991 and 1992-1996 cohorts).

The predictive rule, recalibrated in the 1982-1996 dataset of 19,009 patients, had a ROC area of 0.70 and a Hosmer-Lemeshow P value of 0.53. Table 5.2 shows the risk weights for each level of the prognostic variables derived from the rounded, adjusted odds ratios and the cutpoints of the total risk score used to define the relative risk groups.

Table 5.3 shows the prevalence of individual risk factors as well as the changing distribution of overall risk markers. Combined prevalence of medium- and high-risk patients increased significantly ($P<0.001$) over time for those patients over and under 70 years of age (Figure 5.2).

Overall mortality and risk-group-specific mortality data are presented in Table 5.4 and Figure 5.2. Significantly positive trends are seen for medium and high risk patients, and for most, albeit not all, risk factor subgroups in the non-elderly and elderly. Age-specific mortality improved significantly for persons <70 years of age. Even larger absolute improvements were seen among persons aged 70-74 and ≥75 years. However, these did not reach significance owing to smaller sample sizes. We accordingly turned to the overall logistic regression model for the elderly, as this would allow us to factor in the temporal increases in severity.

For the overall logistic regression, we set aside the risk scores based on rounded odds ratios and used the beta-coefficients, calculating exact adjusted odds ratios and related confidence intervals for all explanatory factors. Among 3,330 elderly patients operated on between 1982 and 1996, there were 165 deaths. Compared to 1982-1986, operations in 1986-1991 and 1992-1996 were each associated with a significant ~50% reduction in relative odds of death (Table 5.5). Adjusted odds ratios and their 95% confidence intervals for all core risk factors are also shown in Table 5.5. Predictive accuracy measured by the area under the ROC curve was 0.69 and precision, measured by the Hosmer-Lemeshow goodness-of-fit statistic was 0.232.

5.4.5 Contemporary Predictors of Operative Mortality in the Elderly

There were 2,002 elderly patients (≥70) who underwent isolated CABG between 1991 and 1996. Operative mortality was 4.6% in this group (n=92). The six core variables, age (70-74, ≥75), sex, previous CABG, LV Grade (1, 2-3, 4), timing of surgery (elective, semi-urgent, emergency) and left main disease (>50% stenosis) were submitted to a logistic regression analysis along with the following additional variables: diabetes, peripheral vascular disease, history of hypertension, previous angioplasty/stent, renal failure, New York Heart Association class, recent preoperative myocardial infarction, preoperative stroke/transient ischemic attack, number of diseased vessels, severity of angina, and body size. The contemporary, independent, multivariable predictors of operative mortality are contained in Table 5.6. Particularly interesting is a risk *reduction* associated with previous angioplasty/stent. The Hosmer-Lemeshow goodness-of-fit P value was 0.932 and area under the ROC curve was 0.713.

5.5 DISCUSSION

While advancing age remains a consistent predictor of operative mortality following isolated CABG, a variety of reports in the literature have demonstrated that elderly patients previously thought to be at very high risk for adverse events can now undergo this beneficial procedure with acceptable post-operative mortality.[153,156,157,160-163,165,173,174] Many of these reports have been limited, however, either by not providing a long-term perspective on outcome trends among the elderly, or by not incorporating risk-adjustment algorithms that take into account the temporal shifts in risk profiles among patients receiving CABG. In the latter respect statistical power

has been a problem for many published reports of CABG in the elderly with the obvious exception of the study by Hannan et al.[168] Given the current convention that there should be at least 10 outcome events for every explanatory variable in an outcomes-prediction model,[88] many studies have sample sizes that permit only two or three variables to be considered adequately for risk-adjustment purposes.

Our analysis has provided a 15-year perspective on 19,009 consecutive isolated CABG procedures at Canada's largest hospital, and includes over 3,300 patients aged 70 and over. We used a previously validated predictive rule[39] as a template for risk adjustment, added an additional explanatory variable (left main disease) and recalibrated the rule across the entire 15-year dataset to create a more level playing field for temporal comparisons of risk factor profiles and outcomes. Our findings confirm that there has not only been a time-related increase in the prevalence of older patients undergoing isolated CABG at our centre, but also an increase in the severity of the preoperative risk profile of those patients. However, risk-adjusted operative mortality has decreased significantly for elderly patients. The current overall mortality rate for elderly patients is less than 5% and only 3% for low- and medium-risk patients.

Given the enthusiasm for outcomes "score-cards", it is also noteworthy that an improvement in outcomes was already evident in the 1987-1991 period for both older and younger patients. This occurred well before Ontario embarked on its current program of systematic outcomes monitoring as described elsewhere.[181]

Because patients' characteristics were not as exhaustively documented in earlier as in later years, we cannot absolutely rule out the possibility that the recent improvements in outcomes are partly an epiphenomenon of unmeasured changes in case

selection. However, the model for mortality of elderly patients between 1982 and 1996 had a ROC curve area similar (0.69 versus 0.71) to that for elderly patients in the 1991-1996 model which drew on additional risk factor data, and the trends to inclusion of higher risk elderly patients are temporally consistent. Therefore, it is exceedingly unlikely that our findings are explained by unmeasured changes in patient characteristics working in the opposite direction, i.e. towards lower-risk case selection. Another hypothetical confounder is declining length of stay. Since patients in recent years would be discharged earlier, some who might otherwise have died in hospital would die at home and not be counted. However, this is also an implausible explanation for the observed trends. The mortality decrement was already evident for the 1987-91 period, which antedates the contemporary move to much shorter lengths of stay after CABG. Second, post-operative stays among the elderly undergoing CABG remain relatively long; for our 1992-96 cohort, the mean was 11.3 days (95% CI: 10.8, 11.8). Third, we have tracked patients after discharge in a major randomized trial.[182] Deaths occurring between discharge and 30 days from the date of surgery were uncommon; over 95% of deaths occurred on the index admission. Thus, any minor decrements in lengths of stay occurring from 1982-86 to 1987-91 are most unlikely to account for the dramatic decline in post-operative mortality observed in the same period.

The reasons for the improved outcomes nonetheless remain speculative. Possible factors include better myocardial protection during surgery (e.g. by use of blood rather than crystalloid cardioplegia or warm/tepid rather than cold cardioplegia temperatures), greater use of left internal thoracic artery conduits, and improved cardiovascular anaesthetic techniques. For that matter, most centres reporting temporal trends in overall

CABG outcomes have noted improvements over the past decade. We found no evidence for further significant changes in outcomes for 1992-1996. However, ongoing improvements may be masked by the increasing prevalence of high-risk patients who are incompletely characterized by the simplified risk-adjustment algorithm used for the 15-year trend analysis.

Analysis of the 1982-1996 dataset was obviously not ideal for delineating predictors in current practice for two reasons. First, a richer set of variables did not become available until 1990; and second, as noted above, outcomes have been improving over time. Thus, to determine the contemporary predictors of operative mortality in the elderly, a logistic regression analysis was performed on a database of 2,002 well-characterized elderly patients undergoing CABG between 1991 and 1996. Strong risk factors were presence of a grade 4 ventricle (odds ratio: 10.7, 95% CI: 4.4 - 26) or previous CABG (odds ratio: 3.73, 95% CI: 2.0-7.0). Some expected predictors -- such as age $\geq$75 years, urgent/emergency surgery, renal failure, number of diseased vessels, presence of left mainstem disease, or recent preoperative myocardial infarction -- fell out of the final model. This is partly a function of statistical power, but it is instructive that other predictors -- such as diabetes and peripheral vascular disease -- took precedence in the risk adjustment algorithm.

The protective effect of previous angioplasty/stenting in the algorithm is strikingly large. This finding is almost counter-intuitive, as one would expect individuals at higher risk from CABG to be more likely to undergo angioplasty as a temporizing measure, rather than proceeding directly to open heart surgery. There are two plausible explanations for the effect, which are not mutually exclusive. The first is that previous

angioplasty/stenting is directly protective by reducing one or more of the critical coronary arterial stenoses prior to CABG. The second is that the prior occurrence of angioplasty/stenting is actually a proxy for less extensive or severe coronary atherosclerosis, independent of the number of diseased vessels. In the latter scenario, patients with a limited number of discrete lesions would be more likely to undergo a prior angioplasty than those with diffuse atheroma and distal vessel involvement.

Limitations of this analysis are those that apply in any observational outcomes analysis. First, miscoding of key risk factors is always a concern. The database in question is very well-established, and subject to systematic logic and range checks. Definitions for risk factors have remained unchanged since 1982 in some cases, and since 1990 for those factors added later. Furthermore, the data are subject to random audits which have consistently shown raw inter-abstractor agreement on major variables to be greater than 98%. The dataset is also acceptably complete in terms of core variable information with only 1.8% of patients missing one or more key data elements.

Second, we have demonstrated the improving outcomes of isolated CABG in the presence of growing numbers of procedures on high-risk elderly patients; but, in contrast to the accumulated trial data on younger persons, the risk-benefit ratios of surgery for the elderly are not precisely defined.[173,183] The elderly are a very diverse group in terms of their physical and mental health, work capacity and economic status.[184] Prudent case selection must obviously take into account the baseline functional capacities and preferences of elderly patients. Furthermore, our results do show that high-risk elderly patients still have a significantly increased operative mortality rate (8.9%); those with poor ventricular function or previous CABG are at particularly high risk of post-operative mortality. These latter findings may help clinicians in counselling elderly

patients about the risks of isolated CABG surgery.

Third, as noted in the *Methods* section, the predictive algorithm for the most recent period is based on consecutive patients from a single centre, without independent validation. With only 92 deaths among the 2,002 elderly patients undergoing isolated CABG at our centre between 1991 and 1996, it was not feasible to split the cohort into derivation and validation samples. That said, most risk-adjustment algorithms in the literature have been validated only in the centres from which they were derived. Other centres should ideally use their own outcomes data to validate and recalibrate the risk factors identified here.

Fourth, we have focused exclusively on mortality as a post-operative complication. Additional work is needed to delineate trends in post-operative morbidity, not the least of which is stroke, an outcome particularly feared by the elderly.

Furthermore, we do not have data on long-term life expectancy gains or quality of life enhancement. Linkage to provincial statistics is planned to address life-expectancy gains, but in the absence of prospective data collection, quality of life improvements cannot be quantified.

In conclusion, previous reports of hospital outcomes in the elderly have been limited either by not providing a long-term perspective on outcome trends or by not incorporating risk-adjustment algorithms that take into account the temporal shifts in risk profiles among patients receiving CABG. This study highlights the value of using a predictive rule for research purposes. The predictive rule was used to stratify patients into relative risk groups allowing us to use simple contingency table analysis to evaluate temporal shifts in risk profiles and outcomes. The results of this simpler analytic format

were corroborated by logistic regression analysis which include temporal predictors in the final model for operative mortality in elderly patients (Table 5.5).

Coronary artery bypass surgery may sometimes be the best of the unattractive options for elderly patients who have a progression of disease and symptoms. Operative mortality after isolated coronary bypass surgery in the elderly declined significantly starting in the late 1980's for this important and growing group of patients, despite an increase in the prevalence and severity of their risk factors; and has been stable since the early 1990's at under 5% overall for patients aged ≥70. A careful weighing of risk, rather than advanced age alone, should determine who is offered surgical revascularization. In this regard, poor ventricular function and redo CABG surgery continue to have the most impact on operative mortality in elderly patients in our centre. These and other risk factors noted here can serve as a starting point for cardiologists and surgeons who wish either to counsel elderly patients about the vital risks of isolated coronary bypass surgery, or to delineate risk factors for adverse events in their own practices. In Chapter Six we will evaluate the use of a predictive rule as a clinical tool for estimating the probability of adverse outcomes following CABG.

Table 5.1 **Incidence and outcomes of CABG surgery in the elderly: Changing profile of cohort studies in the literature**

Author	Age	N	OM
Gann (1977)[136]	≥70	30	6.7
Knapp (1981)[159]	<70 ≥70	2850 121	1.1 1.6
Gersh (1983)[160]	65-69 70-74 ≥75	803 241 42	4.6 6.6 9.5
Hibler (1983)[161]	65-69 70-74 ≥75	65 40 10	4.6 5.0 20.0
Hochberg (1984)[162]	<70 ≥70	75 75	4.0 12.0
MacArthur (1984)[157]	<70 ≥70	21009 1275	2.8 5.8
Horneffer (1987)[163]	<55 55-69 ≥70	228 228 228	2.2 2.2 9.3
Loop (1988)[164]	65-74 ≥75	4603 467	2.0 4.7
Goldman (1988)[118]	<70 ≥70	2887 340	2.9 6.2
Horvath (1990)[165]	<75 ≥75	4385 222	3.1 10.8
Freeman (1991)[158]	≥80	62	12.9
Ko (1991)[166]	≥80	100	12.0
Tsai (1991)[167]	≥80	157	7.0
Hannan (1994)[168]	<75 ≥75	2604 4934	2.3 6.1
Curtis (1994)[169]	≥70	668	5.2
He (1994)[170]	≥70	1399	8.9
Katz (1995)[171]	≥70	628	6.4
Morris (1996)[172]	≥80	474	7.8

OM = operative mortality

Table 5.2 **Odds ratios (OR), 95% confidence intervals (CI), and corresponding risk weights from logistic regression analysis of 19,009 patients undergoing isolated CABG between 1982-1996.**

Variable	OR (CI)	Risk Weight
Female	1.8 (1.5, 2.2)	2
LV Grade: 2 3 4	1.2 (1.0, 1.6) 2.3 (1.8, 2.9) 6.2 (4.4, 8.6)	1 2 6
Timing of Surgery: Semi-Urgent Emergency	1.2 (1.0, 1.5) 3.4 (2.6, 4.4)	1 3
Previous CABG	3.0 (2.3, 3.9)	3
Left Main Stenosis	1.6 (1.3, 2.0)	2
Age: 70-74 ≥75	1.6 (1.4, 2.0) 2.5 (1.9, 3.4)	2 3
Total Possible Score	-	0 - 19
Risk Score Cutpoints: Low Risk Medium Risk High Risk	-	 0 - 3 4 - 6 ≥7

Legend for Table 5.2 OR (CI) = odd ratio and the 95% confidence interval.

Reference categories in order are; male, LV Grade 1, elective surgery, no previous CABG, no left main stenosis, age<70 years.

Table 5.3 **Distribution of risk factors for 19,009 patients undergoing isolated CABG between 1982 and 1996 at the Toronto Hospital**

Variable		1982 - 1986		1987 - 1991		1992 - 1996	
		<70	≥70	<70	≥70	<70	≥70
N		5315	593	5067	1011	5297	1726
N (%)	Sex: Male	4409 (83)	423 (71)	4173 (82)	735 (73)	4342 (82)	1199 (69)
	Female	906 (17)	170 (29)	893 (18)	276 (27)	955 (18) ns	527 (31) ns
	Age: <65	4451 (84)	0	3869 (76)	0	3916 (74)	0
	65-69	864 (16)	0	1197 (24)	0	1381 (26)**	0
	70-74	0	457 (77)	0	699 (69)	0	1072 (62)
	≥75	0	136 (23)	0	312 (31)	0	654 (38) **
	LV Grade: 1	2301 (44)	252 (44)	1783 (36)	315 (32)	1564 (30)	463 (27)
	2	1841 (35)	189 (33)	2061 (42)	394 (40)	2403 (45)	800 (46)
	3	946 (18)	125 (22)	932 (19)	230 (24)	1160 (22)	404 (23)
	4	173 (3.3)	13 (2.2)	178 (3.6)	38 (4.0)	166 (3.1) **	54 (3.1) **
	Previous CABG: No	5028 (95)	584 (98)	4754 (94)	962 (95)	4931 (93)	1607 (93)
	Yes	287 (5.4)	9 (1.5)	293 (6.2)	49 (4.8)	366 (6.9) **	119 (6.9) **
	Timing of surgery: Elective	4586 (86)	438 (74)	3732 (74)	587 (58)	3036 (57)	745 (43)
	Semi-urgent	412 (7.8)	60 (10)	1128 (22)	366 (36)	2093 (40)	926 (54)
	Emergency	317 (6.0)	95 (16)	206 (4.1)	58 (5.5)	168 (3.1) **	55 (3.2) **
	Left Main Disease: No	4676 (88)	479 (81)	4323 (85)	807 (80)	4380 (83)	1303 (75)
	Yes	636 (12)	114 (19)	741 (15)	204 (20)	917 (17) **	423 (25) **
	Risk Groups: *Low*	4023 (77)	163 (28)	3459 (71)	216 (22)	3403 (64)	306 (18)
	Medium	1006 (19)	320 (56)	1182 (24)	557 (58)	1487 (28)	952 (55)
	High	211 (4.0)	93 (16)	247 (5.0)	187 (20)	398 (7.5) **	463 (27) **

Different over TIME, * P<0.05, ** P<0.01, ns=not significant

Table 5.4 **Distribution of Outcomes for 19,009 Patients Undergoing Isolated CABG Between 1982 and 1996 at the Toronto General Hospital**

Variable	1982 - 1986		1987 - 1991		1992 - 1996	
	<70	≥70	<70	≥70	<70	≥70
N	5315	593	5067	1011	5297	1726
(%) **Overall OM**	3.1	7.2	1.9	4.4	1.6 **	4.5 *
Risk-Stratified OM:						
Low Risk	1.9	1.8	1.0	2.3	0.7 **	3.6 ns
Medium Risk	5.8	5.9	2.4	3.4	2.2 **	2.6 *
High Risk	12.8	17.2	9.7	9.1	6.8 *	8.9 *
Sex: Male	2.7	5.4	1.9	3.8	1.4 **	3.8 ns
Female	4.9	11.8	2.1	5.8	2.8 **	6.3 *
Age: <65	2.8	-	1.5	-	1.4 **	-
65-69	4.5	-	3.3	-	2.2 **	-
70-74	-	6.4	-	3.9	-	3.9 ns
≥75	-	10.3	-	5.4	-	5.5 ns
LV Grade: 1	1.8	6.0	1.5	4.1	0.8 **	1.7 *
2	2.8	5.8	1.2	4.3	1.0 **	4.8 ns
3	5.3	11.2	3.1	3.0	3.2 *	5.2 **
4	12.1	7.7	7.9	13.2	7.2 ns	18.5 ns
Previous CABG: No	2.8	7.0	1.7	4.2	1.4 **	4.0 **
Yes	8.0	22.2	6.1	8.2	5.2 ns	11.8 ns
Timing of surgery: Elective	2.5	5.5	1.4	2.0	1.2 **	3.6 *
Semi-Urgent	6.1	5.0	3.0	6.6	1.7 **	5.0 ns
Emergency	7.9	16.8	5.8	13.8	8.3 ns	9.1 ns
Left Main Disease: No	2.9	6.0	1.7	3.8	1.4 **	4.3 ns
Yes	4.6	12.3	3.4	6.4	2.5 ns	5.2 *
Perioperative MI	5.7	8.3	3.7	4.1	2.6 **	3.7 **
LCOS	10.2	20	10.2	16	6.8 **	11 **
Perioperative Stroke	1.5	3.9	1.7	4.7	0.8 **	2.6 **

Legend for Table 5.4 All values are percentages with the exception of the first row (N).

OM = operative mortality, MI = myocardial infarction, LCOS = low cardiac output syndrome.

Different over TIME, * P<0.05, **P<0.01, ns=not significant

Table 5.5 **Logistic Regression of 3,330 Elderly Patients Undergoing CABG Between 1982 and 1996 at the Toronto General Hospital**

Variable	Regression Coefficient	Odds Ratio	95% CI
Year: 1986-1991	-0.635	0.53	0.3-0.9
1992-1996	-0.736	0.48	0.3-0.8
Age ≥75 years	0.335	1.40	1.0-2.0
Female	0.663	1.94	1.4-2.7
Previous CABG	1.194	3.30	2.0-5.6
LV Grade: 2-3	0.511	1.67	1.1-2.5
4	1.796	6.02	3.1-12
Timing of Surgery: Semi-Urgent	0.328	1.39	0.9-2.0
Emergency	0.882	2.41	1.4-4.1
Left Main Disease	0.420	1.52	1.1-2.2
Constant	-3.678	0.03	

Legend for Table 5.5 The reference categories in order are as follows: 1982-1986; 70-74 years; Male; No Previous CABG; LV Grade 1; Elective Surgery; and No Left Main Disease. The ROC curve area for the model is 0.69; the Hosmer-Lemeshow goodness-of-fit P value is 0.232.

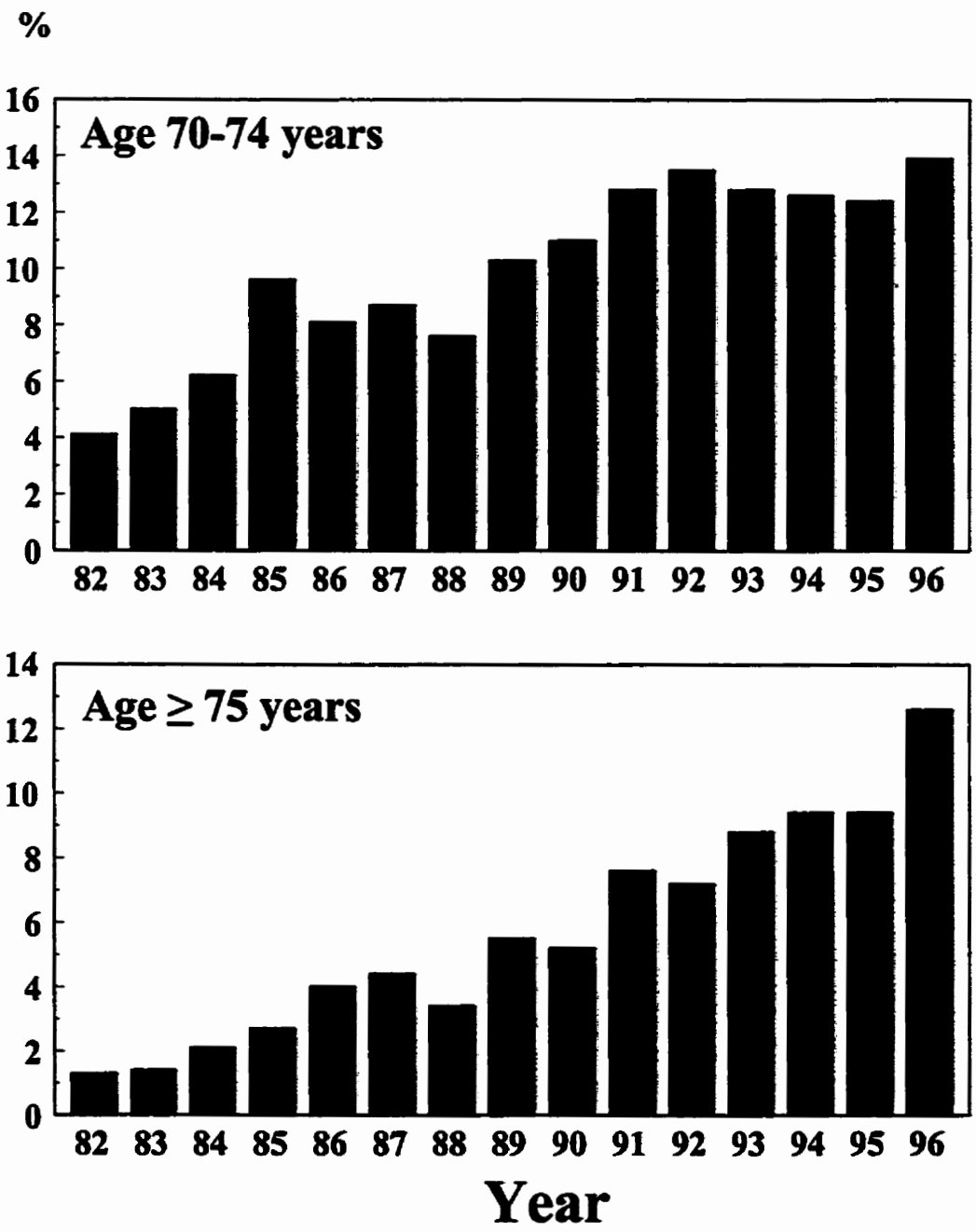
Table 5.6 **Logistic Regression of Contemporary Elderly Patients (1991-1996)**

Variable	Regression Coefficient	Odds Ratio	95% CI
LV Grade: 2-3	0.971	2.64	1.3-5.2
4	2.371	10.7	4.4-26
Previous CABG	1.316	3.73	2.0-7.0
Female	0.570	1.77	1.1-2.8
Peripheral Vascular Disease	0.566	1.76	1.1-2.7
Previous Angioplasty/Stent	-1.342	0.26	0.1-0.9
Diabetes	0.527	1.69	1.1-2.7
Constant	-4.493	0.01	

Legend for Table 5.6 Hosmer-Lemeshow goodness-of-fit P value = 0.932, ROC area = 0.713.

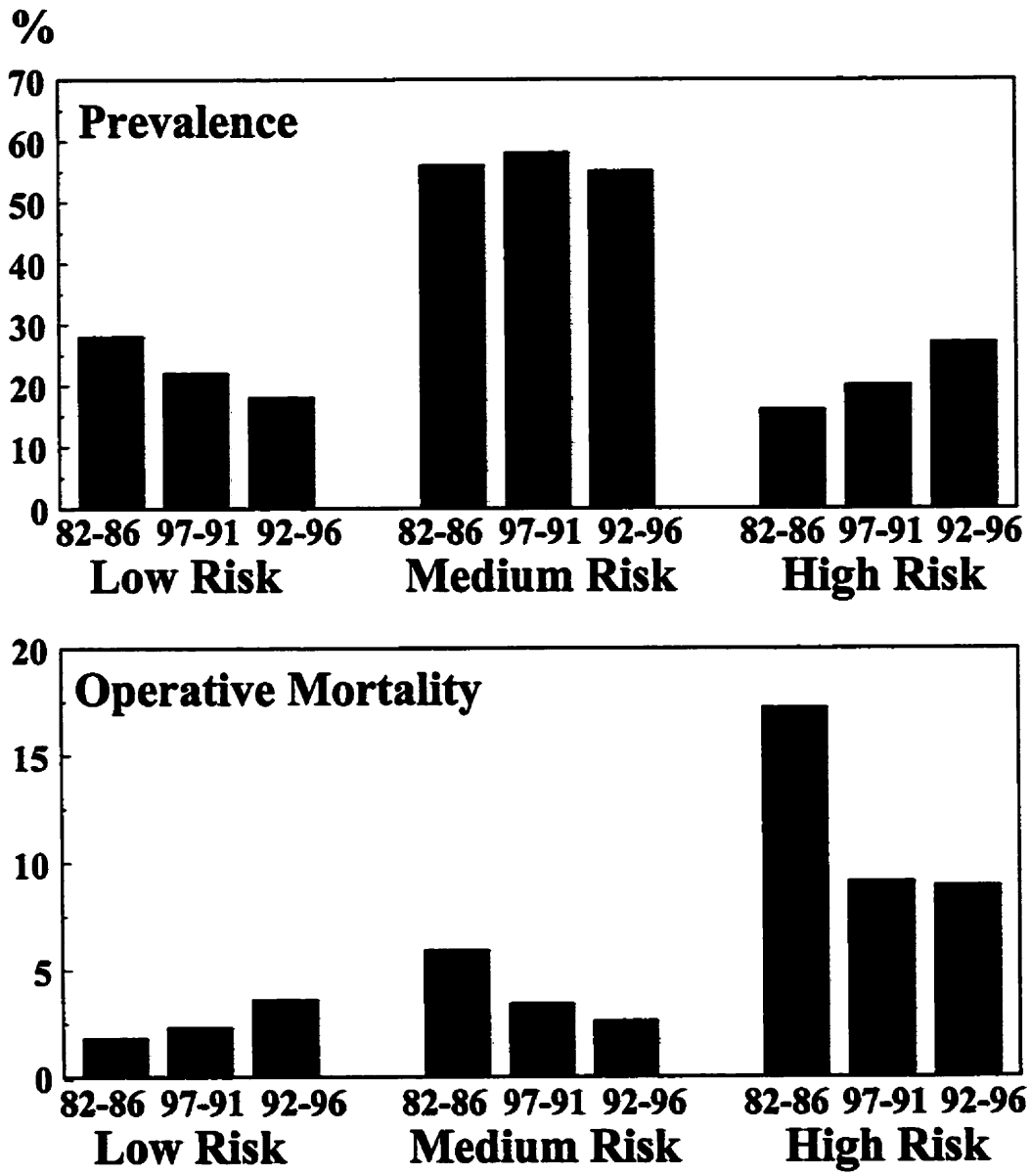
Note the risk reduction associated with previous angioplasty/stent.

Figure 5.1



Legend 5.1 The figure shows the significant increase in the prevalence of elderly patients undergoing isolated CABG at The Toronto Hospital between 1982 and 1996. The temporal change was significant ($P<0.001$) for both those aged 70-74 years and those aged 75 years and older.

Figure 5.2



Legend 5.2 The prevalence (top panel) of high risk elderly patients increased significantly ($P<0.001$) over time. Operative mortality (bottom) panel decreased significantly ($P<0.05$) over time for both medium and high risk elderly patients.

Chapter Six

Predictive Accuracy Study: Comparing a Statistical Model To Clinicians' Estimates of Outcomes After Coronary Bypass Surgery *

* Adapted from:

Ivanov J, Borger MA, David TE, Cohen G, Walton N: Predictive Accuracy Study: Comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery. Ann Thorac Surg 2000;70:162-168.

6.1 ABSTRACT

The purpose of this study was to compare clinicians' prior probability estimates of operative mortality (OM) and prolonged ICU length of stay >48 hr (ICU>48h) following coronary artery bypass graft surgery (CABG) to estimates derived from statistical models alone.

Methods: Nine clinicians estimated the predicted probability of OM and ICU>48h from an abstract of information for each of 100 patients selected from the 1996-97 database of 1904 patients who underwent isolated CABG.. Logistic regression models were used to calculate the predicted probability of OM and ICU>48h for each patient. The study sample was split into two parts; clinicians were randomly given access to a predictive rule to guide their judgements for one part of the study.

Results: Clinicians' estimates were similar with or without access to the rule, and both parts of the study were therefore pooled. Clinicians significantly over-estimated the probability of OM (Model 6.3 ± 1%, Clinicians 7.6 ± 3%, P=0.0001) and ICU>48h (Model 25 ± 2%, Clinicians 28 ± 1%, P=0.0012). Clinicians' estimates of OM were not significantly higher than the model's for non-survivors (0.8±0.7%, P=0.2), but were significantly higher for survivors (1.4±0.3%, P=0.039). Areas under the ROC curves did not differ for either outcome between the model or the clinicians. However, after adjusting for the deliberately skewed prevalence of the outcomes in the sample, the statistical models demonstrated superior discrimination compared to the clinicians: OM at 2% prevalence: (Model ROC 0.974 ● 0.026, Clinicians ROC 0.720 ± 0.069, P<0.001); ICU>48h at 16% prevalence: (Model ROC 0.942 ± 0.014, Clinicians ROC

0.760 ± 0.025, P<0.001).

Conclusions: Clinicians trusted their own empiric estimates rather than a predictive rule and over-estimated the probability of OM and ICU>48h. These findings highlight both the rationale for using predictive rules in practice and the remaining challenges in persuading clinicians to use such rules.

6.2 INTRODUCTION

As noted in preceding chapters, many clinical predictive rules have been developed for cardiac surgery to characterize case-mix and estimate the probability of outcomes. The primary purpose of these predictive rules has been to calculate risk-adjusted outcomes for provider profiling.[14,21,32,38,39,41,65,67,72,95,149,150] Other applications for statistical models include: assisting administrators in determining resource allocation, helping providers assess quality of care, and identifying high-risk subgroups for research purposes.[23,26,32,38,41,73,96,149] As well, patients may ask for probabilistic information regarding diagnosis, prognosis and response to treatment.[15,96,114,185] In this chapter we evaluate the use of a predictive rule as a tool to guide clinicians' estimates of the probability of hospital outcomes.

Statistical models can be used to facilitate discussions between surgeons and their patients regarding the risks and benefits of a particular procedure. However, clinicians sometimes claim that their intuition and experience gives them advantages over any statistical model in predicting adverse outcomes, and that the variables included in models reflect averages, rather than patient-specific clinical features that may put patients at risk. We accordingly undertook a study to 1) compare the judgements made by a statistical model to those made by clinicians, and 2) evaluate the use of a predictive rule as an aid to clinicians in making estimates of the probability of operative mortality and prolonged ICU length of stay greater than 48 hours. To give maximum advantages to clinical judgement, we compared predictive performance for a group of patients assembled purposively to capture a cross-section of both high- and lower-risk individuals who had suffered adverse events, and provided clinicians with additional clinical

information not incorporated into the statistical model.

6.3 METHODS

6.3.1 Sampling frame and Case preparation

We drew 100 cases (study sample) from the 1904 patients who underwent isolated CABG at the Toronto General Hospital between January 1996 and April 1997 (hereafter, the study population). We deliberately over-sampled for 70% of deaths, drawing 29 of 41 post-operative fatalities from the general patient population, covering a range of higher- and lower-risk patients who died after surgery. The remaining 71 patients were sampled from survivors. The 100 study patients were then randomly divided into two 50 patient groups, Parts I and II.

A half-page abstract was prepared for each patient. Again, to support clinical judgements, the abstract contained a table of information covering many more variables than the statistical models which were derived and validated by our group.[39,72,73] The variables drawn from our clinical cardiac surgery database included: age, sex, LV ejection fraction, New York Heart Association functional class, anginal symptoms (i.e., stable, unstable, acute coronary insufficiency), previous CABG, timing of surgery (i.e., elective, same-hospitalization, urgent, emergency), left main disease, positive exercise stress test, recent myocardial infarction within the month prior to surgery, diabetes, peripheral vascular disease (including carotid stenoses), history of hypertension, renal insufficiency, height and weight. Results of coronary angiography were also shown in detail as a diagram along with information regarding the quality of the vessels, i.e., degree of stenosis, and vessel size, quality of distal vessels, and collateral flow, when

available. A brief narrative was prepared which described each patient's relevant history, presenting scenario and other comorbidity. Table 6.1 contains an example of additional, relevant information which was not contained in the database but was added to the narrative.

6.3.2 Predictive Rule

A predictive rule, developed from the combined CABG database of two Toronto hospitals for the years 1993 to 1996, was used as the format for the contemporary, site-specific guidelines. Risk weights for this rule were re-calibrated in the 1993 to 1997 database for the Toronto General Hospital only. Risk weights and cutpoints of the total risk score which defined relative risk groups (low, medium, high) can be found in Table 6.2. Details regarding the development of this rule have been published.[72] By presenting a simple and additive scoring system, we hoped to maximize the system's acceptance by clinicians.

6.3.3 Clinician Sample

All nine cardiac surgeons who performed CABG surgery at the Toronto General Hospital in 1997-1998, as well as two cardiac surgery residents and a specialty nurse-clinician were invited to participate. Two senior clinicians declined the invitation to participate and one senior clinician had difficulty with the format and also decided not to participate. Senior cardiac surgeons are designated in the figures as clinicians #1-#5, #6 is a junior cardiac surgeon, the two speciality residents are #7 and #8 and the speciality nurse is clinician #9.

Clinicians were randomized to receive the predictive rule as an aid for Part I or Part II of the study. After a four month 'wash-out' period, clinicians who received the predictive rule for Part I, did not receive it for Part II of the study, and vice versa.

6.3.4 Outcomes

Clinicians were asked to estimate, for each patient, the probability of 1) operative mortality (OM), defined as any postoperative, in-hospital death, and 2) prolonged ICU length of stay defined as an ICU length of stay greater than 48 hours (ICU>48h), which corresponded to the 83rd percentile of ICU length of stay for 1995-1997.

6.3.5 Statistical Model

The model was essentially similar to the original predictive rule used to assist clinicians. Risk factors from the original rule, in addition to other important prognostic variables, were submitted to a logistic regression analysis to recalibrate regression coefficients and optimize precision.[72]

In 1995, the pattern of practice changed regarding discharge from the ICU: patients were extubated earlier and discharged sooner than previous patients because of "fast-tracking".[186,187] Because of this changing pattern, the logistic regression model for the prolonged ICU stay outcome was re-derived for only the 1995-1997 database of isolated CABG patients. An ICU stay of >48 hours corresponded to the 83rd percentile of total ICU length of stay.

Regression coefficients from the newly derived models were used to calculate the patient-specific predicted probability of each outcome using methods previously

described.[72,73,105,128]

6.3.6 Statistical Analysis

Data were managed in dBASEIV datasets and analyzed using SAS for Windows Version 6.12.[151] Results are presented as means ± standard errors. A two-tailed P value <0.05 indicates statistical significance unless otherwise noted.

The estimates of probability for OM and ICU>48h with and without access to a predictive rule were evaluated by unpaired *t* test. The difference between the clinician's estimate of the probability of OM and ICU>48h and the model's estimate was calculated for each patient (clinician minus model) and evaluated by paired *t* test. The null hypothesis for this analysis was that the difference in the predicted probability of an event equalled zero.

Calibration, or precision at the group level, was evaluated by the Hosmer-Lemeshow goodness-of-fit chi square statistic (HL).[105] Hosmer-Lemeshow goodness-of-fit P values <0.05 indicate a significantly imprecise model (i.e. the model did not fit the data). Calibration is also important because, in contrast to the match between probabilities and categorical events reflected in the ROC curve area, it captures the ability of the model (or clinicians) to predict the absolute event rates on average.

The Brier score (BS) is a quadratic penalty score with values between 0 and 1. The Brier scores measures both accuracy and precision at the individual patient level. The lower the Brier score, the more accurate the judgement. The formula is as follows:

$$BS = (p - d)^2$$

where P=predicted probability of outcome, d (event) 0=event did not occur, 1=event

occurred.

The precision of a set of probabilities are evaluated by calculating the associated Z statistic and P value. A P value <0.05 indicates a model which is significantly imprecise.[110-114,185,188-192] The difference between clinicians' and the models estimates of Brier scores was evaluated by non-parametric analysis of variance, the Kruskal-Wallis non-parametric rank analysis as well as by parametric ANOVA and *t* tests.

Discrimination (or predictive accuracy) was assessed by calculation of the area under the Receiver-Operator Characteristic curve (ROC) for each set of estimates.[108,109] The area under the ROC curve reflects the proportion of randomly paired sets of patients for which the patient experiencing the event has a higher predicted probability of having the event, compared to the patient who does not experience the event. An area under the ROC curve of 50% indicates no discriminatory ability; ROC areas above 70% represent fair discriminatory ability and ROC areas above 80% represent good discriminatory ability. The area under the ROC curve is independent of the prevalence of the outcome in the database. However, since the study sample was deliberately skewed towards patients who had events, we adjusted the prevalence of each outcome by the following strategy: 1) patients who died were ranked in order of their predicted probability of OM (lowest to highest), 2) patients with the lowest predicted OM were incrementally changed from "died" to "survived", thereby reducing the overall prevalence of OM in the dataset, 3) ROC and HL GOF were recalculated for each step, 4) this strategy was repeated for the outcome, ICU>48h. The theory behind this exercise was that, as the prevalence of each outcome was decreased in favour of those having a lower predicted probability of having an event, the ability of the model (or clinicians) to discriminate

patients at higher risk should have approached 100%. The comparisons of ROC curves for clinicians versus the statistical model were performed using methods described by Hanley and McNeil.[108,109]

6.4 RESULTS

6.4.1 Statistical Models

There were 7313 patients who underwent isolated CABG at the Toronto General Hospital between 1993 and 1997. Operative mortality (OM) in this group was 2.3% (n=169). In the 1995 to 1997 database of 4572 patients undergoing isolated CABG, the prevalence of ICU>48h was 17% (n=778).

Logistic regression coefficients for both the OM model and the prolonged ICU stay model are found in Table 6.3. Both models demonstrated excellent discrimination and precision.

6.4.2 Study Population

For the study population of 1904 consecutive patients undergoing isolated CABG between January 1996 and March 1997, OM was 2.3% and ICU>48h was 15.6%. Regression coefficients from the newly derived models for OM and ICU>48h were applied to these data to calculate the predicted probability of each event. The statistical models demonstrated good discrimination and precision in the study population: for OM, ROC was 0.76 and HL P=0.104. For ICU>48h the ROC was 0.71, with HL P=0.51.

6.4.3 Sample characteristics

Table 6.4 depicts the demographics of the study population and the 100 patient study sample. The deliberate skew of the study sample is evident in both the prognostic variables and the prevalence of outcomes. There was no group level difference between Part I and Part II for any prognostic or outcome variable -- on average, the two groups looked the same with regard to risk factors and outcomes.

6.4.4 Impact of access to a Predictive Rule

Access to a predictive rule did not significantly alter overall predictive performance for the group for either outcome (Table 6.5). With the exception of one cardiac surgery resident, there were no statistical differences in clinicians' predicted probabilities of OM made with or without access to the rule. Access to the predictive rule was associated with significantly higher estimates of the probability of ICU>48h for two clinicians and significantly lower estimates for two other clinicians (Figure 6.1) Moreover, when clinicians were surveyed after the study, only one junior staff surgeon and the nurse clinician consistently referred to the predictive rule as a guideline for making their probability estimates. Both parts of the study were therefore combined for comparisons between clinicians' prior probability estimates of outcomes versus those made by a statistical model.

6.4.5 Difference Between Clinicians and the Model

Predicted Probabilities of Outcomes

Figure 6.2 shows the difference by paired *t* test between each clinician's predicted

probability and the model's calculations for OM (top panel) and ICU>48h (bottom panel). The differences for OM were significant for the experienced clinicians (#1, #2, #3, #5, #6) whereas the more inexperienced clinicians (#7 - #9) demonstrated no significant difference from the model's predictions. The junior clinicians (#7-#9) estimates were significantly different from the senior clinicians (#1 - #5) by analysis of variance (P<0.05).

Figure 6.2 (bottom panel) shows the results of the paired *t* tests evaluating the differences between the clinicians and model for predicted probability of ICU<48h. In this instance, there was no relationship between clinical experience in either direction or magnitude of discrepancies.

Table 6.6 shows the results of the pooled clinician sample compared to the statistical model. The clinicians significantly over-estimated both total OM and ICU>48h. However, when we evaluated those having events versus those not having the event, the differences for predictions of OM were significant only for survivors (Clinicians 7.0±9%, Model 5.6±7%, P=0.0001); there was no significant difference in OM predictions for the non-survivors (Clinicians 9.0±10%, Model 8.1±9%, P=0.23). Clinicians estimates of prolonged ICU time were significantly higher than the Model's for both those not experiencing the event (Clinicians 22±22%, Model 20±15%, P=0.03) and those who did have an ICU length of stay >48 hr (Clinicians 39±28%, Model 34±23%, P=0.01).

Accuracy

With the exception of one surgeon, predictive accuracy (or discrimination) as measured by the area under the ROC curve was poor for all sets of judgements for OM

(Figure 6.3). Discrimination was higher, in general for ICU>48h compared to OM. There was no difference in ROC curves between the pooled clinician sample versus the statistical model for either OM or ICU>48h (Table 6.6). However, for OM, six of nine clinician had ROC curves with lower values than the statistical model's ROC.

The prevalence of OM in the Parent database was 2.3%. The prevalence of ICU>48h in the Parent database was 15.3%. By comparison, the prevalence of each outcome in the 100 patient sample was extremely skewed by design: OM 29%, ICU>48h 36%. The prevalence of each outcome was decreased by incrementally reversing the outcome for patients with the lowest probability of experiencing the event. Table 6.7 and Table 6.8 display the results of this strategy for OM and ICU>48h respectively. ROC curves for the statistical models approached 100% as was expected in this manoeuvre. However, ROC curves for the clinicians estimates did not increase as expected and were significantly lower than ROC curves for the statistical models.

There were no significant differences in Brier scores for OM (Figure 6.4, top panel). All Brier scores were associated with Z statistics >1.96 indicating significant imprecision of the judgements. Brier scores for ICU>48h were significantly different between clinicians by ANOVA and the Kruskal-Wallis rank test ($P<0.001$). Although overall precision was poor, two clinicians had judgements associated with Z statistics (1.55 and 0.44) indicating excellent precision (Figure 6.3, bottom panel).

6.5 DISCUSSION

The ability of physicians to make probabilistic judgements may be linked to the quality of care they provide as well as to the allocation and consumption of

resources.[114,185,191,193] As a result, predictive rules have been developed in many areas of medicine to aid clinicians in making probabilistic judgements.[38,96,194] In cardiac surgery, predictive rules have been developed primarily for coronary artery bypass surgery, focusing on post-operative mortality and length of hospital stay.[14,21,32,38,39,41,65,67,95,149,150] It has been suggested that these models may be helpful variously for patient counselling regarding the risks and benefits of a procedure, identification of high risk subgroups for research purposes,[23,26,32,38,41,96,149] and calculation of surgeon-specific, risk-adjusted outcomes of coronary bypass surgery.[72]

In this study we examined the comparative performance of clinicians and a statistical model in predicting adverse outcomes after CABG. We presented a group of clinicians with case abstracts for 100 patients, and deliberately skewed the sample by including 29 patients who died post-operatively. The additional information contained in the narrative section of the patient histories, coupled with the skew towards higher risk patients who experienced adverse events, was designed to give clinicians an advantage over the statistical model's estimates of probability. We postulated that clinicians' conceptual flexibility and intuition might allow them to identify patients at special risk more accurately than a statistical model. To maximize the performance of the statistical model we drew the study sample of 100 patients from the population of patients in which the rule was recalibrated.

Prior to the start of the study, the study protocol was circulated to each clinician and followed by two emails asking for comments or questions. The purpose and details of this study were then presented at divisional rounds on two separate occasions. The predictive rule as depicted in Table 6.2 was provided to each clinician for one phase of

the study. The four month "wash-out" period between the two parts of the study was designed to prevent contamination. We were not particularly surprised that the majority of clinicians elected not to use the rule consistently to calculate patient-specific probabilities of each outcome. Indeed, we found no differences in predictive accuracy whether clinicians were exposed to the rule or not, and whether they claimed to use it consistently or not. We were surprised that two very senior surgeons would not participate in the study despite repeated requests from the primary investigator. These findings highlight the tendency of clinicians to rely on their own intuitive judgements rather than a statistical model.

Given the lack of impact of access to the predictive rule, we analyzed the results for both phases of the study together. Unlike previous studies which have shown that clinicians outperformed quantitative models, [99,189,191] a robust analysis of the paired difference of each clinician's estimates of probability versus those made by the statistical model revealed that senior surgeons significantly over-estimated the probability of operative mortality. One possible explanation for these results is that senior surgeons make typically broad predictions based on their recent experience[96] whereas the junior clinicians, lacking in experience, may rely more closely on published results. However, in a review of 100 studies comparing statistical to clinical judgments, Dawes and colleagues[195] concluded that statistical estimates equalled or surpassed the clinical method even when clinicians were given an information edge.

The relatively poor performance of the statistical model in the 100 patient study sample was understandable given the deliberately skewed prevalence of outcomes in the study sample. Indeed, as noted, we expected that the approach towards a study sample

with an increased prevalence of high risk patients having outcomes would give clinicians an advantage. It did not. Dawes et al[195] noted that a unique ability to observe is not the same as a unique ability to predict based upon the integration of observed information. They stated that research has shown that individuals have difficulty discerning between valid and invalid variables and commonly develop false beliefs in associations between variables. Others have highlighted the importance of disease prevalence when transporting clinical prediction rules.[99] To evaluate underlying discriminatory ability, we explored a strategy to test the predictive accuracy of both the model and the clinicians by incrementally reducing the prevalence of each outcome in the study sample by changing patients with the lowest predicted probability of an outcome from having the event, to not having the event. The results demonstrated that after this adjustment, the statistical models had significantly improved discrimination compared to the clinicians. These results were similar to those found in a study by Poses and colleagues[99] who found that discriminatory power was diminished when three predictive rules were tested in a population whose disease prevalence was significantly lower than the population from whence the rules were derived. After adjusting for disease prevalence by using the actual disease prevalence in their population as the pre-test estimate of disease probability, these investigators found that estimates made by all three rules were significantly more accurate compared to those made by the clinicians. The authors concluded that predictive rules might aid clinicians' decision-making provided the rules are accurate and properly calibrated.

One limitation of this study was that the clinicians were making estimates of mortality and prolonged ICU stay from a written synopsis of patient information rather

than direct examination. However, clinicians were given the advantage by being provided with additional relevant clinical information in the narrative portion of each patient abstract. Also, the diagram of the cardiac catheterization results included distal vessel quality if it was noted to be poor. Despite these advantages, the clinicians did not outperform the statistical models which relied solely on a limited number of variables. We doubt, therefore, that our findings would have been different if the study had taken place prospectively in practice as opposed to relying on written case abstracts.

As noted, the limited impact of access to a predictive rule was not surprising. Clinicians are trained to rely heavily on their own experience and intuitive judgements. Moreover, the literature on changing physician behaviour amply illustrates that clinical decision-making is not readily influenced unless there is acceptance of a new practice norm by 'opinion leaders' and a concerted local effort to address barriers to change, e.g. by automating decision support systems.[196,197] However, surgeons and anaesthetists at our centre are already using an implied form of prior probability estimates to fast-track low risk cardiac surgical patients through the ICU.[186,187] In this regard, resources have been streamlined for those patients deemed to be at a low risk of poor outcome and conversely, there is a greater concentration of resources for higher risk patients. We believe that with appropriate implementation strategies, clinicians will use statistical tools that enable them to make more accurate judgements of patients' risks of adverse outcomes after CABG.

Summary

Predictive rules have the advantage of being able to integrate more information by relating both continuous and categorical variables to an outcome of interest. An accurate and precise model can be used as a concise summary of the relationships between prognostic variables and the outcome.[7,86]

Cardiac surgery clinicians, when given the option, preferred not to use a predictive rule but rather trusted their own judgements to estimate the probability of operative mortality or prolonged ICU length of stay after coronary bypass surgery. Experienced surgeons significantly over-estimated the risk of operative mortality compared to their junior colleagues and compared to the statistical model. Clinicians' predictive accuracy was only fair for operative mortality, and only slightly better for estimates of prolonged ICU length of stay. Although no model can predict the specific individual who will have an adverse event, statistical models do permit reasonably accurate estimates of event rates for subgroups of patients. If methods were devised to ensure that clinicians can and do make use of predictive rules, they would be able to make more accurate judgements of the probability of adverse outcomes following coronary artery bypass graft surgery.

Table 6.1 **Examples of Additional Information Included in the Patient Narrative**

- Patient previously denied surgery because of high risk
- Patient previously refused surgery
- Recent MI with continued post-infarct angina
- Patient required resuscitation preop due to cardiac arrest
- Patient admitted with transmural MI leading to pulmonary edema and shock requiring intubation and resuscitation
- Patient required IV heparin and/or NTG due to severe, unstable angina
- Patient allergic to heparin
- Severe vasculopath
- Angina interfering with quality of life ("afraid to do things")
- SOB on exertion. Presently off work because of angina, four kids at home
- Patient states that he "knows he is going to die"

Legend for Table 6.1 Information included in the patient narrative was designed to give clinicians an advantage by more fully characterizing the patient.

Table 6.2 **Predictive Rule Used As a "Tool" to Aid Clinicians in Making Probability Estimates of Relative Risk of Operative Mortality and Prolonged ICU Stay >48 Hours** [72]

Variable	Risk Weight
LV Grade: 2 3 4	2 3 10
Age: 65-74 ≥75	2 3
CABG Redo	4
Timing: Semi-Urgent/Urgent Emergency	1 3
Female	2
Peripheral Vascular Disease	2
Diabetes	2
Left main stenosis	2
Hypertension	1
Total Possible Risk Score	**0-29**

Legend for Table 6.2 Risk weights which characterize each patient are summed for a "total risk score". Cutpoints of the total risk score describe each patient's relative risk. The percentages for operative mortality (OM) were determined from the original database.[72]

Low risk: Score 0-3 (OM: range 0 - 0.7%; average 0.4%)

Medium risk: Score 4-9 (OM: range 0.7% - 3.0%; average 2.0%)

High risk: Score ≥10 (OM: range >3.0%; average 9.0%)

Table 6.3 **Logistic Regression Coefficients to Calculate Predicted Probability of Operative Mortality and Prolonged ICU Stay >48 Hr.**

Variable	OM 1993-1997	ICU>48h 1995-1997
Age: 65-74	0.461 ± 0.19	0.341 ± 0.10
≥75	1.299 ± 0.22	0.633 ± 0.13
Female	0.624 ± 0.18	0.520 ± 0.10
LV Grade: 2	0.600 ± 0.26	0
3	1.179 ± 0.26	0.247 ± 0.10
4	1.897 ± 0.34	0.247 ± 0.10
Timing: Semi-Urgent/Urgent	0.159 ± 0.18	0.209 ± 0.26
Emergency	0.984 ± 0.32	0.247 ± 0.10
Previous CABG	1.106 ± 0.22	0.526 ± 0.16
Left Main Disease	0.349 ± 0.18	0.245 ± 0.10
Peripheral Vascular Disease	0.486 ± 0.18	ns
Hypertension	0.342 ± 0.17	ns
Renal Failure	0.472 ± 0.28	0.912 ± 0.20
Diabetes	ns	0.228 ± 0.10
Recent MI	ns	0.299 ± 0.11
Preop Intra-Aortic Balloon Pump	ns	1.569 ± 0.20
Congestive Heart Failure	ns	0.729 ± 0.13
Triple Vessel Disease	ns	0.296 ± 0.11
Chronic Obstructive Lung Disease	ns	0.306 ± 0.13
Preop Atrial Fib/Flutter	ns	0.402 ± 0.15
Constant	-5.750 ± 0.28	-2.879 ± 0.12
ROC	0.77	0.71
HL GOF	0.10	0.51

Legend for Table 6.3 To maximize performance, the statistical models were re-derived in the Toronto General Hospital database. OM = Operative Mortality. The OM model was derived from the 1993-1997 database of isolated CABG patients. ICU>48h =

prolonged ICU stay >48 hours. The ICU>48h model was derived from the 1995-1997 database.

ROC = area under the Receiver-Operator Characteristic curve, HL GOF=Hosmer-Lemeshow goodness-of-fit statistic, ns = variable was submitted to the model for consideration but was not a significant predictor of the outcome.

Table 6.4 **Comparing the 100 Patient Sample to the Parent Population**

	Study Population Jan 1996-Mar 1997	Study Sample
N	1904	100
Age: <65	51	37
65-74	37	40
≥75	12	23
Female	22	26
LV Grade: 1	26	22
2	49	38
3	22	26
4	3	14
Timing: Elective	50	39
Semi-Urgent/Urgent	48	37
Emergency	2	24
NYHA: 1	3	2
2	11	2
3	32	29
4	54	67
Previous CABG	6	17
Left Main Disease	18	21
Diabetes	26	27
Hypertension	55	68
Peripheral Vascular Disease	14	24
Renal Failure	1.3	2.0
Operative Mortality	2.2	29
ICU Stay >48 hrs	16	36

Legend for Table 6.4 The study population was comprised of all patients undergoing isolated CABG between Jan 01, 1996 and March 31, 1997. Results are presented as proportions.

Table 6.5 **Clinicians' Results for Using a Predictive Rule as an Aid to Predict Outcomes Versus Not Using a Rule**

	RULE	NO RULE
OM: Predicted Probability (%)	7.6 ● 1	7.6 ± 1
ROC (%)	62 ± 2	60 ± 2
BS (%)	25 ± 2	25 ± 2
ICU>48h: Predicted Probability (%)	28 ± 1	28 ± 1
ROC (%)	71 ± 2	73 ± 2
BS (%)	22 ± 1	23 ± 1

Legend for Table 6.5 ROC = Area under the Receiver-Operator Characteristic curve, BS = Brier Score, OM = Operative Mortality, ICU>48h = prolonged ICU stay greater than 48 hours.

Results are presented as the mean ● standard error. All BS were associated with Z statistics >1.96 and therefore, significantly imprecise. There were no statistically significant differences for clinicians between using a Rule or not using a Rule.

Table 6.6 **Comparing Predicted Probability Estimates of Operative Mortality and Prolonged ICU Stay >48 hr: Statistical Model Versus Clinicians**

	MODEL	CLINICIANS
OM: Predicted Probability (%)	6.3 ± 1	7.6 ± 3*
ROC	0.614 ± 0.021	0.603 ± 0.021
HL GOF p	<0.001	<0.001
BS (%)	25 ± 2	25 ± 2
ICU>48h: Predicted Probability (%)	25 ± 2	28 ● 1**
ROC	0.682 ± 0.020	0.689 ± 0.020
HL GOF p	<0.001	<0.001
BS (%)	22 ± 2	22 ± 1

Legend for Table 6.6 OM = Operative Mortality, ICU>48h = prolonged ICU stay greater than 48 hours, ROC = Area under the Receiver-Operator Characteristic curve, HL GOF = Hosmer-Lemeshow goodness-of-fit statistic. Results are presented as the mean ± standard error.

Clinicians different from Model, *P=0.0001, **P=0.0012. All Brier Scores (BS) were associated with P values <0.05 indicating significantly *imprecise* calibration.

Table 6.7 **Changing the Prevalence of Operative Mortality in the 100 Patient Sample: The Effect on Discrimination and Precision**

Probability Cutpoint	Observed OM	Estimates	ROC	HL
none	29%	Model Clinicians	0.614 ± 0.021 0.603 ± 0.021	<0.001 <0.001
<1.0%	27%	Model Clinicians	0.653 ± 0.022 0.601 ± 0.022	<0.001 <0.001
<2.0%	20%	Model Clinicians	0.779 ± 0.022 0.668 ± 0.024*	<0.001 <0.001
<6.0%	12%	Model Clinicians	0.851 ± 0.024 0.714 ± 0.029*	0.097 <0.001
<10.0%	8%	Model Clinicians	0.950 ± 0.018 0.783 ± 0.033*	0.456 0.001
<15.0%	6%	Model Clinicians	0.953 ± 0.020 0.767 ± 0.039*	0.697 0.002
<23.2%	4%	Model Clinicians	0.966 ± 0.021 0.769± 0.047*	0.752 <0.001
<23.5%	2%	Model Clinicians	0.974 ± 0.026 0.720 ± 0.069*	0.834 <0.001
<30.0%	1%	Model Clinicians	0.990 ± 0.023 0.720 ± 0.097*	0.714 <0.001

Legend for Table 6.7 Patients were ranked in order of their predicted probability of operative mortality (OM) as calculated by the statistical model. Patients with probabilities lower than the "cutpoint" were changed from "died" to "survived" in the database. Receiver-Operator Characteristic curves (ROC) and Hosmer-Lemeshow goodness-of-fit (HL) were recalculated for each step. The Model had statistically higher ROC areas compared to the Clinicians (* P<0.01). As the prevalence of OM decreased,

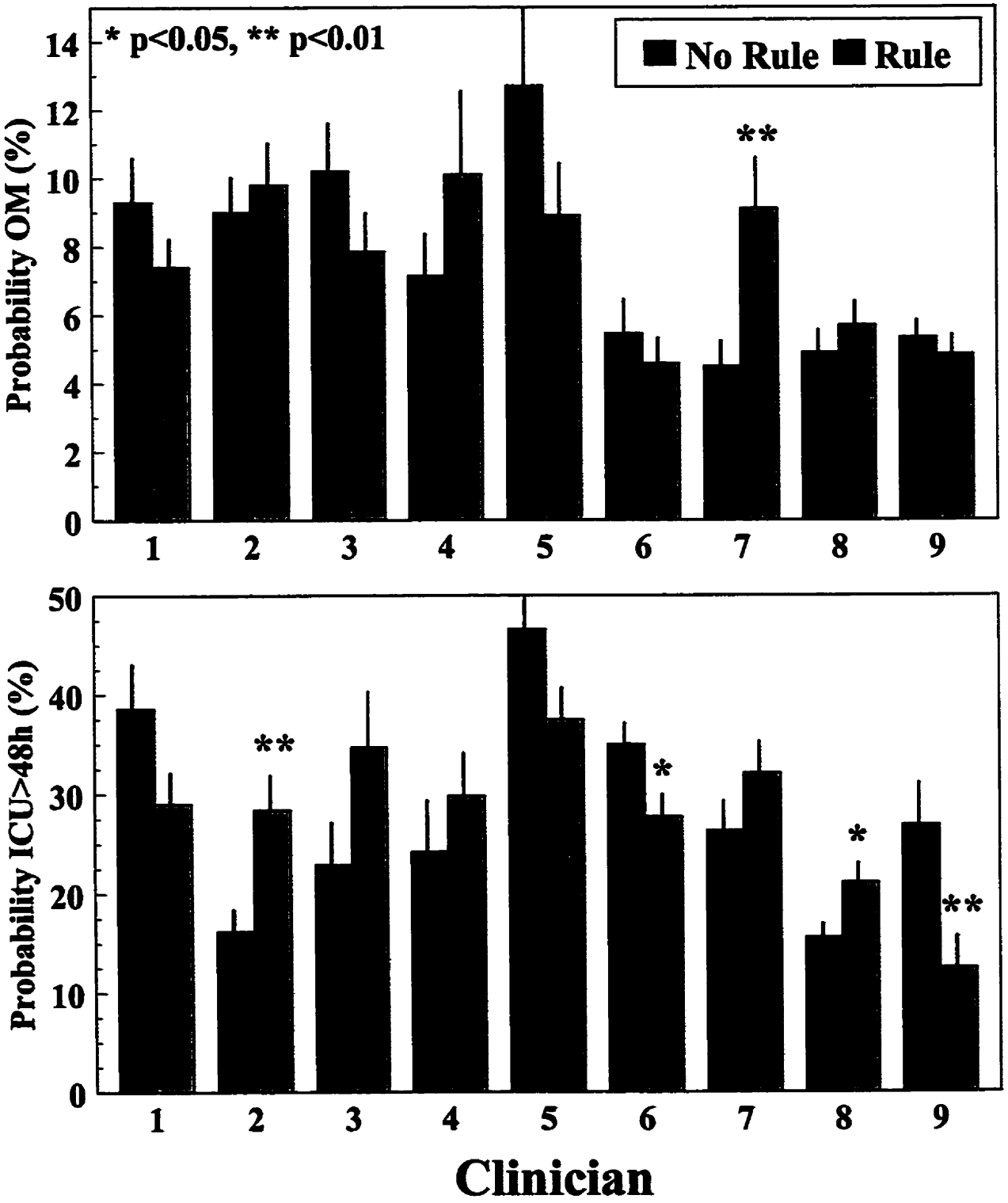
model discrimination did as expected, and improved to almost 100% The clinicians' estimates improved only to a maximum ROC area of 0.783. HL GOF for the clinicians remained significantly imprecise

Table 6.8 **Changing the Prevalence of Prolonged ICU Stay >48 Hr in the 100 Patient Sample: The Effect on Discrimination and Precision**

Probability Estimate	Observed ICU>48h	Estimates	ROC	HL
none	36%	Model Clinicians	0.682 ± 0.020 0.689 ± 0.020	<0.001 <0.001
<5%	34%	Model Clinicians	0.729 ± 0.019 0.668 ± 0.020*	0.017 <0.001
<12%	29%	Model Clinicians	0.814 ± 0.018 0.690 ± 0.021*	0.115 <0.001
<15%	26%	Model Clinicians	0.848 ± 0.017 0.696 ± 0.021*	0.131 <0.001
<20%	21%	Model Clinicians	0.897 ± 0.016 0.719 ± 0.023*	0.414 <0.001
<25%	18%	Model Clinicians	0.925 ± 0.015 0.764 ± 0.023*	0.336 <0.001
<30%	16%	Model Clinicians	0.942 ± 0.014 0.760 ± 0.025*	0.137 <0.001
<36%	12%	Model Clinicians	0.964 ± 0.013 0.767 ± 0.028*	0.102 <0.001
<45%	10%	Model Clinicians	0.990 ± 0.007 0.803 ± 0.029*	0.008 <0.001

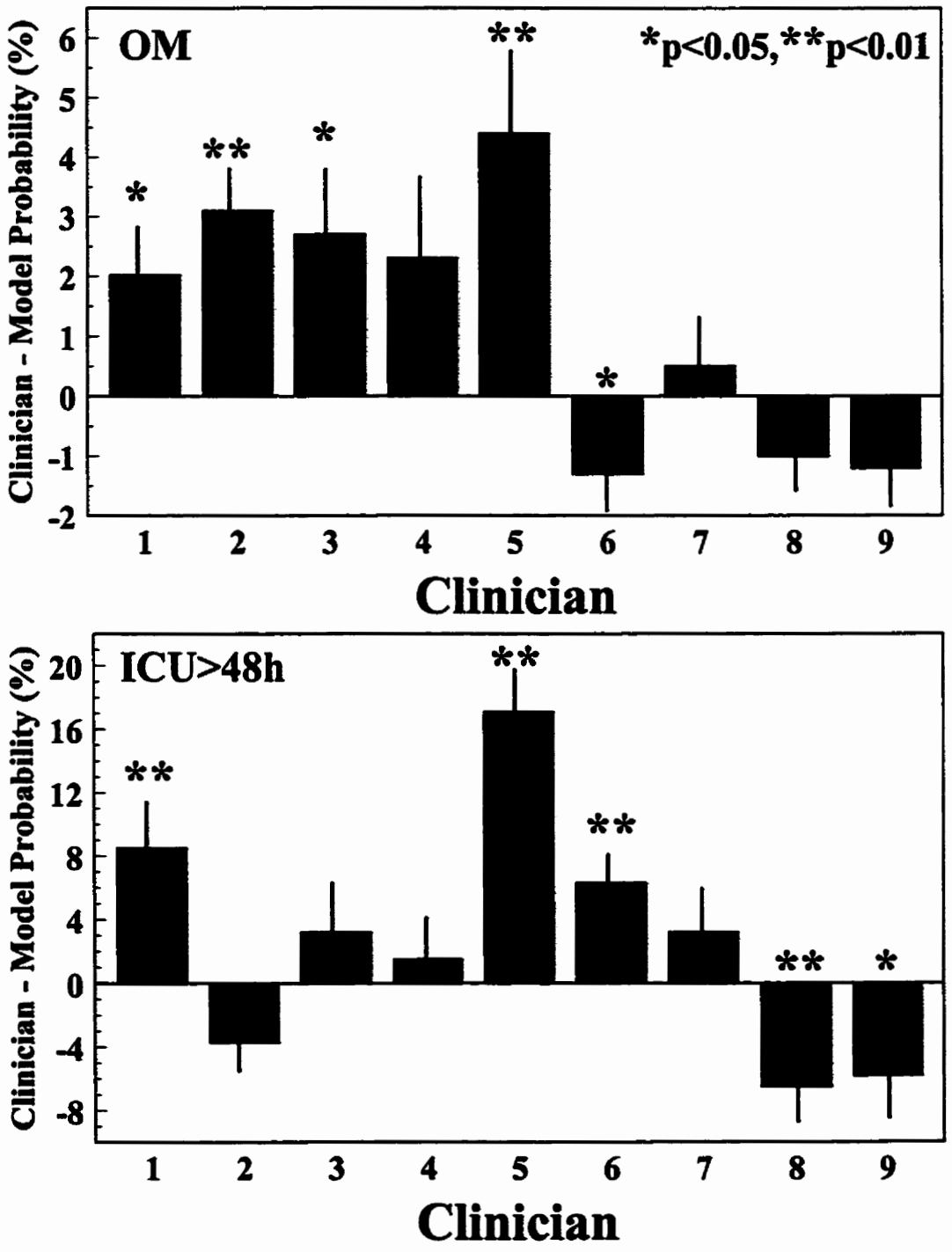
Legend for Table 6.8 The same strategy was employed as is described in Table 6.7. The Model had statistically higher ROC areas compared to the Clinicians (* P<0.01). As the prevalence of ICU>48h decreased, model discrimination did as expected, and improved to almost 100% The clinicians' estimates improved only to a maximum ROC area of 0.783. HL for the clinicians remained significantly imprecise

Figure 6.1



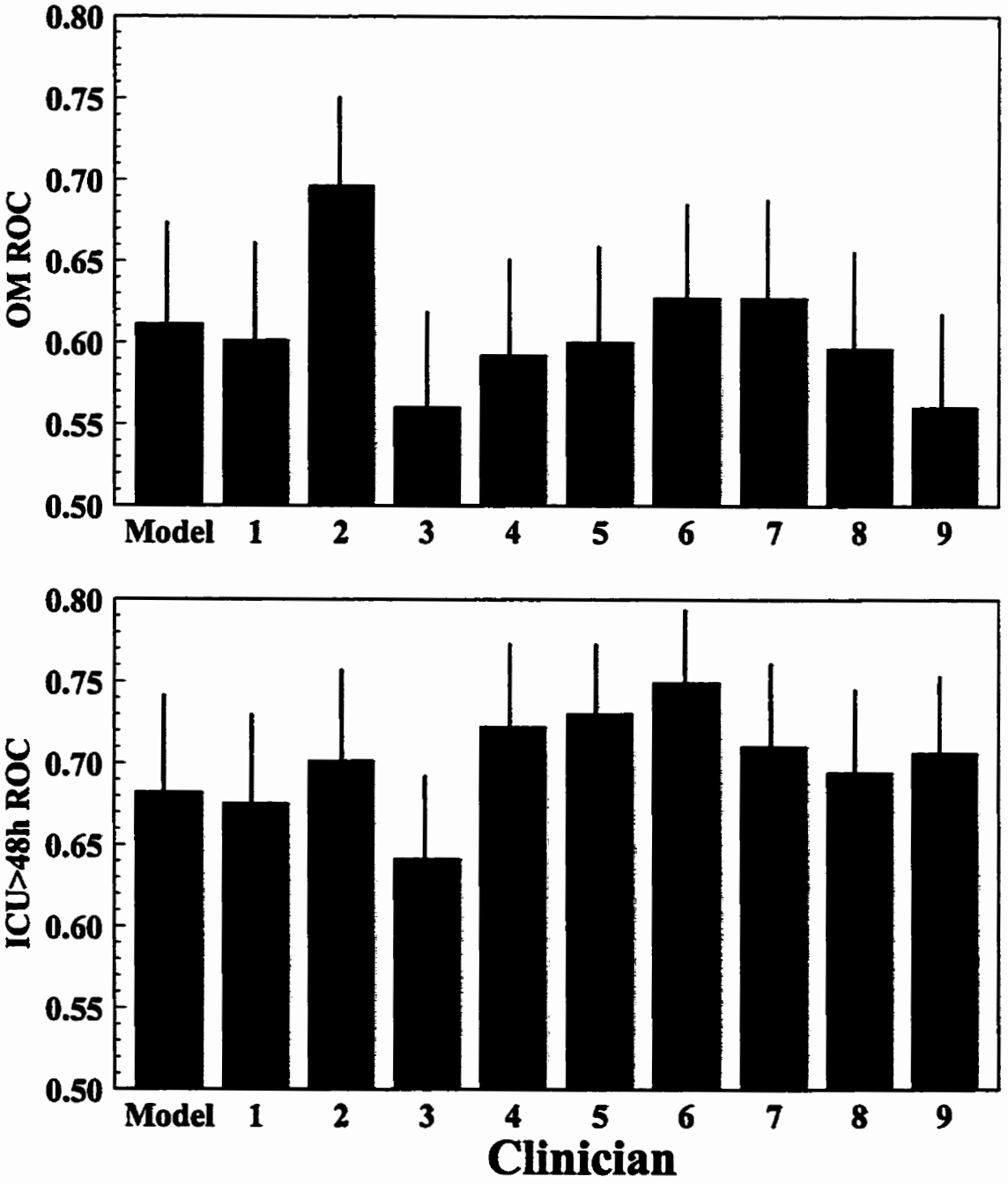
Legend 6.1 Predicted probabilities of operative mortality (OM) made by clinicians with access to the predictive rule (Rule) and without access to the predictive rule (No Rule) are depicted in the top panel. Only clincians #6 and #9 reported that they used the rule when it was made available. Probability estimates for prolonged ICU stay greater than 48 hr (ICU>48h) are depicted in the bottom panel.

Figure 6.2



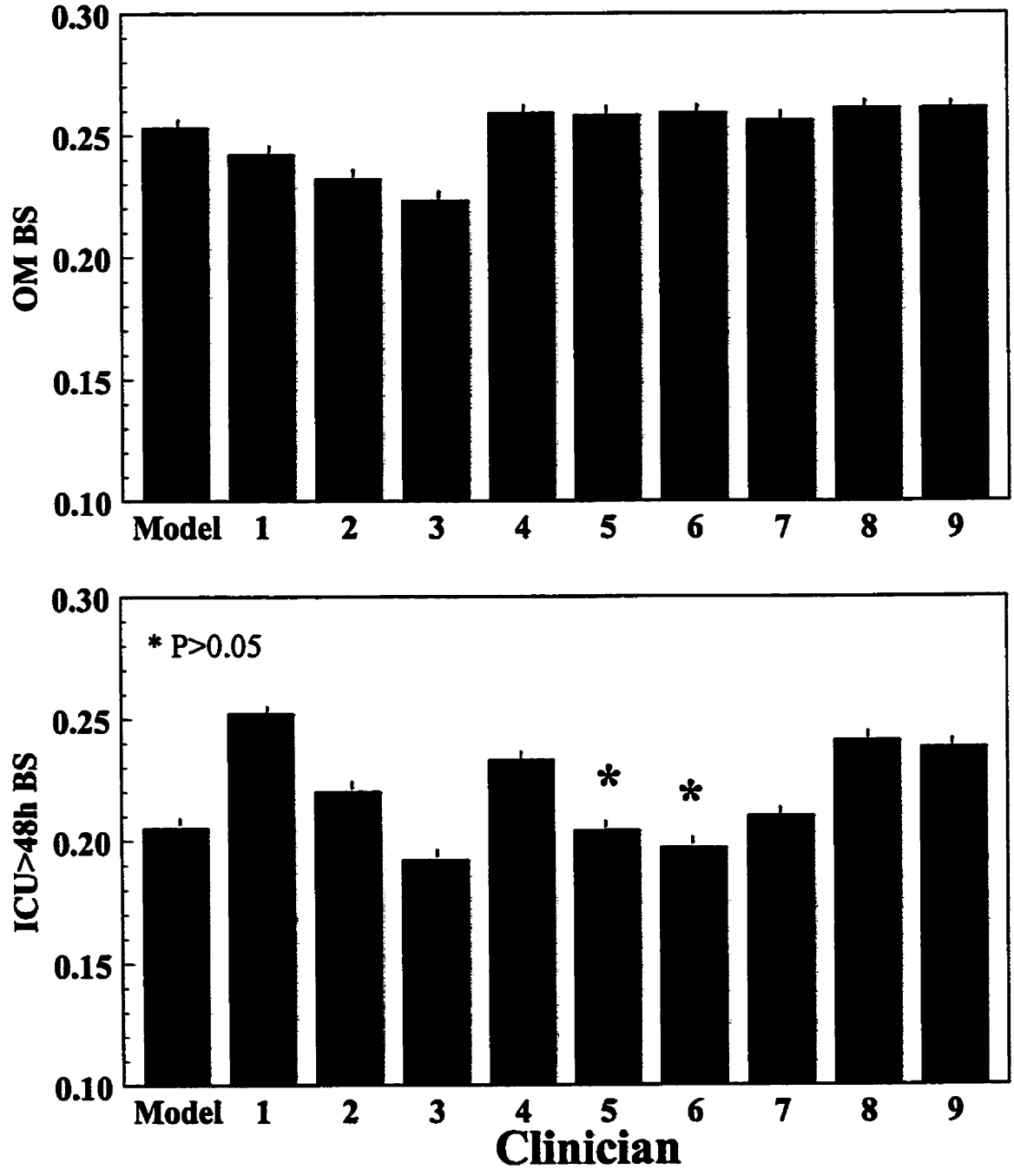
Legend 6.2: The delta of the clinicians minus the statistical model's estimates of the predicted probability of operative mortality (OM, top panel) and prolonged ICU length of stay greater than 48 hours (ICU>48h, bottom panel) is depicted. The senior clinicians (#1 - 5) significantly over-estimated the probability of OM compared to the statistical model. The junior clinicians' (#7-9) estimates were significantly different ($p<0.05$) from the seniors' estimates. There was no discernible pattern for ICU>48h relating to seniority.

Figure 6.3



Legend 6.3 The area under the Receiver-Operator Characteristic curve (ROC) for operative mortality (OM, top panel) and prolonged ICU length of stay greater than 48 hours (ICU>48h, bottom panel). With the exception of one senior surgeon, discrimination was poor for OM. Discrimination was slightly better for estimates of ICU>48h.

Figure 6.4



Legend 6.4 Brier Scores (BS) for operative mortality (OM, top panel) and prolonged ICU length of stay greater than 48 hours (ICU>48h, bottom panel) are depicted. All Brier Scores for OM were associated with P values <0.001, indicating significantly imprecise judgements. With the exception of Clinicians #5 and #6, all Brier Scores for ICU>48h were also associated with P values <0.001.

Chapter Seven

The Development and Application of a Predictive Rule to Evaluate the Long-Term Outcomes of Coronary Artery Bypass Surgery *

* Adapted (in part) from:

Ivanov J, Ralph-Edwards A, David TE, Yau TM and Naylor. The evaluation of long-term outcomes in patients following coronary artery bypass surgery using a clinical database linked to administrative data. Submitted to J Thorac Cardiovasc Surg; May 2000. Revision requested September 2000.

7.1 ABSTRACT

Purpose: To identify the contemporary, independent predictors of late survival and re-admissions for cardiac events following coronary bypass surgery (CABG); and to develop a predictive risk index for long-term survival.

Methods: A clinical database of 7022 patients undergoing isolated CABG between January 1/92 and December 31/96 was linked to the Canadian Institute of Health Information (CIHI) database and the Registered Persons Data Base to determine operative deaths (n=164), and late deaths: in-hospital (n=270) and out-of-hospital (n=137). Discharge diagnostic codes and procedure codes from the CIHI database were used to determine freedom from re-admissions for cardiac events which included recurrent angina, acute MI, congestive heart failure and repeat revascularization (PTCA or CABG). Risk ratios from the Cox regression analysis for survival were rounded to their nearest integer to provide a risk weight for each independent predictor of survival. Risk weights were summed to create a total risk score for each patient. Total risk scores were divided into quartiles to form relative risk groups. The predicted probability of survival to each time interval for each patient was calculated by a log-logistic, risk-adjusted model. Comparisons of predicted versus observed probabilities were made by linear regression.

Results: Data were 99.5% complete for all prognostic risk factors. Mean follow-up was 42 ± 19 months (range 0-78). Overall five year survival was 89±1%. The independent predictors of late survival were (risk ratio in parentheses): age group, defined as <65 years, 65-74 years and ≥75 years (1.7), female (1.2), LV grade (1.5), previous CABG (1.8), surgical priority (1.3), congestive heart failure (1.7), diabetes (1.3), peripheral vascular disease (1.5), hypertension (1.3) and renal failure (1.5). The overall five year freedom from re-admission for cardiac events was 79 ± 1%. The independent predictors of cardiac re-admission were: age group (1.1), female (1.4), LV

Grade (1.4), previous CABG (1.6), surgical priority (1.6), congestive heart failure (1.8), diabetes (1.3), peripheral vascular disease (1.3), hypertension (1.2), and renal failure (1.4). The comparisons of observed to predicted five year survival for each risk group and the R^2 for predicted versus observed survival probability for each patient at the time of death or censor are as follows: ***Low Risk*** (N=1854) OBS=96±1%, PRED=97±1%, R^2=0.52. ***Medium Risk*** (N=2391) OBS=92±1%, PRED=94±1%, R^2=0.61. ***High Risk*** (N=1788) OBS=86±1%, PRED=89±2%, R^2=0.58. ***Extreme Risk*** (N=954) OBS=71±2%, PRED=76±10%, R^2=0.28. All R^2 P values were <0.001.

Conclusions: The linkage of a clinical database with administrative data allowed for the efficient follow-up of a large contemporary cohort of CABG patients and the identification of multivariable predictors of survival and re-admission to hospital for cardiac events. Patient-related risk factors which were predictive of mortality also predicted re-admission to hospital for cardiac events. A predictive rule developed from the Cox regression risk ratios of survival permitted the characterization of patients into relative risk groups. Predicted long-term survival showed moderate precision with observed results for all but the extremely high risk group. This predictive algorithm may provide a valuable tool for medical decision making, internal benchmarking, trial design and patient counselling for the follow-up interval between one and five years. Prospective external validation is still required and should be the focus of future research. This study demonstrates the need to consider long-term as well as short-term outcomes to evaluate the benefit of CABG. Additionally, the true benefit of CABG cannot be evaluated by the examination of mortality alone, but should also include freedom from recurrent cardiac events.

7.2 INTRODUCTION

Since its inception in 1969,[198] coronary artery bypass graft surgery (CABG) has become the most intensely studied surgical procedure in the world. Over the past two decades we, and others, have documented a significant temporal increase in the prevalence of risk factors in patients presenting to CABG which has been accompanied by a paradoxical decrease in hospital mortality.[32,35,73,120] As a result of improved hospital processes of care and outcomes, there has also been a significant shift in the postoperative management of CABG patients, specifically in regards to shorter intensive care unit and hospital lengths of stay.[186,187]

The previous three chapters demonstrated that the surgical literature is replete with studies examining the association between various risk factors and hospital mortality following CABG. However, the effectiveness of CABG cannot be characterized solely by evaluations of hospital outcomes which reflect short-term risks of surgery. The net benefit of surgical therapy for ischemic heart disease must be evaluated by examining long-term survival and freedom from recurring symptoms.

There are many studies which have identified important multivariable risk factors and their impact on long-term outcomes in patients recruited as early as the 1970's.[199-210] Fewer studies have included more contemporary patients operated on after 1985.[211-216] There are no large observational studies which examine postoperative survival and freedom from cardiac events in CABG patients in Canada. Recent medical audits (institutional or surgeon-specific) which evaluate the safety and effectiveness of surgical therapy do not include evaluations of the contemporary, long-term outcomes of therapy for individual patients or those with specific combinations of risk factors.[217] Guidelines developed for the American Heart Association and American College of Cardiology recommended the use of time-related, patient-specific calculations of predicted survival derived from multivariable risk equations of long-term outcomes as a tool to counsel

patients regarding the risks and benefits of various therapies for the treatment of coronary artery disease.[218] Additionally, multivariable risk equations developed for long-term survival may provide a valuable adjunct to the evaluation of quality of care through the calculation of provider-specific, risk-adjusted survival. Predictive risk algorithms may also be valuable tools for the identification of high risk subgroups of patients who may benefit from specific therapies designed to reduce their risk of poor outcomes. Such algorithms may be useful both in practice and in the design of efficient randomized controlled trials.

Postoperative follow-up in large groups of patients by traditional methods is a slow, resource intensive and costly endeavor. In Ontario, Canada's largest province, all patients have a unique Health Card number which becomes part of every medical record. The Canadian Institute of Health Information (CIHI) contains the discharge International Classification of Disease 9th revision (ICD-9) codes on every patient admitted to a Canadian hospital. Since 1982, the Division of Cardiovascular Surgery at the Toronto General Hospital has maintained a prospective database on every patient undergoing cardiac surgery at our institution.[72,120,128] We have linked our clinical database with administrative data to follow all patients undergoing isolated CABG between Jan 1, 1992 and Dec 31, 1996 for survival and re-admission to hospital for cardiac events.

Therefore, the purpose of this study was:

1) To determine the contemporary, independent predictors of long-term survival and re-admissions for cardiac events;
2) To compare risk factors for operative mortality with those for long-term survival and re-admissions for cardiac events;
3) To develop a predictive risk index for long-term survival based on a simple set of key prognostic variables available to most cardiac surgeons at the time of initial patient consultation;

4) To use the predictive risk index as a tool to characterize patients into relative risk groups based on long-term outcomes for greater ease of clinical and research applicability.

7.3 METHODS

7.3.1 Data Sources

Clinical, operative and hospital outcome information for all 7022 consecutive patients undergoing isolated coronary artery bypass graft surgery (CABG) at the Toronto General Hospital (TGH) between Jan 1, 1992 and Dec 31, 1996 were entered prospectively into a computerized cardiovascular surgery clinical database. Permission was obtained from the Ontario Ministry of Health to use patient identifiers, the medical record number, the Ontario health card number and date of operation to link the clinical database to administrative databases: the Canadian Institute of Health Information (CIHI) database and Registered Persons Database (RPDB) of registered deaths. Data linkage and analysis were performed on-site at the Institute for Clinical Evaluative Sciences where the security and confidentiality of the data is protected. The date of entry into the study was defined as the date of operation. The closing date for administrative data was March 31, 1998.

7.3.2 Outcomes

Outcomes for this study were time to mortality from any cause and time to re-admission to hospital for cardiac related events following discharge for the index CABG operation.

Deaths were determined from three sources:

a) operative mortality, defined as any in-hospital death in the clinical registry during the index admission for CABG,

b) in-hospital deaths from the CIHI database were determined from the discharge exit code

c) out-of-hospital deaths were determined from the RPDB

Re-admissions to hospital for cardiac events occurring after the entry date of operation were determined from the ICD-9 codes for the most responsible diagnosis (DXCODE1) and the revised Canadian Classification of Diagnostic Procedure codes (CCP) for the first and second procedures (PRCODE1, PRCODE2). Cardiac events included acute myocardial infarction (DXCODE = 410), angina (DXCODE = 411, 413), congestive heart failure (DXCODE = 428.0 to 428.9), atrial fibrillation or flutter (DXCODES = 427.3), supraventricular tachyarrhythmia (DXCODE = 427.0 to 427.2), percutaneous transluminal angioplasty or stenting (PRCODE = 48.02 to 48.04) and repeat CABG (PRCODE = 48.11 to 48.90).[219,220] The first admission for each cardiac event was captured and used in the analysis, therefore, any one patient may have had more than one type of cardiac event analyzed. The failure time for freedom from any cardiac event was defined as the earliest admission date for any cardiac event.

The RPDB captured 95% of the operative deaths from the clinical registry (which is considered the "gold standard") for a kappa statistic for agreement above chance of 0.95 ± 0.01. Deaths of out-of-province patients would not have been captured by the RPDB. Approximately two dozen foreign patients per year undergo CABG at the Toronto General Hospital. Out-of-country or out-of-province deaths or hospital admissions would not have been captured by either the CIHI or RPDB databases.

The number of follow-up months was calculated by subtracting the date of operation from the closing date of the CIHI database. For patients who died, survival time was calculated by subtracting the date of operation from the death date. The number of months for freedom from re-admission to hospital for a cardiac event was calculated by subtracting the entry date of operation from the closing date for

administrative data. For patients who experienced a cardiac event, the time to the earliest admission for any one event was determined by calculating the interval between the re-admission date and the entry date of operation.

7.3.3 Analysis

General issues

SAS 6.12 for Windows [151] was used for statistical analysis. Data are presented as the means ± standard deviations in text and tables, and as means ● standard errors in figures, unless otherwise noted. Statistical significance was defined as a two-tailed P value <0.05.

Predictors of Operative Mortality

Using previously described methods, the independent predictors of operative mortality (OM) were determined by a stepwise logistic regression analysis.[72] All explanatory variables available in the database were submitted to logistic regression analyses using stepwise selection.[89,105] The variables submitted to all multivariable models included: age, gender, left ventricular (LV) grade based upon the ventriculographic ejection fraction (1=EF>60%, 2=EF 40-59%, 3=EF 20-39%, 4=EF <20%), New York Heart Association functional classification (NYHA), surgical priority (elective, same-hospitalization as a cardiac catheterization or cardiac event, emergency defined as <12 hours from catheterization or a cardiac event), acute coronary insufficiency defined as prolonged chest pain lasting greater than 15 min, recent myocardial infarction within the month prior to surgery, previous bypass surgery, congestive heart failure, diabetes, peripheral vascular disease which included carotid disease, a history of hypertension, renal failure defined as requiring either peritoneal or hemodialysis, hyperlipidemia, history of stroke or transient ischemic attacks, the extent

of coronary artery disease and a significant left main stenosis.

Age was entered as a categorical variable (1=age <65 yr, 2=age 65-74 yr, 3=age ≥75 yr) so that odds ratios for operative mortality for age could be compared to the risk ratios for long-term survival for the predictive model. Age was also evaluated as a continuous variable in a separate model to determine to what extent the odds ratios for other predictor variables changed between the two models.

Two way interactions were tested in all models for combinations of age or LV Grade or redo CABG with all other prognostic variables. Because of high collinearity (Pearson correlation coefficients >0.4) the variables "surgical priority", "acute coronary insufficiency" and "recent myocardial infarction" were ultimately tested individually in the model for survival. Incomplete revascularization was not tested in any model because there were only 16 patients who received fewer grafts than they had diseased vessels.

The best logistic regression model was determined by two diagnostic criteria: the Hosmer-Lemeshow goodness-of-fit statistic (HL)[105] which evaluated the precision between predicted and observed probabilities, and the area under the Receiver-Operator Characteristic (ROC) curve which evaluated model discrimination.[92,108,109]

Univariate Results for Survival and Cardiac Related Re-admissions

The cumulative probability of remaining free of an event for longer than time=t was calculated by Kaplan-Meier analysis for all prognostic variables. The univariate association between survival curves for each prognostic variable (i.e., male vs female, primary vs redo CABG etc.) was evaluated by the log-rank statistic.

Multivariable Cox Regression Models

All prognostic variables were submitted to Cox regression analyses using

stepwise selection to determine the independent, multivariable predictors of each long-term outcome. Because of criticisms from statistical reviewers regarding the potential loss of information when continuous variables are categorized, we used age group, LV grade, NYHA etc as continuous variables and derived one regression parameter for each variable. The regression coefficient would then be multiplied by each value of the prognostic variable and exponentiated to obtain the risk ratio. Model validity was evaluated by the chi square P-value for the last variable retained in the model.

Three different models of survival were constructed: (a) survival to one year, in which all patients were censored at one year if they survived past one year, (b) survival in all patients for the total follow-up period, and (c) survival in just those patients who survived their index hospital admission.

Relative Risk Group Construction

Risk weights for each level of an independent predictor variable were derived by rounding the risk ratios from the Cox regression analysis for survival to the nearest integer. A total risk score was calculated by summing the relevant risk weights which characterized each patient.[39,72,73] Because of small numbers, risk scores >14 were collapsed into a single score of 14. Relative risk groups were constructed based on quartiles of the total risk score: low risk (LR=total risk score 0-2), medium risk (MR=total risk score 3-5), high risk (HR=total risk score 6-8) and extremely high risk (ER=total risk score ≥9).

As a comparative method, the regression coefficients were also used to construct relative risk groups. Regression coefficients which characterized each patient were summed. Quartiles of this risk score were also used to define each relative risk group: LR total score 0-0.63 (N=1682, 24%), MR=total score 0.64-1.09 (N=1827, 26%), HR=total score 1.10-1.59 (N=1673, 24%), and ER=total score ≥1.6 (N=1805, 26%).

Predicted versus Observed Survival

A fully parametric, accelerated failure time model was used so that estimated survivor functions could be directly produced from the model and therefore goodness-of-fit could be checked more thoroughly. The intercept, scale parameter and regression coefficients from the log-logistic analysis were used to calculate the predicted probability of survival to time=t from the formula:

$$S(t|X) = 1 / 1 + [-g(\beta o + X)] * t^{g}$$

where S = survival probability, g = 1/scale parameter, β_0 = regression intercept, X = regression coefficients for covariates which describe each patient, t = follow-up time (months).[151]

Accuracy and Precision of the Survival Model

A single survival estimate for each patient was calculated from the log-logistic model and converted to a cumulative hazard (-log of predicted survival). The number of predicted deaths was determined by summing the cumulative hazards.[217] The percent of accurately predicted deaths was calculated by dividing the predicted number of deaths by the observed number of deaths within each risk group. Model precision was evaluated first by graphically superimposing the predicted probability (± 1 standard error) of survival as calculated by the log-logistic model over a graph of observed survival as calculated by the Kaplan-Meier analysis.[89,92,218] A weighted linear regression was performed comparing the predicted probability of survival versus observed probability of survival grouped by total risk score.[72,221] In addition, the difference (residual) between the mean predicted minus observed probabilities of survival was plotted to evaluate time-related departures from a residual that equalled zero.

A further evaluation of model fit was performed by a weighted linear regression comparing the annual mean predicted versus observed cumulative probability of survival

for each total risk score to evaluate to what degree the relationship was over- or under-estimated. A slope of 1 and intercept of 0 would indicate a perfect fit of predicted to observed outcomes.[89] Model fit was also evaluated by the residual chi square comparing expected (sum of cumulative hazards) to observed deaths within each risk group from the formula:

$$\chi^2 = (O-E)^2 / (E(1-E/N))$$

where χ^2 = chi square, O=observed number of deaths, E=expected number of deaths, N=total number in each group.

Calculation of Provider-Specific Risk-Adjusted Survival

We explored the possibility of using the predicted probabilities of survival derived from the multivariable model in an equation to calculate provider-specific, risk-adjusted outcomes. A predicted survival curve was generated for each patient. The mean predicted survival of each surgeon's patients was divided by the observed cumulative survival and then multiplied by the overall observed survival for all providers for the three year time interval. This calculation of risk-adjusted survival is interpreted as the survival for any one provider based upon the results he/she could anticipate if his/her case-mix was identical to everyone else's.[67,72]

7.4 RESULTS

7.4.1 Patient Sample

Between January 1, 1992 and December 31, 1996, 7022 consecutive patients underwent isolated coronary artery bypass graft surgery (CABG) at the Toronto General Hospital. Data were complete for all but 35 patients who were missing one or more important elements of prognostic information (99.5% complete). Operative mortality was 2.3% (n=164). Patient demographics, risk factors and outcomes are presented in

Table 7.1.

Linking the clinical database with the CIHI database resulted in a total of 18,074 records, of which 11,156 records represented pre-CABG admissions. The remaining 6,918 records represented post-CABG admissions in 3,629 patients. Non-cardiac admissions in this group accounted for 5,026 records. Total post-CABG events therefore accounted for 1,892 total admissions in 1,036 patients.

7.4.2 Predictors of Operative Mortality by Logistic Regression

The independent, multivariable predictors of operative mortality are contained in Table 7.2. The area under the ROC curve was 0.772 and the Hosmer-Lemeshow goodness-of-fit P value was 0.16. When age was tested in a separate model as a continuous variable, the odds ratios for the remaining predictor variables did not change more than ± 10%. In addition, neither model accuracy nor precision was affected by the use of age as either a categorical or continuous variable.

7.4.3 Survival and Freedom From Re-admission for Cardiac Events

Survival

Mean follow up was 42 ± 19 months (range 0 to 78 months). The minimum follow-up for survivors was 15 months. There was a total of 407 late deaths following discharge from the index operation: 270 in-hospital deaths were determined from CIHI data while 137 out-of-hospital deaths were identified from the RPDB data. Overall survival to one and five years was 96 ± 0.4% and 89 ± 1% (Figure 7.1). The risk ratios (RR) and the 95% confidence intervals (CI) for the independent, multivariable predictors of survival to one year, total survival, and survival for the hospital survivors only are found in Table 7.2. Univariate results from Kaplan-Meier analyses for survival are presented in Table 7.3.

Re-admissions for Cardiac Events

There were 1036 (15%) re-admissions for cardiac events over the complete follow-up period. Overall actuarial freedom from re-admission for any cardiac event at one and five years was 94 ± 0.3% and 79 ± 1%. The independent, multivariable predictors of re-admission for cardiac events are contained in Table 7.4. The Kaplan-Meier estimates are contained in Table 7.5. All univariate comparisons for freedom from re-admissions for cardiac events were significant by the log-rank test, P<0.001.

One and five year freedom from re-admission for recurrent chest pain, which included acute myocardial infarction or angina was 97 ± 0.2% and 87 ± 1% (n=608, 8.7%).

Freedom from re-admission at one and five years for congestive heart failure (n=314, 4.5%) was 98 ± 0.2% and 94 ± 0.2%.

There were 27 repeat CABG procedures and 108 percutaneous angioplasty/stents over the follow-up period. The one and five year freedom from revascularization, which included both PTCA and/or CABG (n=135, 1.9%), was 99 ± 0.1% and 97 ● 0.2%.

The one and five year event-free survival (n=1460, 21%) was, 92 ● 0.3% and 72 ● 1% respectively. The independent, multivariable predictors of event-free survival were (risk ratios (RR) and 95% confidence intervals (CI) in parentheses): age group (RR 1.252, CI 1.16 - 1.35), female gender (RR 1.316, CI 1.17 - 1.48), LV Grade (RR 1.374, CI 1.28 - 1.47), NYHA (RR 1.227, CI 1.12 - 1.35), previous CABG (RR 1.609, CI 1.36 - 1.91), preoperative congestive heart failure (RR 1.748, CI 1.48 - 2.06), surgical priority (RR 1.35, CI 1.20 - 1.52), diabetes (RR 1.262, CI 1.13, 1.42), hypertension (RR 1.205, CI 1.08 - 1.34), peripheral vascular disease (RR 1.340, CI 1.17 - 1.53), renal failure (RR 1.453, CI 1.19 - 1.77) and myocardial infarction in the month prior to surgery (RR 0.837, CI 0.73 - 0.96). The P value for the last variable included in the model was 0.013.

7.4.4 Relative Risk Group Construction

The risk ratios for each predictor variable from the Cox regression analysis for survival were rounded to their nearest integer to form a risk weight. (Table 7.6). The total risk score was calculated by summing the risk weights which characterized each patient. Univariate as well as frequency analysis were used to establish the cutpoints of total risk score which defined the relative risk groups: low (LR, score 0-2, N=1854, 27%), medium (MR, score 3-5, N=2391, 34%), high (HR, score 6-8, N=1788, 26%) and extremely high (ER, score ≥9, N=954, 14%). Patient characteristics and outcomes for the relative risk groups are presented in Table 7.7.

A comparison of patients in each risk group constructed from the rounded risk ratios versus the Cox regression coefficients showed agreement above chance was 0.88 ± 0.01. We determined that it would be easier for clinicians to use a simple predictive rule which sums rounded risk ratios to form a total risk score rather than summing regression coefficients and have therefore relied upon the former method.

7.4.5 Predicted versus Observed Survival

Regression coefficients for the log-logistic accelerated failure time model are described in Table 7.8. The yearly predicted probabilities of survival for each risk score are found in Table 7.9. The mean predicted probability of survival was graphically superimposed over observed survival for the total sample of patients (Figure 7.1) and for each relative risk group (Figure 7.2).

Both observed five year survival from the Kaplan-Meier analysis and risk-adjusted predicted survival from the log-logistic accelerated time failure model are shown in Table 7.10. The R^2 from the linear regression analysis of predicted versus observed survival probabilities for the total dataset of all patients was only 0.04; however, due to the large sample size, this regression was statistically significant

(P<0.001). The R^2's for each risk subgroup were all significant at the P<0.001 level. The R^2 for the ER group was only half that of the other three groups. However, the R^2 for the ER group which was constructed from cutpoints of the sum of regression coefficients was only 0.14, suggesting that the use of rounded risk ratios was a more valid approach.

The departure between predicted and observed survival seen at five and six years in the ER group may have been a function of reduced sample size at those time intervals or, more probably, the onset of saphenous vein graft failure which has serious consequences in higher risk patients who are further compromised by poor ventricular function. Figure 7.3 depicts the residual difference between predicted minus observed results for the overall patient sample and for each risk subgroup. There is a clear over-estimation of actual survival for the operative and latter time intervals. Figure 7.4 depicts the calibration curve of the mean predicted versus observed estimates of survival for each risk score at each yearly interval. The weighted linear regression R^2 of 0.86 demonstrated that there was an excellent fit between the model's estimates and actual outcomes when averaged for each risk score.

The predicted number of deaths within each risk group was calculated by summing the cumulative hazard ratios from the log-logistic model. Chi square P values for each risk subgroup indicated no statistical difference between observed and predicted results (Table 7.10). When the sample of patients was evaluated in total, summing the cumulative hazard ratios resulted in 508 predicted deaths for 89% accuracy. However, due to the larger sample size, there was a significant difference between observed and predicted deaths for the whole group (P<0.001).

To further validate the appropriateness of risk group stratification, we examined the univariate association between risk groups and survival (Kaplan-Meier analysis, log-rank P=0.0001) as well as freedom from admissions for chest pain, congestive heart failure, any cardiac event and overall event-free survival (Figure 7.5). There was no

significant, univariate difference between risk groups for revascularization. At six years the freedom from revascularization was (95% confidence intervals): LR 97% (96, 98%); MR 97% (96, 98%); HR 97% (96, 98%); ER (98% (97, 99%), P=0.8.

7.4.6 Provider-Specific, Risk-Adjusted Survival

The results depicted in Figure 7.6 are a preliminary evaluation of the use of a multivariable model of survival in the calculation of risk-adjusted, provider-specific survival. We suggest a cautious interpretation of these results since the providers in this context are surgeons and in all fairness, have limited impact on the long-term management of these patients. However, Figure 7.6 does suggest that it is possible to separate out provider effects. Note the very different results between the upper most and lowest dashed lines. These two providers had identical predicted probabilities of three year survival based on their case-mix but different observed and risk-adjusted outcomes.

7.5 DISCUSSION

The clinical database of 7022 consecutive patients undergoing CABG between 1992 and 1996 at the Toronto General Hospital was linked to provincial and national hospital administrative data as well as to a national data register of deceased persons. This data linkage permitted us to follow our patients for survival and re-admissions to hospital following their index operation. Two other studies by Hartz[45] and Boscarino[76] also used linked clinical and administrative data to evaluate survival data following CABG. The study by Hartz and colleagues was conducted in a large population of Medicare patients for the sole evaluation of late survival. Boscarino et al evaluated late survival in a relatively small cohort of 771 patients following CABG. Thus we present a unique Canadian study which evaluates not only survival in a large population of CABG patients but also their freedom from re-admissions for cardiac events. Also, a predictive

rule was developed to predict late survival following CABG.

The richly characterized clinical data in this study were used for the development of a multivariable statistical model to characterize survival up to six years following surgery. To achieve clinical applicability and relevance to decision making in and around the time of surgery, the variables in the models were those usually available to cardiac surgeons at the time of initial surgical referral.[70] Additionally, the independent, multivariable risk factors for re-admission to hospital for cardiac events were also identified by Cox regression analysis. The results demonstrated that a predictive rule can be used to construct valid relative risk groups for long-term outcomes after CABG and suggest that outcomes can be predicted reliably for up to five years postoperatively in all but the most extremely high risk patients. However, external validation is still required.

7.5.1 Study Strengths

This observational study was conducted in a large cohort of consecutive patients undergoing isolated CABG at one institution over a relatively short recruitment period of five years. The results of this study are therefore generalizable to institutions with similar case-mix and outcomes as the Toronto General Hospital.

The diversity of patient characteristics and data linkage with administrative data extended our ability to capture a wide range of important clinical risk factors as well as outcomes. The framework of this study therefore integrates both classical principles of measurement with statistical methods to model the impact of those risk factors on survival and cardiac-related re-admissions to hospital. Follow-up data were collected by the same method for all patients and therefore reduced outcome ascertainment bias. This study highlights the value of using complementary data sources for clinical research.[45]

7.5.2 Limitations

The most important limitation of this study is the lack of information on long term interventions such as risk factor modification, use of anti-platelet drugs, lipid lowering medications, beta blockers etc. Variations in these long term factors might be important confounders that, if included in a model, would alter the risk weights of certain variables or eliminate their impact altogether. The most optimistic theory would be that the results in this study are unbiased by these factors if they are similar across surgeons and risk groups.

Mis-classification of clinical and outcome variables could bias outcomes. Coding of admissions for myocardial infarction has been addressed in previous studies conducted by the Institute for Clinical Evaluative Sciences.[219] Coding accuracy in previous studies demonstrated a sensitivity of 95% and specificity of 88% for the most responsible diagnosis of acute MI. To mitigate potential errors (e.g. unstable angina diagnosed as an acute MI), we choose to create a composite outcome of recurrent ischemic pain which included both acute MI or angina. Mis-classification in this study would most likely contribute to noise and bias the results towards the null hypothesis for subgroup comparisons.

We did not capture out-of-country deaths or re-admissions to hospital. There are approximately two dozen foreign patients each year who undergo CABG at the Toronto General Hospital. Additionally, Ontario residents may be admitted to hospital for cardiac events while on vacation in other provinces or countries. These factors may have contributed to a slight under-estimation of outcomes but in general, should not have biased the subgroup comparisons.

The predictive rule appeared to be internally valid. However, external validation is still required. The standard methods of index development usually involve creating a derivation and validation dataset using some method of sample splitting, e.g.,

bootstrapping. The creation of derivation-validation sub-sets of patients is a valid exercise for much larger datasets but in the case of only 7000 patients it would have resulted in under-powered observations for the highest risk group. Therefore, our strategy was to use our entire dataset to identify all the potential predictors and their associated risk weights, validate the rule internally, using methods previously described,[90,217] and plan for prospective, external validation.

The first re-admission to hospital for a cardiac event was evaluated and not the first occurrence of symptoms which did not require hospitalization. To this end, due to the nature of our follow-up methods, we may have under-estimated the actual recurrence of symptoms related to ischemic heart disease, although our report shows similar recurrence rates to other studies using more traditional methods of following patients.

7.5.3 Survival

The observed five year survival of 89 ± 1% for all patients was similar to several other published reports despite the fact that patients in those studies were recruited as early as the 1970's.[199-210,216,218] There has been a nation wide increase in the prevalence of risk factors in patients presenting to CABG which has, paradoxically, been accompanied by a significant reduction in hospital mortality. The fact that our observed five year survival is similar to earlier reports suggests that overall risk-adjusted, long-term survival has also improved in parallel with hospital outcomes, possibly due to continued refinements in operative and anaesthetic techniques as well as postoperative management occurring during the interval of follow-up.[199] The predictors of mortality in those patients who survived the immediate postoperative period were different from the predictors in the total patient sample (Table 7.2). These results suggest that the impact of gender and redo bypass on survival occurs in the early postoperative period only.

Survival is the most unbiased event and therefore the risk ratios from the Cox

regression for survival were used to determine the risk weights for the predictive rule. The purpose of using risk ratios was simply to separate patients into clinically relevant risk groups and not to calculate the actual probability of events. Risk group stratification by quartiles of the total regression coefficient score had excellent agreement with the method used based upon the risk scores. However, the risk scores derived from summing regression coefficients resulted in a more complex computational method which would not have been as useful for clinicians. Coste and colleagues[90] suggested that every effort be made to simplify scoring systems and present them in an attractive format to ensure their applicability in diverse settings.

A prognostic rule based simply on coronary anatomy developed at Duke University [199] showed a clear relationship between survival and the extent of coronary artery disease. They found that adjusting survival in cohort studies by baseline characteristics resulted in estimates similar to those found in randomized controlled trials. Although the extent of coronary artery disease was not a significant predictor of survival in this study, the weighting of other clinical characteristics to construct relative risk groups was significantly associated with long-term survival.

Califf and colleagues concluded that five year survival rates were a function of operative risk.[201] Our findings are similar. In this study, the risk factors and indeed most of the risk weights for the predictive rule developed from the Cox model for long-term survival were very similar to the logistic regression-based predictive rule developed for operative mortality in Chapter Four. It is possible that the original rule developed for operative mortality may have been equally effective at stratifying patients into low-high risk groups for long-term outcomes.

But, are five year survival rates primarily a function of operative risk or are other, unmeasured forces exerting their influence over late survival following CABG? The shape of the observed survival curves in our high and extremely high risk patients

showed a "shoulder" of decreasing survival beginning at approximately four years postoperatively. The shape of our survival curves was very similar to those seen in the study from Duke[202] in patients with poor LV ejection fraction (their Figure 3). In the Duke study's lower risk, surgical cohort there is a similar "shoulder" seen in observed survival at seven years (their Figure 5). The VA randomized study in low risk patients also reported a gradual loss of vein graft patency between five and 10 years.[210] In our study, it is possible that we are seeing at four years, the onset of vein graft failure in the higher risk groups: a harbinger of events that are delayed in lower risk patients with good LV function, but having serious consequences in patients with poor ventricular function and little reserve.

Sergeant and Blackstone[217] also evaluated the use of a predictive risk algorithm developed from the ACC/AHA guidelines to predict the probability of survival after CABG.[218] The rule was recalibrated in the Belgium dataset and modified slightly by the addition of other variables. They concluded that the poorer precision between predicted and observed survival seen in higher risk patients was due to unmeasured risk factors which, because of their low prevalence in the database, were not included in the predictive rule. That may be a valid conclusion to some extent, but one cannot ignore the postoperative course in compromised patients, specifically in regards to graft failure, left ventricular reserve, the modification of risk-factors (e.g., cessation of smoking, hypertension control, diabetic control), and compliance with medical therapy (e.g., aspirin and lipid lowering medication).

7.5.4 Re-admissions for Cardiac Events

The recurrence of symptoms of ischemia and congestive heart failure not requiring admission to hospital may have resulted in an under-estimation of the return of symptoms related to ischemic heart disease. However, our five year freedom from

angina was 87 ± 1%, which was very similar to the ACC/AHA Task Force Report five year rate of 83%.[218] A myocardial infarction in the month prior to surgery and left main disease were associated with a risk reduction for postoperative chest pain recurrence, possibly because the area at risk was already infarcted prior to surgery.

Our five year revascularization rate was identical to that reported by others.[205,209,216] These previous studies also found that older age, poor ventricular function and triple vessel disease were associated with a lower risk of revascularization. These results highlight a need to revise the concept of increased risk of revascularization compared to repeat cardiac events when considering these risk factors since the results go in the opposite direction. Sergeant[216] identified a very short early hazard phase for revascularization followed by a longer, late phase of increasing hazard. They note that revascularization is an event which captures the practice of reintervention and not the actual need. It is likely that our results reflect the fact that patients with poorer LV function, more comorbidity or increased age are not offered repeat procedures because they are deemed to be at extraordinarily high operative risk.

Our event-free five year survival rate of 72% was close to the 78% reported by Sergeant[209] despite the difference in follow-up methods between the two studies. Higher risk patients had significantly decreased event-free survival compared to the other cohorts. However, Califf et al[201] and others[199,209] have suggested that high risk patients have more of a survival benefit from CABG than low risk patients. Califf et al were specifically concerned that pressure from providers to select low risk patients would lead to a significant misallocation of funds and result in lower CABG rates in the patients who would benefit most, those at higher risk of poor outcomes. They suggested that patient selection for CABG should be contingent on long-term benefits as well as operative risk. The use of a predictive rule may help to identify high risk patients who would benefit most from surgical therapy as well as helping providers plan for the appropriate

allocation of resources and postoperative patient-management strategies.

7.5.5 Model Development

Our plan was to develop a predictive rule which would encompass the entire follow-up period from the date of operation to the end of the follow-up interval, in this case, 78 months. To maintain simplicity of use, the predictive rule would therefore be required to model hospital outcomes as well as later events. There have been at least two[222] and perhaps three [217,218] distinct hazard phases identified for survival following CABG. The early phase was defined as the early postoperative period (up to one year) which was associated with an increased hazard of dying. In the following phase there was a decreased, but sustained hazard of dying. After five years, there is a gradual increase in the hazard of dying as well as an increased hazard of experiencing cardiac events.[218] The Cox model was used to calculate an adjusted risk ratio for each variable and was therefore used to determine the relative risk weights for each independent predictor variable. The Cox model is semi-parametric and does not have an intercept or shaping parameter. Therefore, the log-logistic model, which is fully parametric with a calculated intercept and shaping parameter, was used to calculate risk-adjusted, expected survival. The log-logistic shape parameter is specifically designed to capture the early postoperative phase of increased hazard and the subsequent constant hazard phase. However, it is apparent from our results that the very late hazard phase is not precisely captured by this equation (Figure 7.3) and there is good argument for a similar lack of precision for the earliest hazard phase in the extremely high risk group. This equation is therefore an appropriate statistical tool to model expected survival following CABG for one to four years or until the late hazard begins to increase. Because of the early onset of increased hazard in the high and extremely high risk groups, clinicians who are counselling high risk patients may wish to tailor their estimates of survival probability

towards the lower confidence intervals in Table 7.9.

This model contained most of the risk factors identified by the Working Panel Group on the Cooperative CABG Database Project as being important in the development and utilization of a predictive rule.[70] Over-estimation of survival in our highest risk groups was similar to results seen by Sergeant.[209] It is likely that some risk factors have a major impact on outcomes but their low prevalence in the database does not permit them to be statistically modeled.[209] Excess mortality may be related to unmeasured comorbidity, the onset of early graft failure, or differences in surgeon-specific, risk-adjusted outcomes that were not accounted for in the model. Sergeant noted that the changing pattern of risk factors may result in the absence of strong predictors in the model. In a recent study from Australia in which patients were recruited between 1985 and 1995,[215] many of the same risk factors associated with decreased survival were identified; however, notable by their absence were the extent of coronary artery disease, surgical priority, hypertension and female gender. We have observed that previously important risk factors such as left main disease and the number of diseased coronary vessels are no longer contemporary independent predictors of mortality.[72,128] Perhaps the previous identification of these important conditions or the increasing wide-spread use of arterial conduits has led to a reduction in the mortality associated with these variables. Indeed, the odds ratios for operative mortality for female gender and previous CABG were greater than the upper 95% confidence intervals for the risk ratios for long-term survival and re-admissions for cardiac events. These results suggest that the influence of these two variables on long-term outcomes may well be concentrated in the early postoperative period with a reduced hazard after the first year. The advent of new surgical technologies such as multiple arterial conduits, off-pump bypass and robotic surgery, combined with improved postoperative management of patients following CABG, will require that we rethink the impact of traditional risk factors in the next

millennium.

Other reasons for a lack of model membership could be due to high collinearity between clinical variables which measure similar domains, such as anginal severity, recent MI, coronary anatomy and surgical priority. In our database, there was a significantly high correlation (Pearson correlation coefficient in parentheses) between surgical priority and recent MI (0.5), NYHA (0.6), and anginal severity (0.6) In Ontario, patients waiting for surgery are managed by a triage system which assigns priority based on several factors, with the greatest weight given to coronary anatomy and anginal instability.[10-13,223] Since the inception of this system in April 1991, we have found that "surgical priority" has become a more robust predictor of hospital outcomes and, because of the triage system, is less susceptible to "up-coding". In our models, the inclusion of "surgical priority" was associated with improved accuracy and precision as compared to its component variables. Another explanation for lack of model membership is that a variable which has a significant univariate association with the outcome, may not contribute additional predictive value to the model over and above the variables already in the equation.

Our predicted versus observed survival R^2 of 0.04 for the entire dataset demonstrated a rather weak relationship despite being statistically significant. When regressions were examined for the risk subgroups, there was a much improved fit between observed and predicted values as evidenced by the R^2 in each panel of Figure 7.2. It is possible that evaluating all patients in one model violated an assumption of linearity which resulted in poorer fit. Using linear regression for the risk subgroups would have, in effect, examined shorter segments of the total calibration curve and therefore resulted in an improved fit. Transformation of the values on the y-axis did not improve precision.

7.5.6 Implications for Patient Counselling

Can outcomes be predicted reliably past four or five years postoperatively? Several researchers have felt that there is an important relationship between preoperative risk factors and long-term survival.[45] However, worsening disease states combined with the onset of graft failure are not included in prognostic models developed from preoperative risk factors. The confidence intervals around estimates of predicted survival were quite broad for those patients having higher risk scores (Table 7.9). Weintraub and colleagues[205] were particularly concerned that the changing pattern of CABG surgery, specifically in regards to the use of mammary artery grafts, aspirin treatment, risk factor modification and the increasing prevalence of high risk patients, would make it difficult to predict very late outcomes. Therefore, when counselling patients regarding risks and net benefits, surgeons must consider not only temporal changes in the prevalence of risk factors and improvements in outcomes but also the patients' contribution to their own well-being. Risk factor modification, (e.g., cessation of smoking, control of hypertension, diabetic control and compliance with anti-platelet and lipid lowering medications) is not captured by a predictive risk index; however, this domain probably has significant impact on long-term survival.

7.5.7 Quality Improvement

Clinical outcome is influenced by patient characteristics but also institutional and provider effects.[79] If we understand the forces, we can identify specific characteristics which may be amenable to strategies which will reduce a patient's risk of poor outcomes. We present this preliminary analysis of risk-adjusted three year survival as one potential tool for the evaluation of provider-specific outcomes and quality of care on a different time-horizon than is the case for most existing models. However, the validity of calculating risk-adjusted, long-term survival requires more intensive evaluation than

given in this report.

7.5.8 Conclusions

The linkage of a clinical database with administrative data allowed for the efficient follow-up of a large contemporary cohort of consecutive CABG patients and the identification of multivariable predictors of survival and re-admission to hospital for cardiac events. Patient-related risk factors which were predictive of mortality also predicted re-admission to hospital for cardiac events. Actuarial estimates from registry and administrative data were similar to estimates made from more traditional follow-up methods in previously published studies.

Survival estimates in this contemporary sample of patients suggest that there may have been a similar shift in the improvement of long-term outcomes which parallels temporal improvements seen with hospital survival. However, statistical models based on preoperative risk factors do not capture information on long term factors which may have acted as important confounders of outcomes.

Patients were characterized by relative risk groups based on Cox regression risk ratios for important prognostic variables. An accelerated time failure model using the log-logistic distribution was used to generate a predicted survival curve for each patient. After stratifying by risk group, the model predicting long-term survival demonstrated a good correlation between predicted and observed outcomes for the majority of patients. However, the results in higher risk patients suggested that the predictive rule was not as precise for estimates made after four to five years. Future research will be directed towards improving the precision of the predictive model by adding in variables which describe process of care or early outcomes.

This predictive algorithm may provide a valuable tool for medical decision making, internal benchmarking, trial design and patient counselling. Prospective external validation is still required and should be the focus of future research.

Table 7.1 **Demographics, Risk Factors and Outcomes for 7,022 Consecutive Patients Undergoing Coronary Artery Bypass Surgery Between 1992 and 1996 at The Toronto General Hospital**

Variable	Total for All Patients N (%)
N	7022
Age: Mean ± SD	62 ± 10
<65 year	3916 (56)
65-74 yr	2453 (35)
≥75 yr	653 (9.3)
Female	1482 (21)
LV Grade: 1	2026 (29)
2	3202 (46)
3	1565 (22)
4	220 (3.1)
Priority: Elective	3780 (54)
Semi-Urgent	3019 (43)
Emergency	221 (3.2)
Previous CABG	485 (6.9)
Preop CHF	524 (7.5)
NYHA: 1	156 (2.2)
2	936 (13)
3	2458 (35)
4	3472 (49)
ACI	2217 (32)
Recent MI	1278 (18)
TVD	5247 (75)
Left main disease	1340 (19)
Diabetes	1726 (25)
PVD	1003 (14)
Hypertension	3563 (51)
Renal Failure	290 (4.1)
Preop Stroke/TIA	585 (8.4)

Variable	Total for All Patients N (%)
Operative Mortality	164 (2.3)
Operative LOS	546 (7.8)
Operative Stroke	90 (1.3)
Total Deaths (6 yrs)	571 (8.1)
AMI	161 (2.3)
Chest Pain (AMI/UA)	600 (8.6)
CHF	314 (4.5)
PTCA	112 (1.6)
CABG	25 (0.4)
Any Cardiac Event	1036 (15)
Mortality/Morbidity	1460 (21)

Legend for Table 7.1 CHF=congestive heart failure, NYHA=New York Heart Classification, ACI=acute coronary insufficiency, MI=myocardial infarction, TVD=triple vessel disease, PVD=peripheral vascular disease, TIA=transient ischemic attack. Semi-urgent priority was defined as surgery occurring during the same admission as a cardiac catheterization or cardiac event. Emergency priority was surgery occurring within 12 hours of a cardiac catheterization or acute cardiac event. LOS=Low cardiac output syndrome, AMI=acute myocardial infarction, CHF=congestive heart failure, UA=unstable angina, PTCA=percutaneous transluminal coronary angioplasty.

Table 7.2 **Multivariable Models for Survival: Operative Mortality (OM), Survival to One Year with Censoring of Patients Surviving Past One Year, Total Survival in the Entire Study Sample, and Late Survival in Patients Who Survived Their Index Hospital Admission**

Variable	SLR OM	1 Year Survival	Total Survival	Hospital Survivors
	OR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)
Age group	1.704 (1.4, 2.1)	1.669 (1.4, 2.0)	1.684 (1.5, 1.9)	1.675 (1.5, 1.9)
Female	1.940 (1.4, 2.7)	1.651 (1.3, 2.2)	1.216 (1.0, 1.5)	...
LV Grade	1.846 (1.5, 2.3)	1.697 (1.4, 2.0)	1.473 (1.3, 1.6)	1.368 (1.2, 1.6)
Redo CABG	3.528 (2.3, 5.3)	2.865 (2.1, 4.0)	1.838 (1.4, 2.4)	...
Surgical Priority	1.397 (1.1, 1.8)	1.496 (1.2, 1.9)	1.285 (1.1, 1.5)	1.254 (1.0, 1.5)
CHF	1.823 (1.2, 2.8)	1.674 (1.2, 2.3)	1.697 (1.3, 2.2)	1.590 (1.2, 2.2)
Diabetes	1.453 (1.0, 2.0)	1.394 (1.1, 1.8)	1.340 (1.1, 1.6)	1.313 (1.1, 1.6)
PVD	1.572 (1.1, 2.3)	1.469 (1.1, 2.0)	1.530 (1.2, 1.9)	1.446 (1.1, 1.9)
Hypertension	...	1.333 (1.0, 1.7)	1.271 (1.1, 1.5)	1.241 (1.0, 1.5)
Renal Failure	...	...	1.532 (1.1, 2.0)	1.691 (1.2, 2.4)
L Main disease	...	1.378 (1.0, 1.8)	...	...
Preop CVA/TIA	...	...	...	1.422 (1.1, 1.9)
NYHA	...	...	...	...
Recent MI	...	...	...	...
TVD	...	...	...	...
P value for last variable in the model	0.032	0.032	0.004	0.039

Legend for Table 7.2 SLR=stepwise logistic regression, OM=operative mortality, OR=odds ratio, RR=risk ratio, 95%CI=confidence interval, Age group: 1=age <65, 2=age 65-74, 3=age ≥75; LV Grade: 1-4, Surgical Priority: 0=Elective, 1=Same hospitalization, 2=Emergency, x=variable was not significant in the

model. CHF=congestive heart failure, DM=diabetes, PVD=peripheral vascular disease, BP=hypertension, RF=renal failure. Variables submitted but not significant in the models: NYHA, left main disease, triple vessel disease, history of stroke/TIA, recent preoperative MI (highly correlated with surgical priority).

Table 7.3 Kaplan-Meier Estimates of Survival

Variable	1 Month	1 Year	5 Years
Age: <65	98.6 ± 0.2	97.8 ± 0.2	93.4 ± 0.5
65-74	97.2 ± 0.3	95.1 ± 0.4	85.4 ● 1.0
≥75	94.6 ● 0.9	91.9 ● 1.0	77.3 ± 3.0
Sex: Male	98.2 ± 0.2	97.0 ± 0.2	90.0 ± 0.5
Female	96.0 ± 0.5	94.0 ± 0.6	86.2 ± 1.2
LV Grade: 1	99.0 ± 0.2	98.3 ± 0.3	92.7 ● 0.8
2	98.1 ± 0.2	96.8 ± 0.3	90.7 ● 0.7
3	96.4 ± 0.5	94.6 ± 0.6	84.6 ● 1.2
4	91.4 ± 1.9	85.0 ± 2.4	69.5 ● 4.2
Previous CABG: N	98.1 ± 0.2	96.8 ± 0.2	89.7 ± 0.5
Y	93.2 ± 1.1	91.1 ± 1.3	84.4 ± 2.0
Surgical Priority: Elective	98.3 ± 0.2	97.6 ± 0.3	91.5 ± 0.6
Same Hospitalization	97.4 ± 0.3	95.5 ± 0.4	87.5 ● 0.8
Emergency	91.4 ± 1.9	87.3 ± 2.2	74.0 ● 4.0
Congestive Heart Failure: N	98.2 ± 0.2	97.0 ± 0.2	90.6 ± 0.5
Y	92.6 ± 1.2	88.2 ± 1.4	64.4 ± 5.7
Diabetes: N	98.0 ± 0.2	96.9 ± 0.2	90.6 ± 0.5
Y	97.2 ± 0.4	94.8 ± 0.5	85.5 ± 1.2
Hypertension: N	98.3 ± 0.2	97.3 ± 0.3	91.4 ± 0.7
Y	97.3 ± 0.3	95.5 ± 0.3	87.4 ± 0.7
Peripheral Vascular Disease: N	98.1 ± 0.2	96.9 ± 0.2	90.8 ± 0.5
Y	95.7 ± 0.6	92.9 ± 0.8	80.0 ± 1.8
Renal Failure: N	97.9 ± 0.2	96.7 ± 0.2	90.1 ± 0.5
Y	95.5 ± 1.2	89.4 ± 1.8	69.2 ± 4.0

Legend for Table 7.3 All comparisons across risk factors are significant by Kaplan-Meier analysis, log-rank $P<0.001$.

Table 7.4 **Multivariable Cox Models for Time to Re-admissions for Cardiac Events**

Variable	Recurrent Ischemic Pain	Congestive Heart Failure	PTCA or CABG	Any Cardiac Event
Age group	...	1.456 (1.2, 1.7)	0.657 (0.5, 0.9)	1.112 (1.0, 1.2)
Female	1.481 (1.2, 1.8)	1.553 (1.2, 2.0)	...	1.424 (1.2, 1.6)
LV Grade	...	2.291 (2.0, 2.7)	0.770 (0.6, 1.0)	1.353 (1.2, 1.5)
Redo CABG	1.536 (1.2, 2.0)	1.550 (1.1, 2.2)	...	1.576 (1.3, 1.9)
Surgical Priority	1.391 (1.2, 1.7)	...	1.838 (1.4, 2.4)	1.585 (1.4, 1.8)
CHF	1.505 (1.1, 2.0)	2.169 (1.6, 2.9)	...	1.828 (1.5, 2.2)
Diabetes	...	1.698 (1.4, 2.1)	...	1.255 (1.1, 1.4)
PVD	1.268 (1.0, 1.6)	1.413 (1.1, 1.9)	...	1.318 (1.1, 1.6)
Hypertension	...	1.331 (1.1, 1.7)	...	1.169 (1.0, 1.3)
Renal Failure	...	2.020 (1.4, 2.8)	...	1.397 (1.1, 1.8)
L Main disease	0.741 (0.6, 0.9)	...	...	...
Preop CVA/TIA	...	...	...	...
NYHA	1.656 (1.4, 1.9)	1.298 (1.1, 1.6)	...	...
Recent MI	0.757 (0.6, 0.9)	...	...	...
TVD	...	...	0.671 (0.5, 1.0)	...
P value for last variable in the model	0.029	0.021	0.024	0.024

Legend for Table 7.4 Cox regression risk ratios and 95% confidence intervals in parentheses. Age group: 1=age <65, 2=age 65-74, 3=age ≥75; LV Grade: 1-4, Surgical Priority: 0=Elective, 1=Same hospitalization, 2=Emergency, x=variable was not significant in the model. CHF=congestive heart failure, DM=diabetes, PVD=peripheral vascular disease, BP=hypertension, RF=renal failure, NYHA=New York Heart Association classification, left main disease, TVD=triple vessel disease, Preop CVA/TIA=history of stroke/transient ischemic attack, Recent MI=preoperative MI within a month prior to surgery.

Table 7.5 **Kaplan-Meier Estimates of Freedom From Re-admissions to Hospital for Cardiac Events**

Variable	1 Month	1 Year	5 Years
Age: <65	99.1 ● 0.2	95.6 ● 0.3	80.5 ● 0.9
65-74	98.5 ± 0.2	93.9 ± 0.5	77.7 ± 1.2
≥75	97.9 ± 0.6	90.2 ± 1.2	71.9 ± 2.4
Sex: Male	99.0 ± 0.1	95.3 ± 0.3	80.5 ± 0.8
Female	98.1 ± 0.4	91.6 ± 0.7	72.9 ± 1.6
LV Grade: 1	99.4 ± 0.2	96.4 ± 0.4	83.8 ● 1.0
2	98.8 ± 0.2	95.2 ● 0.4	81.6 ± 1.0
3	97.8 ± 0.4	92.2 ± 0.7	70.0 ± 1.8
4	97.1 ● 1.2	81.2 ± 2.8	53.8 ± 4.9
Previous CABG: N	98.8 ± 0.1	94.7 ± 0.3	79.6 ± 0.7
Y	98.2 ± 0.6	91.6 ± 1.3	70.4 ± 2.8
Surgical Priority: Elective	99.5 ± 0.1	96.8 ± 0.3	83.8 ± 0.8
Same Hospitalization	98.3 ± 0.2	92.3 ± 0.5	73.2 ± 1.2
Emergency	92.6 ± 1.8	84.7 ± 2.5	68.1 ± 4.0
Congestive Heart Failure: N	99.0 ± 0.1	95.3 ± 0.3	80.7 ± 0.7
Y	95.6 ± 0.9	84.1 ± 1.7	48.2 ± 4.8
Diabetes: N	98.8 ± 0.2	94.8 ± 0.3	80.9 ± 0.7
Y	98.8 ± 0.3?	93.6 ± 0.6	72.3 ± 1.6
Hypertension: N	99.0 ± 0.2	95.5 ● 0.4	81.7 ± 0.9
Y	98.6 ± 0.2	93.6 ± 0.4	76.4 ± 1.0
Peripheral Vascular Disease: N	98.9 ± 0.1	95.0 ± 0.3	80.6 ± 0.7
Y	97.8 ± 0.5	91.7 ± 0.9	68.0 ± 2.2
Renal Failure: N	98.9 ± 0.1	94.7 ± 0.3	79.8 ● 0.7
Y	95.8 ± 1.2	89.0 ± 1.9	59.1 ± 4.4

Legend for Table 7.5 All comparisons across risk factors are significant by Kaplan-Meier analysis, log-rank $P<0.001$.

Table 7.6 **Predictive Rule Risk Weights**

Variable	Risk Weights
Age: 65-74	2
≥75	3
Female	1
LV Grade: 2	1
3	2
4	3
Redo CABG	2
Surgical Priority: Same Hosp.	1
Emergency	2
Congestive heart failure	2
Peripheral vascular disease	2
Renal failure	2
History of hypertension	1
Diabetes	1

Legend for Table 7.6 The referent values (risk weight=0) are: age<65 years, male, LV Grade 1, primary CABG, elective priority, no congestive heart failure, no diabetes, no peripheral vascular disease, no hypertension, and no renal failure. Risk weights which characterize each patient were summed to create a total "risk score". The total risk score was divided into quartiles to construct four relative risk groups: Low Risk (LR=score 0-2), Medium Risk (MR=score 3-5), High Risk (HR=score 6-8), and Extreme Risk (ER=scores ≥9).

Table 7.7 **Demographics, Risk Factors and Outcomes for Relative Risk Groups Constructed From the Cox Regression Risk Ratios for Survival**

Variable	Low Risk N (%)	Medium Risk N (%)	High Risk N (%)	Extreme Risk N (%)
N	1854 (27)	2391 (34)	1788 (26)	954 (14)
Age: Mean ± SD	55 ● 7	60 ± 9	67 ± 8	71 ± 7
<65 year	1854 (100)	1517 (63)	416 (23)	115 (12)
65-74 yr	0	850 (36)	1093 (61)	496 (52)
≥75 yr	0	24 (1)	279 (16)	343 (36)
Female	133 (7.2)	440 (18)	527 (29)	370 (39)
LV Grade: 1	931 (50)	712 (30)	323 (18)	52 (5.5)
2	923 (50)	1146 (50)	852 (48)	275 (29)
3	0	523 (22)	542 (30)	490 (51)
4	0	10 (0.4)	71 (4.0)	137 (14)
Priority: Elective	1351 (73)	1403 (59)	786 (44)	237 (25)
Semi-Urgent	503 (39)	922 (39)	948 (53)	625 (66)
Emergency	0	66 (2.8)	54 (3)	92 (9.6)
Previous CABG	6 (0.3)	143 (6)	187 (10)	146 (15)
Preop CHF	0	30 (1.2)	130 (7.3)	359 (38)
NYHA: 1	63 (3.4)	55 (2.3)	30 (1.7)	5 (0.5)
2	456 (25)	324 (14)	124 (6.9)	28 (2.9)
3	768 (41)	935 (39)	572 (32)	180 (19)
4	567 (31)	1077 (45)	1062 (59)	741 (78)
ACI	340 (18)	651 (27)	718 (40)	491 (52)
Recent MI	166 (9)	381 (16)	360 (20)	354 (37)
TVD	1254 (68)	1727 (72)	1443 (81)	798 (84)
Left main disease	281 (15)	415 (17)	387 (22)	246 (26)
Diabetes	203 (11)	575 (24)	545 (30)	395 (41)
PVD	16 (0.9)	207 (8.9)	349 (20)	427 (45)
Hypertension	626 (34)	1174 (49)	1086 (61)	670 (70)
Renal Failure	0	36 (1.5)	77 (4.3)	177 (19)
Preop Stroke/TIA	44 (2.4)	147 (6.2)	197 (11)	197 (21)
Operative Mortality	6 (0.3)	33 (1.4)	52 (2.9)	71 (7.4)

Variable	Low Risk N (%)	Medium Risk N (%)	High Risk N (%)	Extreme Risk N (%)
Operative LOS	46 (2.5)	133 (5.6)	182 (10)	178 (19)
Operative Stroke	3 (0.2)	20 (0.8)	32 (1.8)	35 (3.6)
Total Deaths (6 yrs)	42 (2.3)	141 (5.9)	183 (10)	198 (21)
AMI	37 (2.0)	46 (1.9)	42 (2.4)	32 (3.4)
Chest Pain (AMI/UA)	131 (7.1)	184 (7.7)	174 (9.7)	111 (12)
CHF	8 (0.4)	60 (2.5)	106 (5.9)	139 (15)
PTCA	33 (1.8)	46 (1.9)	22 (1.2)	10 (1.0)
CABG	10 (0.5)	3 (0.1)	8 (0.4)	4 (0.4)
Any Cardiac Event	170 (9.2)	286 (12)	308 (17)	263 (28)
Mortality/Morbidity	207 (11)	399 (17)	450 (25)	389 (41)

Legend for Table 7.7 CHF=congestive heart failure, NYHA=New York Heart Classification, ACI=acute coronary insufficiency, MI=myocardial infarction, TVD=triple vessel disease, PVD=peripheral vascular disease, TIA=transient ischemic attack. Semi-urgent priority was defined as surgery occurring during the same admission as a cardiac catheterization or cardiac event. Emergency priority was surgery occurring within 12 hours of a cardiac catheterization or acute cardiac event. LOS=Low cardiac output syndrome, AMI=acute myocardial infarction, CHF=congestive heart failure, UA=unstable angina, PTCA=percutaneous transluminal coronary angioplasty.

Table 7.8 Regression Coefficients for the Log-Logistic Predictive Model

Variable	Regression Coefficient ± Std Err
Intercept	11.243 ± 0.425
Age group	-0.892 ± 0.107
Female	-0.243 ± 0.170
LV Grade	-0.573 ± 0.096
Previous CABG	-0.796 ± 0.241
Preoperative CHF	-0.820 ± 0.220
Surgical Priority	-0.395 ± 0.130
Diabetes	-0.507 ± 0.157
Hypertension	-0.362 ± 0.149
Peripheral vascular disease	-0.668 ± 0.177
Renal failure	-0.804 ± 0.258
Scale parameter	1.505 ± 0.063

Legend for Table 7.8 Age group: 1= age<65 years, 2=age 65-74 yrs, 3=age ≥75 years. LV Grade: 1=EF>60%, 2=EF 40-59%, 3=EF 20-39%, 4=EF <20%. CHF=congestive heart failure. Surgical priority: 0=Elective, 1=same hospitalization as the cardiac catheterization, 2=emergency (<12 hr) surgery. For all other binomial variables the values are 0 if the risk factor is not present and 1 if the risk factor is present.

The intercept, scale parameter and regression coefficients from the log-logistic analysis were used to calculate the predicted probability of survival to time=t from the formula:

$$S(t|X) = 1/1+[-g(\beta o + X)] * t^{g}$$

where S = survival probability, g = 1/scale parameter, β_0 = regression intercept, X = regression coefficients for covariates which describe each patient, t = follow-up time (months).

Table 7.9 Predicted Survival Calculated From the Multivariable Statistical Model

Risk Score (N)	1 Yr % (95% CI)	2 Yr % (95% CI)	3 Yr % (95% CI)	4 Yr % (95% CI)	5 Yr % (95% CI)
0 (275)	99.2 (100, 96)	98.8 (100, 96)	98.4 (100, 96)	98.1 (100, 96)	97.8 (100, 96)
1 (726)	98.9 (100, 89)	98.3 (100, 83)	97.8 (100, 79)	97.4 (100, 76)	97.0 (100, 73)
2 (853)	98.6 (100, 85)	97.8 (100, 79)	97.1 (100, 74)	96.5 (100, 70)	96.0 (100, 67)
3 (768)	98.3 (100, 82)	97.3 (100, 75)	96.5 (100, 69)	95.8 (99, 65)	95.2 (99, 62)
4 (843)	97.9 (100, 79)	96.7 (100, 70)	95.7 (99, 64)	94.9 (99, 60)	94.1 (99, 56)
5 (780)	97.4 (100, 75)	95.9 (100, 65)	94.7 (99, 59)	93.7 (99, 54)	92.8 (99, 50)
6 (742)	96.7 (100, 70)	94.9 (99, 59)	93.5 (99, 52)	92.2 (99, 48)	91.1 (99, 44)
7 (559)	95.9 (100, 64)	93.6 (99, 52)	91.8 (99, 46)	90.2 (99, 41)	88.8 (99, 37)
8 (487)	94.8 (99, 57)	92.0 (99, 46)	89.8 (99, 46)	87.9 (99, 39)	86.2 (99, 32)
9 (337)	93.5 (99, 51)	90.0 (99, 40)	87.3 (98, 33)	85.1 (98, 29)	83.1 (98, 26)
10 (230)	91.9 (99, 44)	87.7 (99, 34)	84.5 (98, 28)	81.9 (98, 24)	79.6 (97, 22)
11 (141)	89.9 (99, 38)	84.9 (98, 28)	81.1 (98, 23)	78.0 (97, 20)	75.3 (97, 17)
12 (102)	87.4 (99, 32)	81.4 (98, 23)	76.9 (97, 19)	73.4 (96, 16)	70.4 (96, 14)
13 (52)	84.6 (98, 27)	77.6 (97, 19)	72.6 (96, 15)	68.6 (96, 13)	65.4 (95, 11)
≥14 (92)	81.6 (97, 18)	73.7 (95, 12)	68.2 (94, 10)	63.9 (93, 8)	60.4 (92, 7)

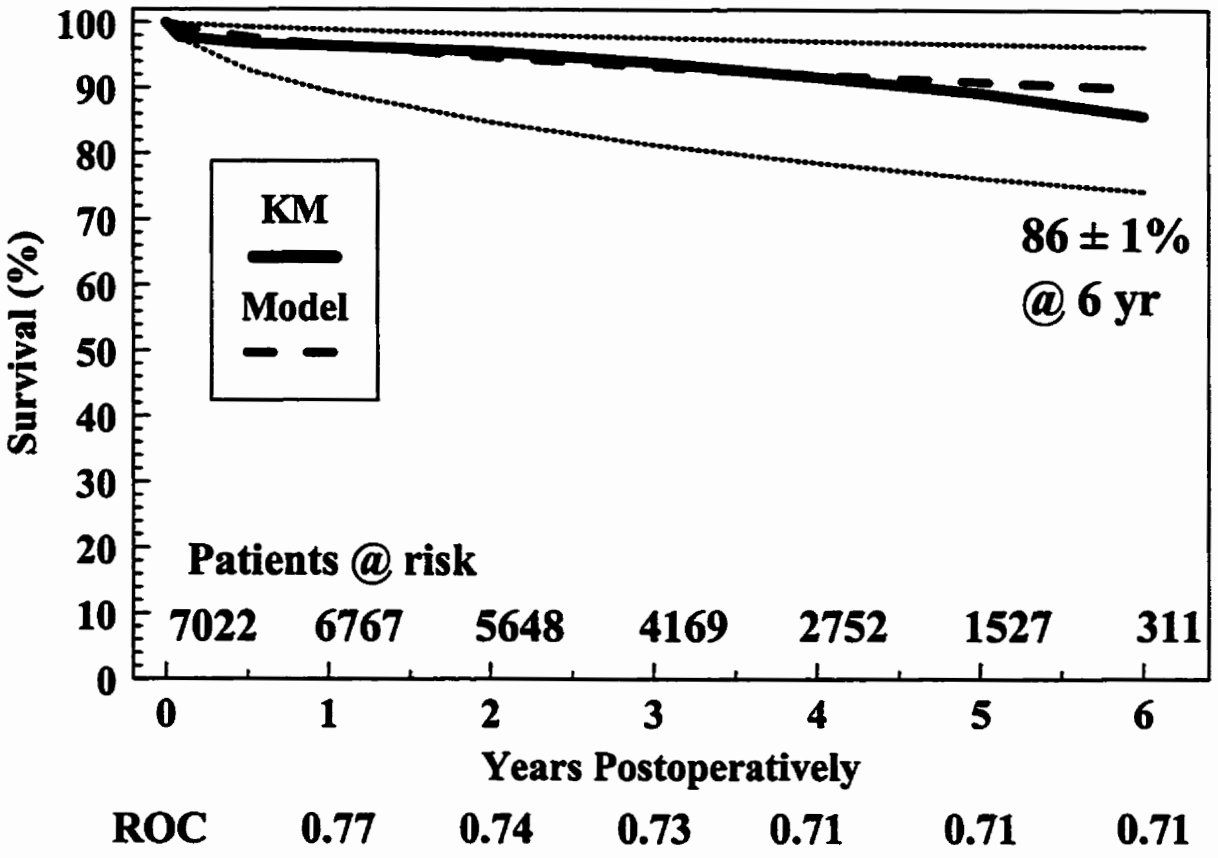
Legend for Table 7.9 Predicted survival (%) and the 95% confidence intervals (CI) in parentheses as calculated by the multivariable log-logistic model.

Table 7.10 **Comparison of Predicted versus Observed Survival Probabilities and Number of Deaths**

Risk Group (Score)	OBS 5 yr Survival (%)	PRED 5 yr Survival (%)	R^2	# OBS Deaths	# PRED Deaths	% PRED	χ^2 P
LR (0 - 2)	96 ± 1	97 ● 0.02	0.52	49	52	106	0.92
MR (3 - 5)	92 ± 1	94 ± 0.02	0.61	141	116	82	0.06
HR (6 - 8)	86 ± 1	89 ± 0.05	0.58	183	158	86	0.11
ER (≥9)	71 ± 2	76 ± 0.31	0.28	198	182	92	0.42

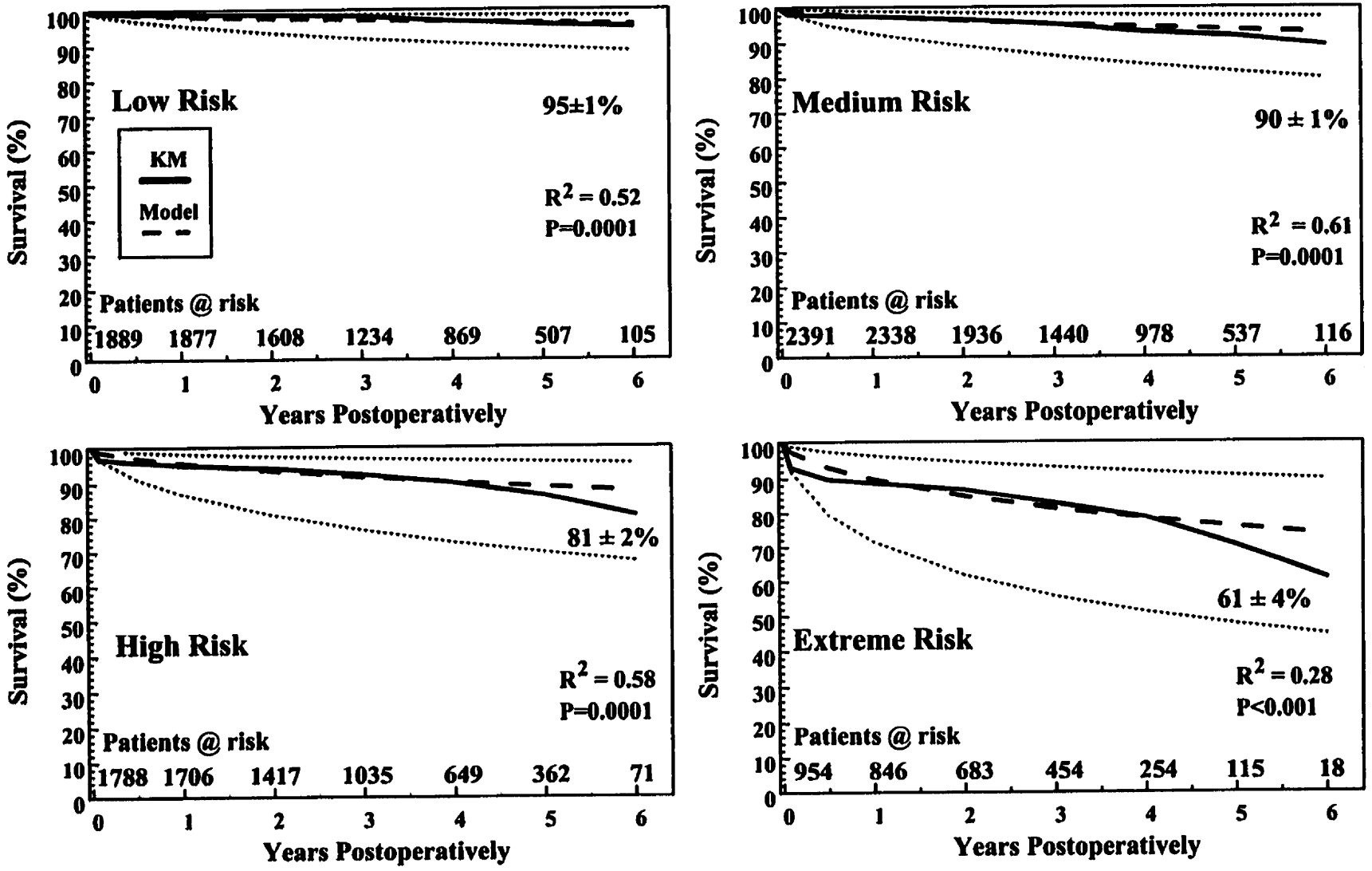
Legend for Table 7.10 LR=low risk, MR=medium risk, HR=high risk, ER=extremely high risk. Observed (OBS) and predicted (PRED) survival probabilities are reported as the mean ± standard error. All P values associated with the R^2 results were significant at the $P<0.001$ level. %PRED = #PRED/#OBS. Chi square (χ^2) P values >0.05 indicate *no* statistical difference between predicted and observed number of events.

Figure 7.1



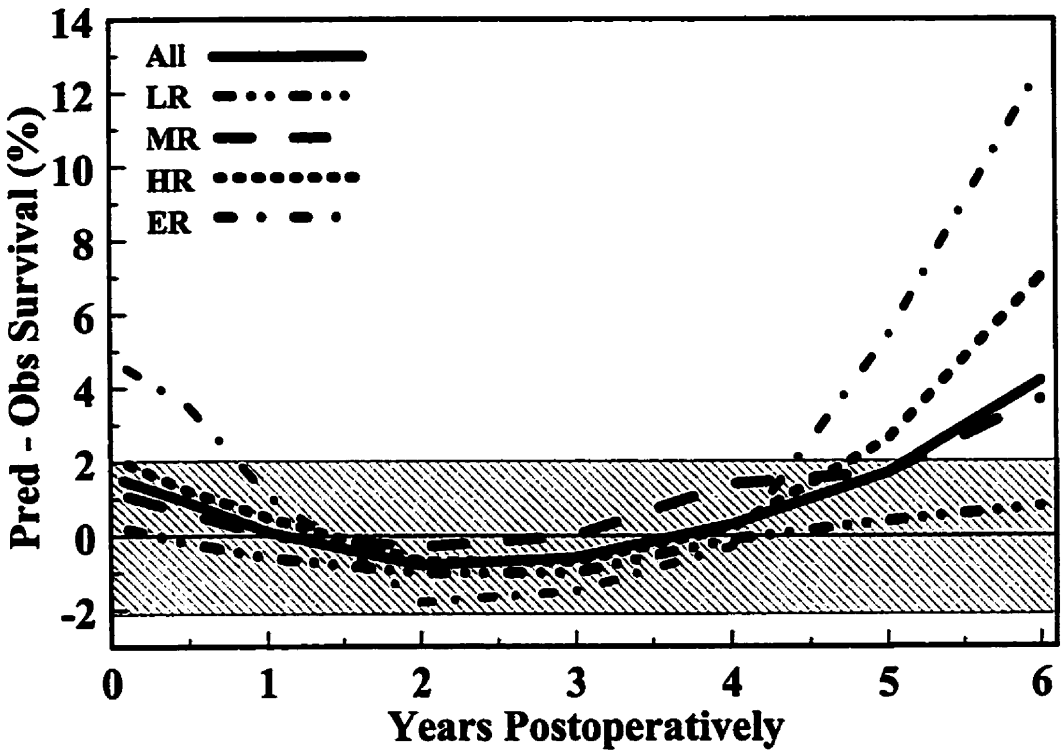
Legend 7.1 The observed cumulative survival probability as calculated by the Kaplan-Meier (KM) analysis for all patients is depicted by the solid, heavy line. The average predicted probability as calculated by the multivariable statistical model is depicted by the heavy dashed line ± 1 standard error (light, dotted lines). The area under the Receiver-Operator Characteristic (ROC) curve for model estimates is given at each yearly interval.

Figure 7.2



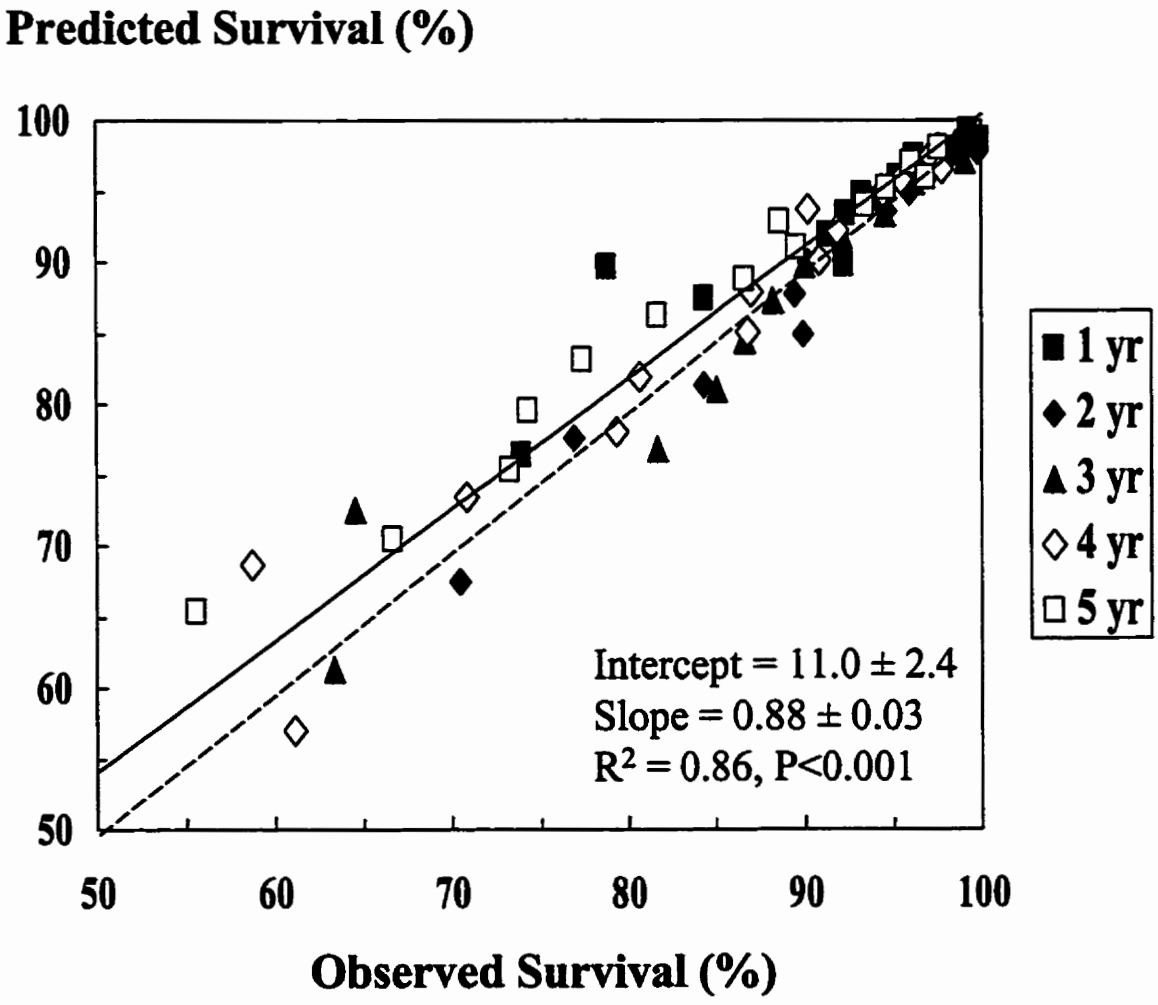
Legend 7.2 Same as the legend for Figure 2. In these figures, the predicted probability of survival as calculated by the statistical model is superimposed over the observed survival for each relative risk group. The 6 yr observed value +/- 1 standard error is presented as text on each graph. The results of the linear regression analysis of observed versus predicted survival are given as the R square and associated P values.

Figure 7.3



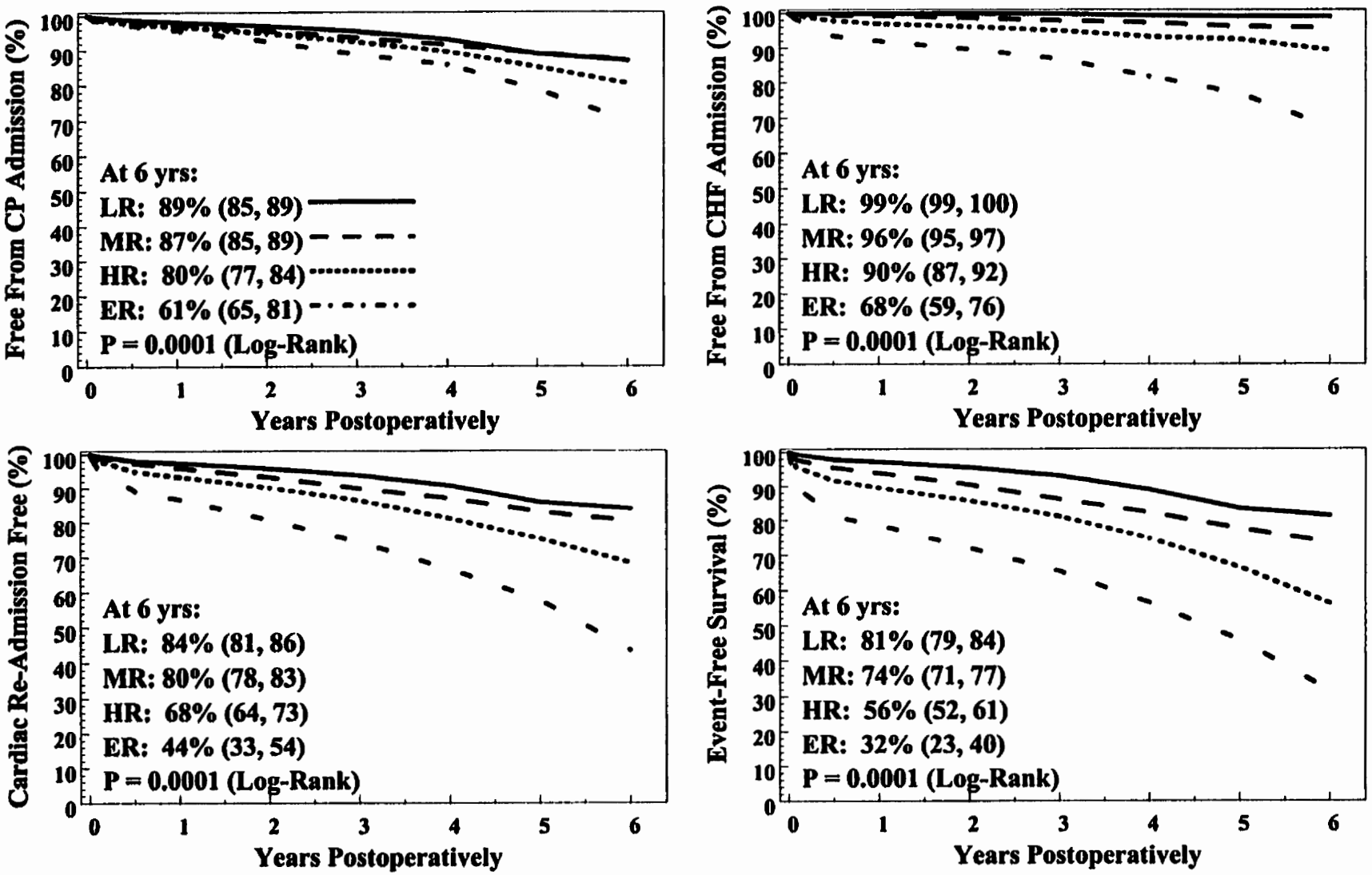
Legend 7.3 The average difference between predicted minus observed survival is depicted for each relative risk group at each time interval. The statistical model over-estimated survival after 4 postoperative years. This loss of precision was most dramatic in the high risk (HR) and extremely high risk (ER) groups. There was also a significant over-estimation of survival in the ER group in the early postoperative period. The stipled area represents ±2% from a zero difference between observed and predicted probabilities.

Figure 7.4



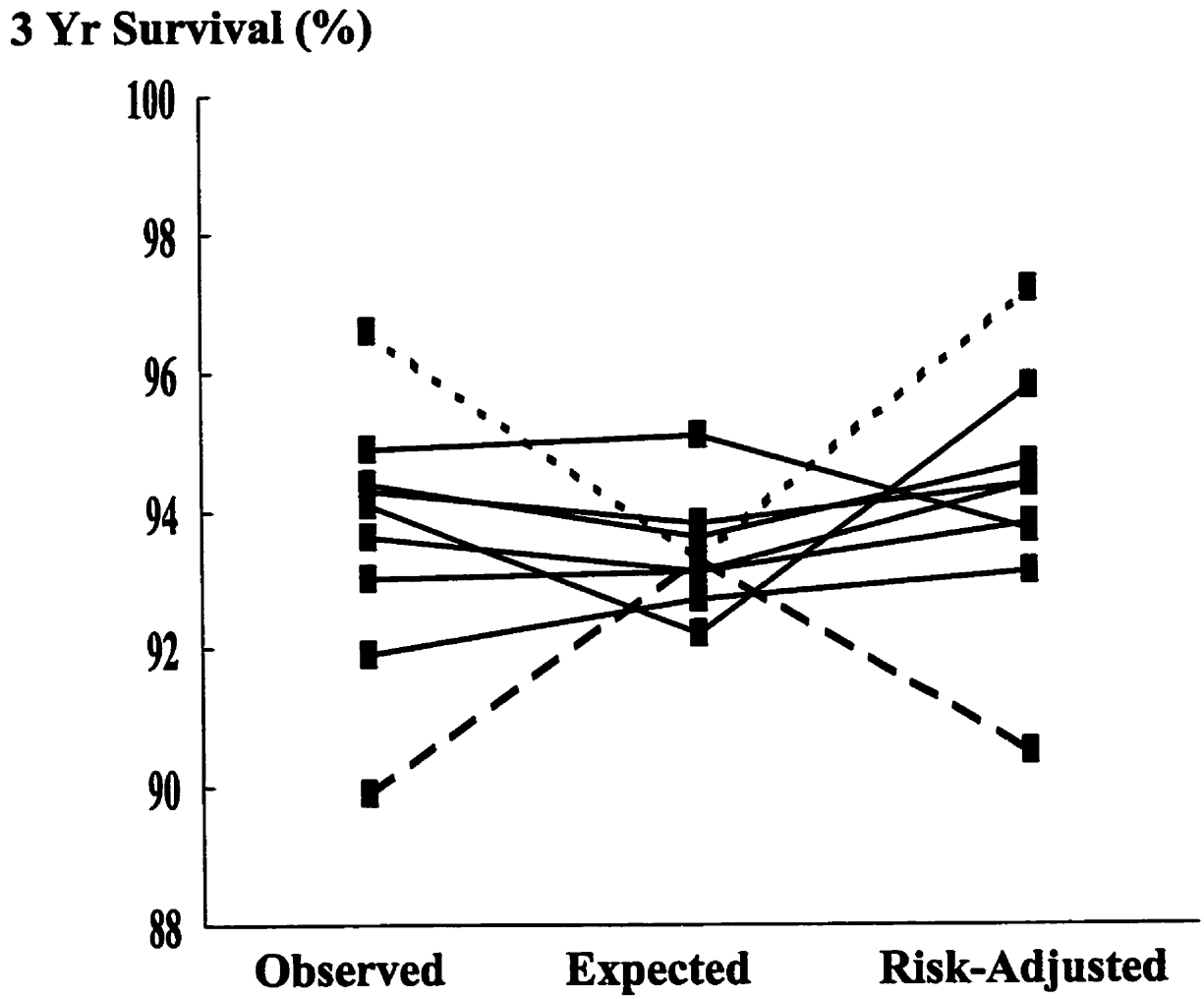
Legend 7.4 The average predicted probability of survival as calculated by the multivariable statistical model is plotted against the mean observed survival for each yearly interval and each total risk score (represented by each symbol). Results are given for the weighted linear regression (solid diagonal line). The dashed diagonal line represents a perfect fit between predicted and observed survival.

Figure 7.5



Legend 7.5 Kaplan-Meier estimates of the cumulative probability of freedom from re-admissions for chest pain (CP), congestive heart failure (CHF), any cardiac event (Cardiac) and event-free survival are depicted for each relative risk group. The 6 year estimates are given in the text with their 95% confidence intervals.

Figure 7.6



Legend 7.6 Comparison of observed, expected and risk-adjusted three year survival for each provider. Note how the two providers represented by dashed lines have the same expected survival based on their case-mix, but different observed and risk-adjusted survival estimates.

Chapter Eight

Summary

8.1 Overview

Methodological issues which surround the development of predictive rules within the contextual framework of coronary artery bypass surgery include, but are not limited to: 1) prognostic variable selection; 2) statistical under- or over-fitting; 3) model discrimination and calibration; 4) selection of an external or internal model; and 5) model validation. Among the applications for predictive rules for operative mortality following CABG are: 1) risk-adjusted outcomes profiling for surgeons or centres, 2) risk group stratification, and 3) decision-support for pre-operative patient counselling. These methodological issues have been evaluated in this applied thesis.

8.2 Chapter Four: Evaluation of Three Modelling Strategies in the Use of Predictive Risk Indices for Assessing Mortality After Coronary Artery Bypass Graft Surgery

In Chapter Four we demonstrated that temporal and inter-centre differences in case-mix make it difficult to achieve optimal predictive performance with "ready-made", external risk indexes used for provider profiling through evaluations of risk-adjusted outcomes. This observation argues against the proliferation of published risk indices in the clinical literature that either affirm well-known prognostic factors or add new variables with minimal marginal impact. We have also demonstrated that recalibration of existing indices may sometimes be sufficient to ensure adequate risk prediction even when models are parsimonious. As a precaution, however, we suggested that centres collect data fastidiously on a modest-sized set of key variables such as those suggested by the Working Group Panel,[70] continue to evaluate the model's validity, and undertake intermittent remodelling to ensure that emerging risk factors are not inadvertently overlooked.

In addition to these conceptual contributions, this chapter also introduced a

refinement in the quantitative evaluation of model calibration. Several authors have previously used linear regression analysis to calculate the coefficient of determination (R^2) of predicted versus observed results for subgroups of patients. We showed that the intercept and slope of the regression line should also be evaluated because they give direct assessments of the amount of over- or under-estimation of the model. Additionally, we introduced analysis of covariance as an effective method to compare competing models in the same dataset. This study illustrated that poorly calibrated risk algorithms can bias the calculation of risk-adjusted operative mortality.

8.3 Chapter Five: Application of a Predictive Rule Used for Risk Stratification: Evaluation of Fifteen-year Trends in Risk Severity and Operative Mortality in Elderly Patients Undergoing Coronary Bypass Surgery

In Chapter Five we demonstrated the application of a predictive rule used for risk stratification. The construction of relative risk groups from cutpoints of the total risk score permitted the evaluation of temporal trends in risk factors and outcomes using simple contingency table analysis. To the average reader, this format is clearer compared to the relative complexity of interpreting adjusted odds ratios from logistic regression analyses.

In this chapter we also demonstrated that the inclusion of additional variables to small, robust models did not significantly improve discrimination: the model for mortality in elderly patients between 1982 and 1996 had a ROC curve area similar (0.69 versus 0.71) to that for elderly patients in the 1991-1996 model which drew on additional risk factor data.

The importance of statistical power for the evaluation of outcomes in relatively low prevalence patient groups such as the elderly was an important element of this study. Most previous reports have had insufficient sample sizes to adequately explore

the multivariable impact of risk factors on operative mortality. Additionally, they have been limited by not providing a long-term perspective on outcome trends among the elderly, and by not incorporating risk-adjustment algorithms that take into account the temporal shifts in risk profiles among patients receiving CABG. Our logistic regression analysis of operative mortality in 3,330 patients from 1982 to 1986 included a temporal marker and revealed that most elderly patients have an acceptable risk of surgery, however, poor ventricular function and previous CABG remain the most important risk factors in this group.

In addition to providing a long-term perspective on outcome trends, this study also showed how a clinical prediction rule can be combined with simple contingency table analysis to document temporal shifts in risk profiles and outcomes.

8.4 Chapter Six: Predictive Accuracy Study: Comparison of a Statistical Model to Clinicians' Probability Estimates of Adverse Outcomes Following Coronary Bypass Surgery

The potential application of clinical prediction rules for decision-support for pre-operative patient counselling was evaluated in Chapter Six. This study was flawed from two important perspectives. First, the study sample of 100 patients was not randomly selected from the database. This was deliberate but severely compromised the ability of both the statistical model and the clinicians to discriminate between those having, and those not having events. However, the evaluation of discrimination through a unique prevalence adjustment exercise demonstrated that the model ultimately did out-perform the clinicians.

Secondly, the clinicians did not make use of the predictive rule for the part of the study in which they were randomized to have access to it. This was despite at least three discussions of the study during rounds and several written communications regarding the

purpose and methods of the study. When given the option, neither the surgeons nor the residents preferred to use a predictive rule but rather trusted their own judgements to estimate the probability of operative mortality or prolonged ICU length of stay after coronary bypass surgery. Reasons for not using the predictive rule may have included: 1) the rule was too time consuming, 2) too complex, or 3) clinicians preference to use their own empiric estimates. The findings of this study highlight both the rationale for using predictive rules as a guide to decision-making in practice, as well as the continuing challenges in persuading clinicians to use such rules.

8.5 Chapter Seven: The Development and Application of a Predictive Rule to Evaluate the Long-Term Outcomes of Coronary Artery Bypass Surgery

Two recurrent issues in the use of clinical prediction rules are the short-time horizon of many studies and the costs of collecting data for longer-term studies. Chapter Seven presents the methods and findings for a unique Canadian study. The linkage of a clinical database with administrative data allowed for the efficient follow-up of a large contemporary cohort of CABG patients and the identification of multivariable predictors of survival and re-admission to hospital for cardiac events. This study highlights the advantage of a universal health care system which allows researchers to build population-based inclusive linked datasets of this nature. Actuarial estimates from registry and administrative data were similar to estimates made from more traditional follow-up methods in previously published studies.

Additionally, an accelerated failure time model was used to generate a predicted survival curve for each patient. After stratifying by risk group, the model predicting long-term survival demonstrated a good correlation between predicted and observed outcomes for the majority of patients. However, the results in higher risk patients suggested that the predictive rule was not as precise for estimates made after four to five

years. Future studies will evaluate whether the addition of postoperative events improves predictive accuracy of the survival model. We also explored a method of comparing provider-specific long-term risk-adjusted survival.

This predictive algorithm may provide a valuable tool for medical decision making, internal benchmarking, trial design and patient counselling. Prospective external validation is still required and will be the focus of future research.

This study illustrates the feasibility of linking clinical and administrative data in the development of more sophisticated clinical prediction rules. The findings highlight the need to evaluate long-term, as well as short-term outcomes to assess the benefit of surgical revascularization and the utility of adding some measurement of outcomes other than mortality.

8.6 Conclusions

Patients with the same disease can have very different outcomes. Predictive rules which adjust for differences in patient characteristics can offer valuable insights for clinicians, administrators and researchers. However, Parsonnet's cautionary note is well founded: the validity of the model can greatly influence the calculation of risk-adjusted outcomes. Krumholz[224] suggested in a recent editorial, that the proposed application of a predictive rule should dictate the necessity of evaluating calibration. He stated that relative comparisons between providers did not require the recalibration of a previously validated external predictive model. We disagree with this philosophy. Models used for provider profiling should be well-calibrated across the full range of estimated outcomes. Models that over-estimate probabilities at the lower range of outcomes and under-estimate probabilities for higher risk patients may obscure the differences between providers with different case mixes. A precise model will truly "level the playing field" and result in a fairer comparison of risk-adjusted outcomes.

Previously well-validated models do not need to be recalibrated if the application is benchmarking temporal trends (e.g., the study presented in Chapter Five). If the application of the model is for the purpose of patient counselling, recalibration of an external model is essential to provide valid, contemporary estimates of outcomes.

Regardless of the application, all predictive rules must discriminate well. Good discrimination is the most fundamental element of a valid predictive rule and is directly influenced by the number of variables in the model and the quality of the data. The issue regarding the basic building block of any predictive rule--the prognostic variable--has not been well addressed. As with any predictive rule developed from an existing database, we are constrained by the data at hand. Many measurement issues surround the collection of data, not the least of which are outcome ascertainment, prognostic variable selection and using continuous versus categorical variables to describe disease states or risk factors (e.g. age). In the clinical database at the Toronto General Hospital we generally categorize most of the risk factor information. Statisticians argue that the practice of categorizing continuous variables results in a loss of information. It may also negatively impact model discrimination. However, clinicians argue that this practice enhances "usability".

Another major issue surrounding variable selection is the lack of information in most databases regarding the quality of the distal coronary arteries. This piece of information may be one of the most important predictors of outcome. However, no existing predictive rule for OM following CABG includes a measure of distal vessel quality. The inclusion of some measure of distal vessel quality could possibly improve the discrimination of most models to well above the apparent ceiling of ~80%.

Additionally, in none of our models have we included the surgeon as a potential independent predictor of outcome. The inclusion of surgeon or other markers of process of care may improve model discrimination, unless the purpose is to benchmark surgeons,

in which case, surgeon could not be included in the model.

We have illustrated that: 1) the potential bias in the calculation of risk-adjusted operative mortality from poorly calibrated models can affect provider-profiling; 2) risk stratification via predictive rules can be combined with simpler analyses for research and administrative purposes; 3) there are continuing challenges for using predictive rules as guides for decision-making in practice and in persuading clinicians to use such rules; and 4) a predictive rule designed for long-term outcomes provides a novel method to evaluate risk-adjusted survival following surgical revascularization.

8.7 Future Directions

We wish to develop similar predictive rules for outcomes following valve surgery. Edmunds and colleagues[225] explicitly defined the variables which should be included in evaluations of mortality and morbidity following valvular heart surgery. This paper was printed simultaneously in the *Journal of Thoracic and Cardiovascular Surgery* and in the *Annals of Thoracic Surgery* and serves as a template for all authors who are considering submission of a valve-related manuscript to any major cardiac surgery journal. As a direct result of this effort, it has been possible for us to link our clinical database with data from Vancouver General Hospital and Stanford University Medical Centre to facilitate multicentre evaluations of valve surgery. These studies (still in the analytic phase) represent large patient samples with sufficient events to produce valid multivariable models of outcomes. Additionally, the generalizability of the results is greatly enhanced by this collaborative effort.

Future studies are planned to prospectively evaluate surgeons' risk assessment and to determine whether the use of a predictive rule at the time of patient assessment improves their predictive accuracy and precision. We have developed a user-friendly, Windows-based, Visual Basic program (written by Matthew Weisel), which incorporates

the regression parameters from the models to calculate the predicted probabilities and 95% confidence intervals of operative mortality and prolonged ICU length of stay for patients undergoing isolated coronary artery bypass graft surgery. The program allows clinicians to check off patient risk factors on a computer screen during the pre-admission clinic visit. Each surgeon at the Toronto General Hospital has been given this program as an aid to counsel patients at the time of obtaining informed consent. We have demonstrated its use at one meeting and plan for several more demonstrations. The importance of these studies is to determine whether or not the assumption of risk provides a barrier to the care of high risk patients, and to evaluate a strategy of streaming high risk patients towards surgeons with better outcomes in an effort to reduce overall mortality and morbidity.

The evaluation of long-term outcomes is important to establish the effectiveness and efficiency of surgical revascularization. We plan to link the Province of Ontario Cardiac Care Network database to administrative databases (CIHI and RPDB) to evaluate long-term outcomes for coronary artery bypass surgery. Although we have used purely prognostic variables in all of our predictive models, we will evaluate whether the inclusion of some measures of process of care will improve precision and discrimination of long-term outcomes of surgery. The identification of non-cardiac morbidity will present the biggest challenge in this study. This study will allow us to examine regional differences in processes and outcomes of care.

In conclusion, drawing on data for coronary artery bypass graft surgery, this thesis addresses several methodological and practical issues surrounding the development and application of predictive rules. We believe that these findings are generalizable to a wide range of medical diagnoses and surgical procedures where prediction rules can be used to enhance our insights into prognosis, as well as the effectiveness and efficiency of clinical interventions.

Bibliography

1. Naftel DC: Do different investigators sometimes produce different multivariable equations from the same data? [Editorial]. *J Thorac Cardiovasc Surg* 1994;107:1528-1529

2. Bannwort Dd, Rebello CM, Sadick LS, Pontes MD, Ramos JL, Leone CR: Prediction of length of stay in neonatal units for very low birth weight infants. *J Perinatol* 1999;19:92-96

3. Koheler D, Arbel E, Goldberg M, Arlazoroff A: Brainstem auditory evoked response in newborns and infants. *J Child Neurol* 2000;15:33-35

4. Prasad K: The Glasgow Coma Scale: A critical appraisal of its clinimetric properties. *J Clin Epidemiol* 1996;49:755-763

5. Cooper A, Hannan EL, Bessey PQ, Farrell LS, Cayten CG, Mottley L: An examination of the volume-mortality relationship for New York State trauma centers. *J Trauma* 2000;48:16-23

6. Herbert JS, Burnham RS: The effect of polytrauma in persons with traumatic spine injury. A prospective database of spine fractures. *Spine* 2000;25:55-60

7. Laupacis A, Sekar N, Stiell IG: Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-494

8. Becker RB, Zimmerman JE, Knaus WA, Wagner DP, Seneff MG, Draper EA, Higgins TL, Estafanous FG, Loop FD: The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery by-pass surgery. *J Cardiovasc Surg* 1995;36:1-11

9. Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ: A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. *Crit Care Med* 1994;22:750-762

10. Naylor CD, Baigrie RS, Goldman BS, Basinski A: Assessment of priority for coronary revascularisation procedures. *Lancet* 1990;335:1070-1073

11. Naylor CD, Baigrie RS, Goldman BS, Cairns JA, Beanlands DS, Berman N, Borts D, Fitchett DH, Hag A, Hess A, Hughes W, Swan J, Tinmouth A, Gunstensen J, Lichtenstein SV, Salter D, Weisel RD, Anderson GM, Basinski A, Lomas J: Assigning priority to patients requiring coronary revascularization: Consensus principles from a panel of cardiologists and cardiac surgeons. *Can J Cardiol* 1991;7(5):207-213

12. Naylor CD, Morgan CD, Levinton CM, Wheeler S, Hunter L, Klymciw K, Baigrie RS, Goldman BS: Waiting for coronary revascularization in Toronto: 2 years' experience with a regional referral office. *Can Med Assoc J* 1993;149(7):955-962

13. Naylor CD, Sykora K, Jaglal SB, Jefferson S: Waiting for coronary artery bypass surgery: population-based study of 8517 consecutive patients in Ontario, Canada. The Steering Committee of the Adult Cardiac Care Network of Ontario [see comments]. *Lancet* 1995;346:1605-1609

14. Wasson JH, Sox HC, Neff RK, Goldman L: Clinical prediction rules: applications and methodologic standards. *N Engl J Med* 1985;313:793-799

15. Shapiro AR: An evaluation of clinical prediction: A method and initial application. *N Engl J Med* 1977;296:1509-1514

16. Spiegelhalter DJ: Probabilistic prediction in patient management and clinical trials. *Statistics in Medicine* 1986;5:421-433

17. Parsonnet V, Dean D, Bernstein AD: A method of uniform stratification of risk factors for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79:I3-12

18. Parsonnet V: Risk stratification in cardiac surgery: is it worthwhile? *J Cardiac Surg* 1995;10:690-698

19. Parsonnet V, Bernstein AD, Gera M: Clinical usefulness of risk-stratified outcome analysis in cardiac surgery in New Jersey. *Ann Thorac Surg* 1996;61:S8-11; discussion S33-4

20. Hannan EL: Using mortality data for profiling hospital quality of care and targeting substandard care. [Review]. *Journal of the Society for Health Systems* 1989;1:31-48

21. 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;264:2768-2774

22. Hannan EL, Kilburn H Jr, Bernard H, O'Donnell JF, Lukacik G, Shields EP: Coronary artery bypass surgery: The relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. *Med Care* 1991;29:1094-1107

23. Hannan EL, Kilburn H Jr, Lindsey ML, Lewis R: Clinical versus administrative data bases for CABG surgery: Does it matter? *Med Care* 1992;30:892-907

24. 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;58:1852-1857

25. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR: Improving the outcomes of coronary artery bypass surgery in New York state. *JAMA* 1994;271:761-766

26. 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;273:209-213

27. Hannan EL, Racz MJ, Jollis JG, Peterson ED: Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Services Research* 1997;31:659-678

28. Hannan EL, Siu AL, Kumar D, Racz M, Pryor DB, Chassin MR: Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients? *Med Care* 1997;35:49-56

29. Edwards FH, Graeber GM: The theorem of Bayes as a clinical research tool. *Surgery, Gynecology & Obstetrics* 1987;165:127-129

30. Edwards FH, Schaefer PS, Callahan S, Graeber GM, Albus RA: Bayesian statistical theory in the preoperative diagnosis of pulmonary lesions. *Chest* 1987;92:888-891

31. Edwards FH, Cohen AJ, Bellamy RF, Thompson L, Weston L: Risk assessment in urgent/emergent coronary artery surgery. *Chest* 1990;97:1125-1129

32. Edwards FH, Clark RE, Schwartz M: Coronary artery bypass grafting: The Society of Thoracic Surgeons National Database experience. *Ann Thorac Surg* 1994;57:12-19

33. Edwards FH, Grover FL, Shroyer AL, Schwartz M, Bero J: The Society of Thoracic Surgeons National Cardiac Surgery Database: current risk assessment. *Ann Thorac Surg* 1997;63:903-908

34. Edwards FH, Albus RA, Zajtchuk R, Graeber GM, Barry MJ, Rumisek JD, Arishita G: Use of a Bayesian statistical model for risk assessment in coronary artery surgery. *Ann Thorac Surg* 1988;45:437-440

35. Clark RE: The Society of Thoracic Surgeons National Database Status Report. *Ann Thorac Surg* 1994;57:20-26

36. Clark RE: Calculating risk and outcome: The Society of Thoracic Surgeons database. *Ann Thorac Surg* 1996;62:S2-5; discussion S31-2

37. O'Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, Nowicki ER, Levy DG, Tryzelaar JF, Hernandez F, Adrian L, Casey KJ, Bundy D, Soule DN, Marrin CAS, Nugent WC, Charlesworth DC, Clough R, Katz S, Leavitt BJ, Wennberg JE: Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. *Circulation* 1992;85:2110-2118

38. O'Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, Nowicki ER, Tryzelaar JF, Hernandez F, Adrian L, Casey KJ, Soule DN, Marrin CAS, Nugent WC, Charlesworth DC, Clough R, Katz S, Leavitt BJ, Wennberg: A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. *JAMA* 1996;275:841-846

39. Tu JV, Jaglal SB, Naylor CD, the Steering Committee of the Provincial Adult Cardiac Care Network of Ontario: Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation* 1995;91:677-684

40. Orr RK, Naini BS, Sottile FD, Dumas EM, O'Mara P: A comparison of four severity-adjusted models to predict mortality after coronary artery bypass graft surgery. *Arch Surg* 1995;130:301-306

41. Higgins TL, Estafanous FG, Loop FD, Berck GJ, Blum JM, Paranandi L: Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients: A clinical severity score. *JAMA* 1992;267:2344-2348

42. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Lee JC, Starr NJ, Knaus, WA, Cosgrove DM, 3rd. ICU admission score for predicting morbidity and mortality risk after coronary artery bypass grafting. *Ann Thorac Surg* 1997;64:1050-1058

43. Hartz AJ, Kuhn EM, Kayser KL, Pryor DP, Green R, Rimm AA: Assessing providers of coronary revascularization: a method for peer review organizations. *Am J Public Health* 1992;82:1631-1640

44. Hartz AJ, Kuhn EM, Green R, Rimm AA: The use of risk-adjusted complication rates to compare hospitals performing coronary artery bypass surgery or angioplasty. *International Journal of Technology Assessment in Health Care* 1992;8:524-538

45. Hartz AJ, Kuhn EM, Pryor DB, Krakauer H, Young M, Heudebert G, Rimm, AA. Mortality after coronary angioplasty and coronary artery bypass surgery (the national Medicare experience). *Am J Cardiol* 1992;70:179-185

46. Hartz AJ, Gottlieb MS, Kuhn EM, Rimm AA: The relationship between adjusted hospital mortality and the results of peer review. *Health Services Research* 1993;27:765-777

47. Hartz AJ, Kuhn EM: Comparing hospitals that perform coronary artery bypass surgery: the effect of outcome measures and data sources. *Am J Public Health* 1994;84:1609-1614

48. Hartz A, Guse C, Kayser K, Kuhn E, Johnson D: Use of postoperative information to predict mortality rates for patients who have long stays in the intensive care unit after coronary artery bypass grafting. *Heart & Lung* 1998;27:22-30

49. 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; discussion 1306

50. Hammermeister KE: Participatory continuous improvement. *Ann Thorac Surg* 1994;58:1815-1821

51. Hammermeister KE, Daley J, Grover FL: Using outcomes data to improve clinical practice: what we have learned. *Ann Thorac Surg* 1994;58:1809-1811

52. Hammermeister KE, Johnson R, Marshall G, Grover FL: Continuous assessment and improvement in quality of care. A model from the Department of Veterans Affairs Cardiac Surgery. *Ann Surg* 1994;219:281-290

53. Grover FL, Johnson RR, Marshal G, Hammermeister KE: Factors predictive of operative mortality among coronary artery bypass subsets. *Ann Thorac Surg* 1993;56:1296-1307

54. Hattler BG, Madia C, Johnson C, Armitage JM, Hardesty RL, Kormos RL, Pham SM, Payne DN, Griffith BP: Risk stratification using the Society of Thoracic Surgeons Program. *Ann Thorac Surg* 1994;58:1348-1352

55. Dziuban SW, Jr., McIlduff JB, Miller SJ, Dal Col RH: How a New York cardiac surgery program uses outcomes data. *Ann Thorac Surg* 1994;58:1871-1876

56. Griffith BP, Hattler BG, Hardesty RL, Kormos RL, Pham SM, Bahnson HT: The need for accurate risk-adjusted measures of outcome in surgery. Lessons learned through coronary artery bypass. *Ann Surg* 1995;222:593-8; discussion 598-9

57. Shroyer AL, London MJ, VillaNueva CB, Sethi GK, Marshall G, Moritz, TE, Henderson WG, McCarthy MJ, Jr., Grover FL, Hammermeister KE: The Processes, Structures, and Outcomes of Care in Cardiac Surgery study protocol. *Med Care* 1995;33:OS17-25

58. Shroyer AL, London MJ, Sethi GK, Marshall G, Grover FL, Hammermeister, KE. Relationships between patient-related risk factors, processes, structures, and outcomes of cardiac surgical care. Conceptual models. *Med Care* 1995;33:OS26-34

59. Shroyer AL, Dauber I, Jones RH, Daley J, Grover FL, Hammermeister KE: Provider perceptions in using outcomes data to improve clinical practice [see comments]. *Ann Thorac Surg* 1994;58:1877-1880

60. Shroyer AL, Grover FL, Edwards FH: 1995 coronary artery bypass risk model: The Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg* 1998;65:879-884

61. Edmunds LH, Stephenson LW, Edie RN, Ratcliffe MB: Open-heart surgery in octogenarians. *N Engl J Med* 1988;319:131-136

62. Henderson WG, Moritz TE, Shroyer AL, Johnson R, Marshall G, Ellis NK, Sethi GK, Grover FL, Hammermeister KE: An analysis of interobserver reliability and representativeness of data from the Veterans Affairs Cooperative Study on Processes, Structures, and Outcomes in Cardiac Surgery. *Med Care* 1995;33:OS86-101

63. Magovern JA, Sakert T, Magovern GJ, Jr., Benckart DH, Burkholder JA, Liebler GA, Magovern GJ, Sr. A model that predicts morbidity and mortality after coronary artery bypass graft surgery. *J Am Coll Cardiol* 1997;28:1147-1153

64. Tu JV, Guerriere MR: Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Proceedings - the Annual Symposium on Computer Applications in Med Care* 1992;666-672

65. Tu JV, Mazer CD, Levinton C, Armstrong PW, Naylor CD: A predictive index for length of stay in the intensive care unit following cardiac surgery. *Can Med Assoc J* 1994;151:177-185

66. Tu JV: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. [Review] [37 refs]. *J Clin Epidemiol* 1996;49:1225-1231

67. Tu JV, Sykora K, Naylor CD: Assessing the outcomes of coronary artery bypass graft surgery: How many risk factors are enough? *J Am Coll Cardiol* 1997;30:1317-1323

68. Tu JV, Weinstein MC, McNeil BJ, Naylor CD: Predicting mortality after coronary artery bypass surgery: what do artificial neural networks learn? The Steering Committee of the Cardiac Care Network of Ontario. *Medical Decision Making* 1998;18:229-235

69. Daley J: Criteria by which to evaluate risk-adjusted outcomes programs in cardiac surgery. *Ann Thorac Surg* 1994;58:1827-1835

70. Jones RH, Hannan EL, Hammermeister KE, DeLong ER, O'Connor GT, Luepker RV, Parsonnet V, Pryor DB: 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. [Review] [35 refs]. *J Am Coll Cardiol* 1996;28:1478-1487

71. Weightman WM, Gibbs NM, Sheminant MR, Thackray NM, Newman MA: Risk prediction in coronary artery surgery: a comparison of four risk scores. *Med J Australia* 1997;166:408-411

72. Ivanov J, Tu JV, Naylor CD: Ready-made, recalibrated, remodelled? Issues in the use of risk indices for assessing mortality after coronary artery bypass graft surgery. *Circulation* 1999;99:2098-2104

73. Ivanov J, Weisel RD, David TE, Naylor CD: Fifteen-year trends in risk severity and operative mortality in elderly patients undergoing coronary artery bypass graft surgery. *Circulation* 1998;97:673-680

74. Ivanov J, Borger MA, David TE, Cohen G, Walton N: Predictive Accuracy Study: Comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery. *Ann Thorac Surg* 2000;70:162-168

75. Ghali WA, Quan H, Brant R: Coronary artery bypass grafting in Canada: national and provincial mortality trends, 1992-1995 [see comments]. *CMAJ* 1998;159:25-31

76. Boscarino JA, Chang J: Survival after coronary artery bypass graft surgery and community socioeconomic status: clinical and research implications. *Med Care* 1999;37:210-216

77. Harrell FE, Jr., Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. [Review]. *Statistics in Medicine* 1996;15:361-387

78. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA: Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine* 1984;3:143-152

79. Dudley RA, Harrell FE, Jr., Smith LR, Mark DB, Califf RM, Pryor DB, Glower D, Lipscomb J, Hlatky M: Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *J Clin Epidemiol* 1993;46:261-271

80. Harrell FE, Lee KL, Matchar DB, Reichart TA: Regression models for prognostic prediction: Advantages, problems and suggested solutions. *Cancer Treat Rep* 1985;69:1071-1077

81. Harrell FE, Jr., Lee KL, Pollock BG: Regression models in clinical studies: determining relationships between predictors and response. [Review]. *Journal of the National Cancer Institute* 1988;80:1198-1202

82. Hlatky MA, Califf RM, Harrell FE, Jr., Lee KL, Mark DB, Muhlbaier LH, Pryor DB: Clinical judgment and therapeutic decision making. *J Am Coll Cardiol* 1990;15:1-14

83. Diamond D: Future imperfect: the limitations of clinical prediction models and the limits of clinical prediction. *J Am Coll Cardiol* 1989;14:12A-22A

84. Marshall G, Grover FL, Henderson WG, Hammermeister KE: Assessment of predictive models for binary outcomes: an empirical approach using operative death from cardiac surgery. *Statistics in Medicine* 1994;13:1501-1511

85. Lemeshow S, Hosmer DW, Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115:92-106

86. Concato J, Feinstein AR, Holford TR: The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201-210

87. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373-1379

88. Hsieh FY: Sample size tables for logistic regression. *Statistics in Medicine* 1989;8:795-802

89. Greenland S: Modeling and variable selection in epidemiologic analysis. *Am J Public Health* 1989;79:340-349

90. Coste J, Bouyer J, Job-Spira N: Construction of composite scales for risk assessment in epidemiology: An application to ectopic pregnancy. *Am J Epidemiol* 1997;145:278-289

91. Braitman LE, Davidoff F: Predicting clinical states in individual patients. *Ann Intern Med* 1996;125:406-412

92. DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB: Comparing risk-adjustment methods for provider profiling. *Statistics in Medicine* 1997;16:2645-2664

93. Mackillop WJ, Quint CF: Measuring the accuracy of prognostic judgements in oncology. *J Clin Epidemiol* 1997;50:21-29

94. Sun GW, Shook TL, Kay GL: Inappropriate use of bivariate analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996;49:907-916

95. Charlson M, Ales KL, Simon R: Why predictive indexes perform less well in validation studies: is it magic or methods? *Arch Intern Med* 1987;147:2155-2161

96. Teres D, Lemeshow S: Why severity models should be used with caution. *Critical Care Clinics* 1994;10:93-110

97. Hosmer DW, Taber S, Lemeshow S: The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991;81:1630-1635

98. Heckerling PS, Conant RC, Tape TG, Wigton RS: Reproducibility of predictor variables from a validated clinical rule. *Med Decis Making* 1992;12:280-285

99. Poses RM, Cebul RD, Collins M, Fager SS: The importance of disease prevalence in transporting clinical prediction rules. The case of streptococcal pharyngitis. *Annals of Internal Medicine* 1986;105:586-591

100. Deyo RA, Centor RM: Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *J Chron Dis* 1986;11:897-906

101. Shahian DM, Williamson WA, Svensson LG, Restuccia JD, D'Agostino RS: Applications of statistical quality control to cardiac surgery. *Ann Thorac Surg* 1996;62:1351-8; discussion 1358-9

102. Katz D, Foxman B: How well do prediction equations predict? Using receiver operating characteristic curves and accuracy curves to compare validity and generalizability. *Epidemiology* 1993;4:319-326

103. Dawson NV: Physician judgement in clinical settings: Methodological influences and cognitive performance. *Clin Chem* 1993;39:1468-1480

104. Feinstein AR: The theory and evaluation of sensibility. in: *Clinimetrics*. New Haven and London, Yale University Press, 1987, pp 141-165

105. Hosmer DW, Lemeshow S: *Applied Logistic Regression*. New York, John Wiley & Sons, 1989

106. Tu JV, Naylor CD: Clinical prediction rules. *J Clin Epidemiol* 1997;50:743-744

107. Daly LE, Lonergan M, Graham I: Predicting operative mortality after coronary artery bypass surgery in males. *Quarterly Journal of Medicine* 1993;86:771-778

108. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36

109. Hanley JA, McNeil BJ: A method for comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839-843

110. Redelmeier DA, Bloch DA, Hickam DH: Assessing predictive accuracy: How to compare Brier Scores. *J Clin Epidemiol* 1991;44:1141-1146

111. Spiegelhalter DJ: Statistical methodology for evaluating gastrointestinal symptoms. *Clinics Gastroenterology* 1985;14:489-515

112. Goldman L, Waternaux C, Garfield F, Cohn PF, Strong R, Barry WH, Cook EF, Rosati R, Sherman H: Impact of a cardiology data bank on physicians' prognostic estimates: Evidence that cardiology fellows change their estimates to become as accurate as the faculty. *Arch Intern Med* 1981;141:1631-1634

113. Epstein ES, Murphy AH: A note on the attributes of probabilistic predictions and the probability score. *J Applied Meteorology* 1965;4:297-299

114. Poses RM, Cebul RD, Centor RM: Evaluating physicians' probabilistic judgments. *Medical Decision Making* 1988;8:233-240

115. Shroyer ALW, Grover FL, Edwards FH: 1995 Coronary artery bypass risk model: The Society of Thoracic Surgeons Adult Cardiac Surgery Database. *Ann Thorac Surg* 1998;65:879-884

116. D'Agostino RB: Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statist Med* 1998;17:2265-2281

117. Teoh KH, Christakis GT, Weisel RD, Katz AM, Tong CP, Mickleborough LL, Scully HE, Baird RJ, Goldman BS: Increased risk of urgent revascularization. *J Thorac Cardiovasc Surg* 1987;93:291-299

118. Goldman BS, Scully HE, Tong CP, Lee JY, Weisel RD, Mickleborough LL, Baird RJ: Coronary artery bypass graft surgery in the elderly. *Geriatric Cardiovasc Med* 1988;1:201-207

119. Christakis GT, Weisel RD, David TE, Salerno TA, Ivanov J, and the Cardiac Surgeons at the University of Toronto: Predictors of hospital survival after valve replacement. *Circulation* 1988;78:25-34

120. Christakis GT, Ivanov J, Weisel RD, Birnbaum PL, David TE, Salerno TA: The changing pattern of coronary artery bypass surgery. *Circulation* 1989;80(suppl I):I151-161

121. Fremes SE, Goldman BS, Ivanov J, Weisel RD, David TE, Salerno TA, and the Cardiac Surgeons at the University of Toronto: Valve surgery in the elderly. *Circulation* 1989;80:77-90

122. Ivanov J, Weisel RD: Adult cardiac surgery mortality and morbidity program. *J Am Coll Cardiol* 1990;15:271A(Abstract)

123. Fremes SE, Goldman BS, Weisel RD, Ivanov J, Christakis GT, Salerno, TA, David TE: Recent preoperative myocardial infarction increases the risk of surgery for unstable angina. *J Cardiac Surg* 1991;6:2-12

124. Fremes SE, Goldman BS, Christakis GT, Ivanov J, Weisel RD, Salerno TA, and the Cardiac Surgeons at the University of Toronto: Current risk of coronary bypass for unstable angina. *Eur J Cardio Surg* 1991;5:235-243

125. Ikonomidis JS, Weisel RD, Mouradian M, Treleaven DJ, Ivanov J, Mickleborough LL, Scully HE, Baird RJ: Thoracic aortic surgery. *Circulation* 1991;84:1-6

126. Christakis GT, Weisel RD, Fremes SE, Ivanov J, David TE, Goldman BS, Salerno TA: Coronary artery bypass grafting in patients with poor ventricular function. Cardiovascular Surgeons of the University of Toronto. *J Thorac Cardiovasc Surg* 1992;103:1083-91; discussion 1091-

127. Rao V, Christakis GT, Weisel RD, Ivanov J, Peniston CM, Ikonomidis, JS, Shirai T: Risk factors for stroke following coronary bypass surgery. *J Cardiac Surg* 1995;10:468-474

128. Rao V, Ivanov J, Weisel RD, Ikonomidis JS, Christakis GT, David TE: Predictors of low cardiac output syndrome after coronary artery bypass. *J Thorac Cardiovasc Surg* 1996;112:38-51

129. Christakis GT, Weisel RD, Buth KJ, Fremes SE, Rao V, Panagiotopoulos, KP, Ivanov J, Goldman BS, David TE: Is body size the cause for poor outcomes of coronary artery bypass operations in women? *J Thorac Cardiovasc Surg* 1995;110:1344-56; discussion 1356-

130. Borger MA, Rao V, Weisel RD, Ivanov J, Cohen G, Shiono N, Scully HE: Deep sternal wound infection: risk factors and treatment. *Ann Thorac Surg* 1998;65:1050-1056

131. Gordon RS, Ivanov J, Cohen G, Ralph-Edwards AC: Permanent cardiac pacing after a cardiac operation: Predicting the use of permanent pacemakers. *Ann Thorac Surg* 1998;66:1698-1704

132. Borger MA, Ivanov J, Weisel RD, Peniston CM, Rambaldini G, Rao V, Cohen G, Mickleborough LL, Feindel CM, David TE: Decreasing incidence of stroke during valvular surgery. *Circulation* 1998;98:37-43

133. Ralph-Edwards AC, Robinson AG, Gordon RS, Ivanov J: Valve surgery in octogenarians. *Can J Cardiol* 1999;15:1113-1119

134. Yau TM, El-Ghoneimi YA, Armstrong S, Ivanov J, David TE: Mitral valve repair and replacement for rheumatic disease. *J Thorac Cardiovasc Surg* 2000;119:53-61

135. Mickleborough LL, Carson S, Tamariz M, Ivanov J: Results of revascularization in patients with severe left ventricular dysfunction. *J Thorac Cardiovasc Surg* 2000;119:550-557

136. Gann D, Colin C, Hildner FJ, Samet P, Yahr WZ, Greenberg JJ: Coronary artery bypass surgery in patients seventy years of age and older. *J Thorac Cardiovasc Surg* 1977;73(2):237-241

137. Morgan CD, Sykora K, Naylor CD: Analysis of deaths while waiting for cardiac surgery among 29, 293 consecutive patients in Ontario, Canada. The Steering Committee of the Cardiac Care Network of Ontario. *Heart* 1998;79:345-349

138. Naylor CD, Szalai JP, Katic M: Benchmarking the vital risk of waiting for coronary artery bypass surgery in Ontario. *CMAJ* 2000;162:775-779

139. Naylor CD, Rothwell DM, Tu JV, Austin P, and the Cardiac Care Network Steering Committee: Outcomes of Coronary Artery Bypass Surgery in Ontario. in Naylor CD, Slaughter PM (eds): *Cardiovascular Health and Services in Ontario: An ICES Atlas.* Toronto, Institute for Clinical Evaluative Sciences, 1999, pp 189-198

140. Carey JS, Dziuban SW, Arom KV, Cimochowskl GE, Plume SK, Grover FL: Quality improvement in thoracic surgery. *Bulletin of the American College of Surgeons* 1998;83:24-29

141. Green J, Wintfeld N: Report cards on cardiac surgeons. Assessing New York State's approach [see comments]. *N Engl J Med* 1995;332:1229-1232

142. Tu JV, Naylor CD: Coronary artery bypass mortality rates in Ontario. A Canadian approach to quality assurance in cardiac surgery. Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. *Circulation* 1996;94:2429-2433

143. Gabrielle F, Roques F, Michel P, Bernard A, de Vicentis C, Roques X, Brenon R, Bauder E, David M: Is the Parsonnet's score a good predictive score of mortality in adult cardiac surgery: assessment by a French multicentre study. *Eur J Cardio Surg* 1997;11:406-414

144. Boyd WD, Desai ND, Del Rizzo DF, Novick RJ, McKenzie FN, Menkis AH: Off-pump surgery decreases postoperative complications and resource utilization in the elderly. *Ann Thorac Surg* 1999;68:1490-1493

145. Sharma AC, Fogelson BG, Nawas SI, Vigneswaran WT, Sam ADI, Alden KJ, Ferguson JL, Law WR: Elevated coronary endothelin-1 but not nitric oxide in diabetes during CABG. *Ann Thorac Surg* 1999;67:1659-1663

146. Bridgewater B, Steyn RS, Ray S, Hooper T: Minimally invasive aortic valve replacement through a transverse sternotomy: a word of caution. *Heart* 1998;79:605-607

147. O'Connor GT, Plume SK, Olmstead EM, Morton JR, Maloney CT, Nugent WC, Hernandez F, Jr., Clough R, Leavitt BJ, Coffin LH, Marrin CA, Wennberg, D, Birkmeyer JD, Charlesworth DC, Malenka DJ, Quinton HB, Kasper JF: A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group [see comments]. *JAMA* 1996;275:841-846

148. Christakis GT, Lichtenstein SV, Buth KJ, Fremes SE, Weisel RD, Naylor CD: The influence of risk factors on the results of warm heart surgery: a substudy of a randomized trial. *Eur J Cardiothorac Surg* 1997;11:515-520

149. Luft HS, Romano PS: Chance, continuity and change in hospital mortality rates. *JAMA* 1993;270:331-337

150. Silber JH, Rosenbaum PR, Schwartz S, Ross RN, Williams SV: Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. *JAMA* 1995;274:317-323

151. *SAS Language Guide for Personal Computers.* Cary, NC, SAS Institute Inc. 1988

152. Dixon WJ (ed): *BMDP Statistical Software Manual.* Berkeley, CA, University of California Press, 1992

153. Arora R, Sager J, Butler RN: Therapeutic goals: quality of life. *Cardiol Clin* 1986;4(2):305-312

154. Guralnik JM, Fitzsimmons SC: Aging in America: A demographic perspective. *Cardiol Clinics* 1986;4(2):175-183

155. Rowe JW: Health care of the elderly. *N Engl J Med* 1985;312(13):827-835

156. Naylor CD, Ugnat AM, Weinkauf D, Anderson GM, Wielgosz A: Coronary artery bypass grafting in Canada: What is its rate of use? Which rate is right? *Can Med Assoc J* 1992;146(6):851-859

157. MacArthur AE, Hall RJ, Gray AG, Mathur VS, Cooley DA: Coronary revascularization in the elderly patient. *JACC* 1984;3(6):1398-1402

158. Freeman WK, Schaff HV, O'Brien PC, Orszulak TA, Naessens JM, Tajik AJ: Cardiac surgery in the octogenarian: Perioperative outcome and clinical follow-up. *JACC* 1991;18(1):29-35

159. Knapp WS, Douglas JS Jr, Craver JM, Jones EL, King SB III, Bone DK, Bradford JM, Hatcher CR Jr: Efficacy of coronary artery bypass grafting in elderly patients with coronary artery disease. *Am J Cardiol* 1981;47:923-930

160. Gersh BJ, Kronmal RA, Frye RL, Schaff HV, Ryan TJ, Gosselin AJ, Kaiser GC, Killip T: Coronary arteriography and coronary artery bypass: Morbidity and mortality in patients ages 65 years or older. *Circulation* 1983;67(3):483-491

161. Hibler BA, Wright JO, Wright CB, Ehrenhaft JL, Doty DB, Rossi NP: Coronary artery bypass surgery in the elderly. *Arch Surg* 1983;118:402-404

162. Hochberg MS, Levine FH, Daggett WM, Akins CW, Austen WG, Buckley MJ: Isolated coronary artery bypass grafting in patients seventy years of age and older. *J Thorac Cardiovasc Surg* 1982;84:219-223

163. Horneffer PJ, Gardner TJ, Manolio TA, Hoff SJ, Rykiel MF, Pearson TA, Gott VL, Baumgartner WA, Borkon AM, Watkins L, Reitz BA: The effects of age on outcome after coronary bypass surgery. *Circulation* 1987;76 (Suppl V):V6-12

164. Loop FD, Lytle BW, Cosgrove DM, Goormastic M, Taylor PC, Golding LAR, Stewart RW, Gill CC: Coronary artery bypass graft surgery in the elderly. *Cleve Clin J Med* 1988;55:23-34

165. Horvath KA, DiSesa VJ, Peigh PS, Couper GS, Collins JJ, Cohn LH: Favorable results of coronary artery bypass grafting in patients older than 75 years. *J Thorac Cardiovasc Surg* 1990;99:92-96

166. Ko W, Krieger KH, Lazenby WD, Shin YT, Goldstein M, Lazzaro R, Isom OW: Isolated coronary bypass grafting in one hundred consecutive otogenarian patients. *J Thorac Cardiovasc Surg* 1991;102:532-538

167. Tsai TP, Nessim S, Kass RM, Chaux A, Gray RJ, Khan SS, Blanche C, Utley C, Matloff JM: Morbidity and mortality after coronary bypass in octogenarians. *Ann Thorac Surg* 1991;51:983-986

168. Hannan EL, Burke J: Effect of age on mortality in coronary artery bypass surgery in New York, 1991-1992. *Am Heart J* 1994;128:1184-1191

169. Curtis JJ, Walls JT, Boley TM, Schmaltz RA, Demmy TL, Salam N: Coronary revascularization in the elderly: determinants of operative mortality. *Ann Thorac Surg* 1994;58:1069-1072

170. He GW, Acuff TE, Ryan WH, Bowman RT, Douthit MB, Mack MJ: Determinants of operative mortality in elderly patients undergoing coronary artery bypass grafting. Emphasis on the influence of internal mammary artery grafting on mortality and morbidity. *J Thorac Cardiovasc Surg* 1994;108:73-81

171. Katz NM, Hannan RL, Hopkins RA, Wallace RB: Cardiac operations in patients aged 70 years and over: mortality, length of stay, and hospital charge. *Ann Thorac Surg* 1995;60:96-100

172. Morris RJ, Strong MD, Grunewald KE, Kuretu ML, Samuels LE, Kresh JY, Brockman SK: Internal thoracic artery for coronary artery grafting in octogenarians. *Ann Thorac Surg* 1996;62:16-22

173. Tsai TP, Matloff JM, Gray RJ, Chaux A, Kass RM, Lee ME, Czer LSC: Cardiac surgery in the octogenarian. *J Thorac Cardiovasc Surg* 1986;91:924-928

174. Pryor DB, Harrell FE, Jr., Rankin JS, Lee KL, Muhlbaier LH, Oldham HN, Hlatky MA, Mark DB, Reves JG, Califf RM: The changing survival benefits of coronary revascularization over time. *Circulation* 1987;76:V13-21

175. Goel V, Williams JI, Anderson GM, Blackstein-Hirsh P, Fooks C, Naylor CD: Naylor CD (ed): *Patterns of Health Care in Ontario: The ICES Practice Atlas.* Toronto, Canadian Medical Association, 1997, pp 99-104

176. Tu JV, Naylor CD, Kumar D, DeBuono BA, McNeil BJ, Hannan EL: Coronary artery bypass graft surgery in Ontario and New York State: which rate is right? Steering Committee of the Cardiac Care Network of Ontario. *Ann Intern Med* 1997;126:13-19

177. Stason WB, Sanders CA, Smith HC: Cardiovascular care of the elderly: Economic considerations. *JACC* 1987;10(2):18-21

178. Gersh BJ, Kronmal RA, Schaff HV, Frye RL, Ryan TJ, Myers WO, Athearn MW, Gosselin AJ, Kaiser GC, Killip T: Long-term (5 year) results of coronary bypass surgery in patients 65 years old or older: A report from the Coronary Artery Surgery Study. *Circulation* 1983;68(suppl II):II190-199

179. Cheitlin MD: Coronary bypass surgery in the elderly. *Clinics in Geriatric Medicine* 1996;12:195-205

180. Cane ME, Chen C, Bailey BM, Fernandez J, Laub GW, Anderson WA, McGrath LB: CABG in octogenarians: early and late events and actuarial survival in comparison with a matched population. *Ann Thorac Surg* 1995;60:1033-1037

181. Tu JV, Naylor CD, the Steering Committee of the Provincial Adult Cardiac Care Network of Ontario: Coronary artery bypass mortality rates in Ontario. A Canadian approach to quality assurance in cardiac surgery. *Circulation* 1996;94:2429-2433

182. The Warm Heart Investigators: Normothermic versus hypothermic cardiac surgery: A randomized trial in 1732 coronary bypass patients. *Lancet* 1994;343:559-563

183. Dustan HP, Hamilton MP, McCullough L: Sociopolitical, ethical considerations in the treatment of cardiovascular disease in the elderly. *JACC* 1987;10(2):14A-17A

184. Naylor CD, Levinton CM, Wheeler S, Hunter L: Queueing for coronary surgery during severe supply-demand mismatch in a Canadian referral centre: A case study of implicit rationing. *Soc Sci Med* 1993;37(1):61-67

185. Poses RM, Bekes C, Winkler RL, Scott WE, Copare FJ: Are two (inexperienced) heads better than one (experienced) head? Averaging house officers' prognostic judgments for critically ill patients. *Archives of Internal Medicine* 1990;150:1874-1878

186. Cheng DC, Karski J, Peniston C, Asokumar B, Raveendran G, Carroll J, Nierenberg H, Roger S, Mickle D, Tong J, Zelovitsky J, David T, Sandler A: Morbidity outcome in early versus conventional tracheal extubation after coronary artery bypass grafting: a prospective randomized controlled trial. *J Thorac Cardiovasc Surg* 1996;112:755-764

187. Cheng DC, Karski J, Peniston C, Raveendran G, Asokumar B, Carroll J, David T, Sandler A: Early tracheal extubation after coronary artery bypass graft surgery reduces costs and improves resource use. A prospective, randomized, controlled trial. *Anesthesiology* 1996;85:1300-1310

188. Lee KL, Pryor DB, Harrell FE, Jr., Califf RM, Behar VS, Floyd WL, Morris JJ, Waugh RA, Whalen RE, Rosati RA: Predicting outcome in coronary disease. Statistical models versus expert clinicians. *American Journal of Medicine* 1986;80:553-560

189. Dolan JG, Bordley DR, Mushlin AI: An evaluation of clinicians' subjective prior probability estimates. *Medical Decision Making* 1986;6:216-223

190. Poses RM, Bekes C, Copare FJ, Scott WE: What difference do two days make? The inertia of physicians' sequential prognostic judgments for critically ill patients. *Medical Decision Making* 1990;10:6-14

191. McClish DK, Powell SH: How well can physicians estimate mortality in a medical intensive care unit? *Medical Decision Making* 1989;9:125-132

192. Poses RM, Bekes C, Copare FJ, Scott WE: The answer to "What are my chances, doctor?" depends on whom is asked: prognostic disagreement and inaccuracy for critically ill patients. *Crit Care Med* 1989;17:827-833

193. Detsky AS, Stricker SC, Mulley AG, Thibault GE: Prognosis, survival, and the expenditure of hospital resources for patients in an intensive-care unit. *N Engl J Med* 1981;305:667-672

194. Myers DG, Lamm H: The group polarization phenomenon. *Psychological Bulletin* 1976;83:602-627

195. Dawes RM, Faust D, Meehl PE: Clinical versus actuarial judgment. *Science* 1989;243:1668-1674

196. Oxman AD, Thomson MA, Davis DA, Haynes RB: No magic bullets: A systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995;153:1423-1431

197. Davis DA, Thomson MA, Oxman AD, Haynes RB: Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA* 1995;274:700-705

198. Favaloro RG: Saphenous vein graft in the surgical treatment of coronary artery disease: operative technique. *J Thorac Cardiovasc Surg* 1969;58:178-185

199. Jones RH, Kesler K, Phillips HR, 3d, Mark DB, Smith PK, Nelson CL, Newman MF, Reves JG, Anderson RW, Califf RM: Long-term survival benefits of coronary artery bypass grafting and percutaneous transluminal angioplasty in patients with coronary artery disease. *J Thorac Cardiovasc Surg* 1996;111:1013-1025

200. Lawrie GM, Morris GC, Jr., Baron A, Norton J, Glaeser DH: Determinants of survival 10 to 14 years after coronary bypass: analysis of preoperative variables in 1,448 patients. *Ann Thorac Surg* 1987;44:180-185

201. Califf RM, Harrell FE, Jr., Lee KL, Rankin JS, Mark DB, Hlatky MA, Muhlbaier LH, Wechsler AS, Jones RH, Oldham HN, Jr., et al. Changing efficacy of coronary revascularization. Implications for patient selection. *Circulation* 1988;78:I185-91

202. Muhlbaier LH, Pryor DB, Rankin JS, Smith LR, Mark DB, Jones RH, Glower DD, Harrell FE, Jr., Lee KL, Califf RM, et al. Observational comparison of event-free survival with medical and surgical therapy in patients with coronary artery disease. 20 years of follow-up. *Circulation* 1992;86:II198-204

203. Rahimtoola SH, Fessler CL, Grunkemeier GL, Starr A: Survival 15 to 20 years after coronary bypass surgery for angina. *J Am Coll Cardiol* 1993;21:151-157

204. Rahimtoola SH, Bennett AJ, Grunkemeier GL, Block P, Starr A: Survival at 15 to 18 years after coronary bypass surgery for angina in women. *Circulation* 1993;88:II71-8

205. Weintraub WS, Jones EL, Craver JM, Guyton RA: Frequency of repeat coronary bypass or coronary angioplasty after coronary artery bypass surgery using saphenous venous grafts. *Am J Cardiol* 1994;73:103-112

206. Davis KB, Chaitman B, Ryan T, Bittner V, Kennedy JW: Comparison of 15-year survival for men and women after initial medical or surgical treatment for coronary artery disease: a CASS registry study. Coronary Artery Surgery Study [see comments]. *J Am Coll Cardiol* 1995;25:1000-1009

207. Weintraub WS, Jones EL, Craver JM, Grosswald R, Guyton RA: In-hospital and long-term outcome after reoperative coronary artery bypass graft surgery. *Circulation* 1995;92:II50-7

208. Cameron AA, Davis KB, Rogers WJ: Recurrence of angina after coronary artery bypass surgery: predictors and prognosis (CASS Registry). Coronary Artery Surgery Study. *J Am Coll Cardiol* 1995;26:895-899

209. van Brussel BL, Plokker HW, Voors AA, Ernst JM, Knaepen PJ, Koomen EM, Tijssen JG, Vermeulen FE: Multivariate risk factor analysis of clinical outcome 15 years after venous coronary artery bypass graft surgery. *European Heart Journal* 1995;16:1200-1206

210. Peduzzi P, Kamina A, Detre K: Twenty-two-year follow-up in the VA Cooperative Study of Coronary Artery Bypass Surgery for Stable Angina. *Am J Cardiol* 1998;81:1393-1399

211. Gray RJ, Nessim S, Khan SS, Denton T, Matloff JM: Adverse 5-year outcome after coronary artery bypass surgery in blacks. *Archives of Internal Medicine* 1996;156:769-773

212. Hammar N, Sandberg E, Larsen FF, Ivert T: Comparison of early and late mortality in men and women after isolated coronary artery bypass graft surgery in Stockholm, Sweden, 1980 to 1989. *J Am Coll Cardiol* 1997;29:659-664

213. Barsness GW, Peterson ED, Ohman EM, Nelson CL, DeLong ER, Reves JG, Smith PK, Anderson RD, Jones RH, Mark DB, Califf RM: Relationship between diabetes mellitus and long-term survival after coronary bypass and angioplasty. *Circulation* 1997;96:2551-2556

214. Weintraub WS, Stein B, Kosinski A, Douglas JS, Jr., Ghazzal ZM, Jones, EL, Morris DC, Guyton RA, Craver JM, King SB, 3rd. Outcome of coronary bypass surgery versus coronary angioplasty in diabetic patients with multivessel coronary artery disease [see comments]. *J Am Coll Cardiol* 1998;31:10-19

215. Buxton BF, Komeda M, Fuller JA, Gordon I: Bilateral internal thoracic artery grafting may improve outcome of coronary artery surgery. Risk-adjusted survival. *Circulation* 1998;98:II1-6

216. Sergeant P, Blackstone E, Meyns B, Stockman B, Jashari R: First cardiological or cardiosurgical reintervention for ischemic heart disease after primary coronary artery bypass grafting. *European Journal of Cardio-Thoracic Surgery* 1998;14:480-487

217. Sergeant P, Blackstone E, Meyns B: Can the outcome of coronary bypass grafting be predicted reliably? *European Journal of Cardio-Thoracic Surgery* 1997;11:2-9

218. Kirklin JK, and the ACC/AHA Task Force: Guidelines and indications for coronary artery bypass graft surgery. A report of the American College of Cardiology/American Heart Association Task Force on assessment of diagnostic and therapeutic cardiovascular procedures. *JACC* 1991;17:543-589

219. Tu JV, Austin P, Naylor CD, Iron K, Zhang H: Acute Myocardial Infarction Outcomes In Ontario. in Naylor CD, Slaughter PM (eds): *Cardiovascular Health and Services in Ontario: An ICES Atlas.* Toronto, Institute for Clinical Evaluative Sciences, 1999, pp 83-110

220. Alter DA, Austin P, Tu JV: Use of Coronary Angiography, Angioplasty and Bypass Surgery After Acute Myocardial Infarction in Ontario. in Naylor CD, Slaughter PM (eds): *Cardiovascular Health and Services in Ontario: An ICES Atlas.* Toronto, Institute for Clinical Evaluative Sciences, 1999, pp 141-164

221. Iezzoni LI, Ash AS, Shwartz M, Landon BE, Mackiernan YD: Predicting in-hospital deaths from coronary artery bypass graft surgery. Do different severity measures give different predictions? [see comments]. *Med Care* 1998;36:28-39

222. Smith LR, Harrell FE, Rankin JS, Califf RM, Pryor DB, Muhlbaier LH, Lee KL, Mark DB, Jones RH, Oldham HN, Glower DD, Reves JG, Sabiston DC Jr: Determinants of early versus late cardiac death in patients undergoing coronary artery bypass graft surgery. *Circulation* 1991;84(Suppl III):III245-253

223. Naylor CD, Levinton CM, Baigrie RS: Adapting to waiting lists for coronary revascularization. Do Canadian specialists agree on which patients come first? *Chest* 1992;101:715-722

224. Krumholz HM: Mathematical models and the assessment of performance in cardiology. [Editorial]. *Circulation* 1999;99:2067-2069

225. Edmunds LH, Clark RE, Cohn LH, Grunkemeier GL, Miller DC, Weisel RD: Guidelines for reporting morbidity and mortality after cardiac valvular operations. *J Thorac Cardiovasc Surg* 1996;112:708-711

UNIVERSITY HEALTH NETWORK - DIVISION OF CARDIOVASCULAR SURGERY 2000

ADMISSION TO HOSPITAL: _____/_____/_____
DATE OF OPERATION: _____/_____/_____
DATE OF DISCHARGE: _____/_____/_____
DAYS IN ICU: _____
MINUTES ON VENT: _____
MINUTES IN ICU: _____
SURGEON: _____ 1st ASSIST: _____ A
B

APPENDIX 1

Date Accepted for Surgery: _____/_____/_____

FAMILY DR.: _____

CARDIOLOGIST: _____

Triage Case	Y	N
Same Day Admit	Y	N
Readmission to ICU	Y	N

GENERAL PATIENT DATA

TIMING:

( ) 1 = Elective

( ) 2 = Same Hospitalization
( ) 3 = Urgent <72 hrs from event
( ) 4 = Emergent <12 hrs from event

NON-ELECTIVE SURGERY FROM:

( ) 1 = Floor
( ) 2 = CCU
( ) 3 = Cath Lab
( ) 5 = Other: _____

PREVIOUS HEART SURGERY:

_____

Date: ___/___/____

Place: _____

# of Previous Ops: ____

PREVIOUS INTERVENTION:

( ) 0 = None
( ) 1 = Angioplasty Date: ___/___/___
( ) 2 = Valvotomy Date: ___/___/___
( ) 3 = Thrombolysis Date: ___/___/___

HEART CATHETERIZATION:

( ) TTH
( ) Scarborough Centenary
( ) Sunnybrook
( ) St. Michael's
( ) Mount Sinai
( ) Other
_____

ANGINA PECTORIS:

( ) 0 = None

( ) 1 = Stable

( ) 2 = Crescendo
( ) 3 = A.C.I.

M.I. < 30 DAYS:
( ) 0 = No
( ) 1 = Non-Q Wave
( ) 2 = Transmural

DATE OF M.I.: ____/____/____

N.Y.H.A.:

( ) I
( ) II
( ) III
( ) IV

L.V. EJECTION FRACTION:

( ) 1 = ≥ 60%
( ) 2 = 40-59%
( ) 3 = 20-39%
( ) 4 = < 20%

STRESS TEST:

( ) 0 = Not done

( ) 1 = Negative
( ) 2 = Positive

C.A.D. RISKS:

	YES	NO
Diabetes	( )	( )

( ) 1 = Diet ( ) 2 = Oral ( ) 3 = Insulin

Hypertension	( )	( )
Hyperlipid	( )	( )
Family History	( )	( )

SMOKING:

( ) 0 = Never smoked
( ) 1 = Stopped smoking
( ) 2 = Still smoking

ASSOC. DISEASES:

	YES	NO
C.O.P.D. severe	( )	( )
Stroke / TIA	( )	( )
P.V.D.	( )	( )
Dialysis	( )	( )
Marfan	( )	( )

ELECTROCARDIOGRAM:
( ) 0 = Normal Sinus Rhythm
( ) 1 = A. Fib. / Flutter
( ) 2 = CHB / Pacer

OTHER SYMPTOMS:

	YES	NO
Heart Failure	( )	( )
Shock	( )	( )
Syncope	( )	( )

PREOPERATIVE:	YES	NO
Aspirin < 7 days	( )	( )

Date: _____/_____/_____

Creatinine level _____

VALVE PATIENT DATA

VALVULAR LESIONS:

( ) AS ( ) AI _____
( ) MS ( ) MI _____
( ) TS ( ) TI _____

INFECTIVE ENDOCARDITIS:

( ) 0 = No
( ) 1 = Remote
( ) 2 = Active
( ) 3 = Active abscess

REASON FOR URGENT SURGERY:

( ) 0 = Not Urgent
( ) 1 = Acute endocarditis
( ) 2 = Acute pump failure
( ) 3 = Acute prosthetic failure
( ) 4 = Failed balloon valvuloplasty
( ) 5 = Life threatening arrhythmia
( ) 6 = Ischemic mitral regurgitation
( ) 7 = Dissection

AORTIC VALVE SURGERY:

( ) 0 = None
( ) 1 = Repair
( ) 2 = Replacement
( ) 3 = Repair of an in-situ prosthetic valve

( ) D1 = reimplantation ( ) 1 = STJ
( ) D2 = remodelling ( ) 2 = 1 Sinus
( ) 3 = 2 Sinuses
( ) 4 = 3 Sinuses

AV PROSTHESIS:
TYPE: _____
SIZE: _____ (mm)

X PLANTED VALVE PATHOLOGY:

( ) 1 = Thrombosed
( ) 2 = Calcified
( ) 3 = Cusp Tear
( ) 4 = Infected
( ) 5 = Mechanical Dysfunction
( ) 6 = Other _____

AV PATHOLOGY: Check all that apply

( ) 1 = Rheumatic
( ) 2 = Tricuspid, calcific
( ) 3 = Bicuspid
( ) 4 = Congenital, other
( ) 5 = Aortoannular ectasia
( ) 6 = Prosthetic dysfunction
( ) 7 = Dissection
( ) 8 = Other: _____

AV ANNULUS ENLARGED:

( ) 0 = No
( ) 1 = Yes, annular
( ) 2 = Yes, sinuses
( ) 3 = Yes, both

X PLANTED VALVES:

TYPE: _____
SIZE: _____ (mm)

MITRAL VALVE SURGERY:

( ) 0 = None
( ) 1 = Repair
( ) 2 = Replacement
( ) 3 = Repair of an in-situ prosthetic valve

MV PROSTHESIS:

TYPE: _____
SIZE: _____ (mm)

MV ANN. RECONSTRUCTION:
( ) 0 = No
( ) 1 = Yes, autologous
( ) 2 = Yes, bovine
( ) 3 = Yes, Dacron

X PLANTED VALVE PATHOLOGY:

( ) 1 = Thrombosed
( ) 2 = Calcified
( ) 3 = Cusp Tear
( ) 4 = Infected
( ) 5 = Mechanical Dysfunction
( ) 6 = Other _____

MV PATHOLOGY: Check all that apply

( ) 1 = Rheumatic
( ) 2 = Ischemic
( ) 3 = Myxomatous
( ) 4 = Prosthetic dysfunction
( ) 5 = Annular dilatation
( ) 6 = Other: _____

CHORDAE PRESERVED:
( ) 0 = No
( ) 1 = Yes, posterior only
( ) 2 = Yes, posterior and anterior

Goretex sutures: ( ) YES ( ) NO

X PLANTED VALVES:

TYPE: _____
SIZE: _____ (mm)

TRICUSPID VALVE SURGERY:

( ) 0 = None
( ) 1 = Repair
( ) 2 = Replacement
( ) 3 = Repair of an in-situ prosthetic valve

TV PROSTHESIS:

TYPE: _____
SIZE: _____(mm)

TV PATHOLOGY: Check all that apply

( ) 1 = Rheumatic
( ) 2 = Secondary
( ) 3 = Congenital
( ) 4 = Prosthetic dysfunction
( ) 5 = Other: _____

X PLANTED VALVE PATHOLOGY:

( ) 1 = Thrombosed
( ) 2 = Calcified
( ) 3 = Cusp Tear
( ) 4 = Infected
( ) 5 = Mechanical Dysfunction
( ) 6 = Other _____

X PLANTED VALVES:

TYPE: _____
SIZE: _____(mm)

ASSOCIATED PROCEDURES:
(non-cardiac)

( ) 0 = No
( ) 1 = Abdominal Aortic Surgery
( ) 2 = Carotid Endarterectomy
( ) 3 = Other: _____

OTHER CARDIAC OPERATIONS

LEFT VENTRICULAR ANEURYSM:

( ) 0 = No
( ) 1 = Anterior
( ) 2 = Posterior
( ) 3 = False Aneurysm

L.V.A. Patch: ( ) YES ( ) NO

Location: _____

Material: _____

CONGENITAL HEART DISEASE:

( ) 0 = No
( ) 1 = ASD
( ) 2 = VSD
( ) 3 = Both
( ) 4 = Other: (NOTES)

ASC. AORTA REPLACED:

( ) 0 = No
( ) 1 = Yes, separate
( ) 2 = Yes, composite

OTHER PROCEDURES:

( ) 0 = No
( ) 1 = Ischemic VSD - Patch
( ) 2 = Myxoma
( ) 3 = Myectomy
( ) 4 = Mapping / Arrhythmia
( ) 5 = Heart Transplant
( ) 9 = Maze Procedure
( ) 8 = Other:

NOTES

ACB PATIENT DATA

CORONARY STENOSIS ≥ 50%:

( ) LAD ( ) INT ( ) RCA
( ) DIA1 ( ) OM1 ( ) PIV
( ) DIA2 ( ) OM2 ( ) PL
( ) PER ( ) CX ( ) AM

( ) Left Main
NO

LITA: ( ) YES:_____ ( ) NO
NO
RITA: ( ) YES:_____ ( ) NO

RADIAL ARTERY: ( ) YES:_____ ( ) NO

ARTERIES GRAFTED:

( ) LAD ( ) INT ( ) RCA
( ) DIA1 ( ) OM1 ( ) PIV
( ) DIA2 ( ) OM2 ( ) PL
( ) PER ( ) CX ( ) AM

Endarterectomy : ( ) YES:_____ ( )

Coronary Vein : ( ) YES:_____ ( )

Patch Angioplasty

Other conduit: _____

GENERAL OPERATIVE DATA

PUMP CASE ( ) YES ( ) NO ( ) STANDBY

MINIMALLY INVASIVE SURGERY ( ) YES ( ) NO

OR TIME:_____ min.

BYPASS TIME: _____ min.

X-CLAMP TIME: _____ min.

CIR. ARREST TIME:_____ min.

B.S.A.:_____ m^2

HEIGHT:_____ cm

WEIGHT:_____ kg

MYOCARDIAL PROTECTION:

Primary form of myocardial protection:

( ) 0 = None, coronary perfusion
( ) 1 = Warm blood cardioplegia (35-38°C)
( ) 2 = Tepid blood cardioplegia (20-35°C)
( ) 3 = Cold blood cardioplegia (< 20°C)
( ) 4 = Crystalloid cardioplegia

TECHNIQUE Method of Cardioplegia:

( ) 1 = Intermittent
( ) 2 = Continuous
( ) 3 = Both

DIRECTION of Cardioplegia:
( ) 1 = Antegrade
( ) 2 = Retrograde
( ) 3 = Both

HYPOTHERMIA:

( ) 0 = Normothermia
( ) 1 = Mild (30-35°C)
( ) 2 = Mod (20-29°C)
( ) 3 = Profound (≤ 20°C)

OFF C.P.B.: # 1

( ) 0 = Well
( ) 1 = Inotropes / Balloon
( ) 2 = OR Death

COMPLICATIONS

IABP:

( ) 0 = No
( ) 1 = Preop CCU
( ) 2 = Preop OR
( ) 3 = OR
( ) 4 = Postop ICU

REOPENING:

( ) 0 = No
( ) 1 = Bleeding
( ) 2 = Tamponade
( ) 3 = Shock / arrest
( ) 4 = Infection
( ) 5 = Dehiscence
( ) 6 = Redo Surgery
( ) 7 = Other: _____

POSTOP E.C.G.:

( ) 0 = No change
( ) 1 = New ST/T waves
( ) 2 = New Q waves
( ) 3 = Complete LBBB
( ) 4 = Loss of R waves

Perioperative M.I.: ( ) YES ( ) NO

Inotropes: ( ) YES ( ) NO

LOS: ( ) YES ( ) NO

Renal Failure: ( ) YES ( ) NO

Pacemaker: ( ) YES ( ) NO

Vent. Dysrhyth: ( ) YES ( ) NO

A Fib. ( ) YES ( ) NO

D.V.T.: ( ) YES ( ) NO

Pulm. Complic's: ( ) YES ( ) NO

Seizures: ( ) YES ( ) NO

TIA: ( ) YES ( ) NO

Hb: pre _____

post _____

Blood _____ units

STROKE:

( ) 0 = No
( ) 1 = Intra-operative
( ) 2 = Post-operative

INFECTION:

Leg ( ) 0 = No
( ) 1 = Superficial
( ) 2 = Deep

Sternum ( ) 0 = No
( ) 1 = Superficial
( ) 2 = Deep

Arm ( ) YES ( ) NO

Sepsis ( ) YES ( ) NO

HOSPITAL SURVIVOR:
( ) YES ( ) NO

D/C TO: H C O D

PROCEDURE CODE:

( ) 1 ( ) 2 ( ) 3

Appendix 2

NOTE: Fill out forms on all PUMP AND NON-PUMP cases.

The Toronto General Hospital
Division of Cardiovascular Surgery
CODE BOOK TGH (MS Access)

Printed on: March 6, 2001

DEMOGRAPHICS:

PTID Patient identification. First three letters of last name, first letter of first name, and date of birth (LLLFMMDDYY)
(NOTE: See last page for conventions)

CHART Hospital chart number

LNAME Last name of patient

FNAME (>12/31/97) First name of patient

AGE Age on date of surgery (yrs)

SEX Sex of patient:
1 - Male
2 - Female

AREA (>12/31/93) Resident location of patient
0 - Ontario
1 - Out of province
2 - Out of country

TRIAGE (>12/31/96) Non-elective admission (excluding heart transplants) (YES/NO)

SDA (>12/31/96) Patient seen in same-day admit clinic and admitted to hospital the day of surgery (YES/NO)

DATES & DOCTORS:

ADDATE Admission to hospital (MM/DD/YY)

DATEOR Date of operation (MM/DD/YY)

DISDATE Date of discharge from hospital (MM/DD/YY)

DAYSPOST	Post-op days in hospital (excluding OR day)
ICUNUM (>12/31/98)	Number of admissions to ICU
ICU	Total number of days in ICU (excluding OR day)
ICUTIME (>06/05/94)	Total number of minutes in ICU (from arrival in ICU till departure to floor)
VENT (<07/01/93)	Days on ventilator (excluding OR day)
VENTTIME (>06/30/93)	Minutes on ventilator (from arrival in ICU till extubation time)
SURG	Surgeon's initials
ASSIST	First assistant (senior resident or other)
FDOC	Patient's family doctor
CDOC	Outside cardiologist following patient post-op

GENERAL PATIENT DATA :

TIMING Timing of surgery - approximate length of time between the event resulting in hospitalization and surgery

1 - Elective
2 - Same hospitalization: cannot go home because of urgency or anatomy (semi-urgent)
3 - Urgent: < 72 hrs from event
4 - Emergency: < 12 hrs from event

FROM Pre-op location of non-elective patients

1 - Floor
2 - CCU
3 - Cath Lab
4 - Emergency (<01/01/98)
5 - Other

PREVIOUS HEART SURGERY:

ACBREDO Previous ACB (YES/NO)

AVREDO Previous aortic valve surgery
- 0 - No
- 1 - Repair
- 2 - Replacement
- 3 - Repair of an in-situ prosthetic valve

MVREDO Previous mitral valve surgery
- 0 - No
- 1 - Repair
- 2 - Replacement
- 3 - Repair of an in-situ prosthetic valve

TVREDO Previous tricuspid valve surgery
- 0 - No
- 1 - Repair
- 2 - Replacement
- 3 - Repair of an in-situ prosthetic valve

OTHREDO Any other previous cardiac surgery (YES/NO)

PREVIOUS (NON-SURGICAL) INTERVENTION:

PRECARD Previous non-surgical intervention
- 1 - Angioplasty
- 2 – Balloon valvotomy

PIDATE Date of previous intervention (MM/DD/YY)

PITHROMB Previous thrombolysis (YES/NO)

PITDATE Date of previous thrombolysis (MM/DD/YY)

HEART CATHETERIZATION:

CATH Hospital where patient was catheterized
- TGH - The Toronto Hospital, General Division. (<01/01/96)
- TWH - The Toronto Hospital, Western Division. (<01/01/96)
- TTH - The Toronto Hospital (>12/31/95)
- SMC - Sunnybrook Health Science Centre
- SMH - St. Michael's Hospital
- SCH - Scarborough Centenary Hospital

MSH - Mount Sinai Hospital
OTH - Other

CATHDATE Date of catheterization (MM/DD/YY)

CLINICAL PRESENTATION:

ANGINA Most severe angina pectoris within one month prior to surgery
0 - None
1 - Stable: predictable exertional angina
2 - Crescendo: increasing frequency or severity of symptoms
3 - Acute coronary insufficiency: prolonged episodes of unprovoked pain (>15 mins) despite medical therapy

PREOPMI Most recent myocardial infarction within 30 days of OR date
0 - None
1 - Non-Q wave infarction
2 - Transmural infarction

MIDATE Date of most recent MI (MM/DD/YY)

NYHA NYHA classification (cardiac disability)
1 - No restrictions
2 - Symptoms provoked by exertion beyond daily activity
3 - Symptoms provoked by normal daily activity
4 - Unprovoked symptoms

LVGRADE LV grade based on LV ejection fraction
1 - > or = 60%
2 - 40-59%
3 - 20-39%
4 - <20%

STRESS Stress test
0 - Not done
1 - Negative
2 - Positive

C.A.D. RISKS:

DIABETES Diabetes Mellitus (insulin or non-insulin dependent) (YES/NO)

DOI Control of diabetes
(>12/31/97) 1 - Diet

2 - Oral medication
3 - Insulin
4 - None (patient recently diagnosed)

HYPER Medically treated hypertension (YES/NO)

CHLSTRL Diet or medically treated hyperlipidemia (YES/NO)

FHX Family history of heart disease including stroke, PVD, MI, angina, or heart surgery (YES/NO)

SMOKE Smoking history
0 - Never smoked
1 - Stopped smoking
2 - Still smoking

ASSOCIATED DISEASES:

COPD Chronic obstructive pulmonary disease (YES/NO)
(<01/01/96)

COPDS Chronic obstructive pulmonary disease - severe (YES/NO)
(>12/31/95) (Patient must be receiving daily inhalational or oral medication to improve breathing)

THROMB Previous stroke or T.I.A. (YES/NO)

PVD Peripheral vascular disease (YES/NO)

RF Significant renal failure
(<01/01/95)
0 - No
1 - Not on dialysis pre-op
2 - Chronic peritoneal or hemodialysis

NEWRF Significant renal failure
(>12/31/94)
(<01/01/97)
0 - No
1 - Yes (as defined by the surgeon or patient is on dialysis)

DIAL Pre-op dialysis
(>12/31/97)
0 - No
1 - Yes

MARFAN Patient has been diagnosed with Marfan syndrome
(>06/20/91)
0 - No
1 - Yes

RECG Pre-op electrocardiogram (rhythm)
0 - Normal sinus rhythm

1 - Atrial fibrillation or flutter
2 - Complete heart block/ pacemaker

OTHER SYMPTOMS:

CHF Congestive Heart Failure (YES/NO)

SHOCK Shock (YES/NO)
NOTE: If Shock=YES, NYHA=4

SYNCOPE Fainting spells (YES/NO)

PRE-OPERATIVE MEDICATIONS:

ASP Aspirin stopped <7 days prior to surgery (YES/NO)
(>12/31/94)

AMI Taking amiodarone at time of admission (YES/NO)
(>12/31/94 and <01/01/96)

CREAT Creatinine level measured preoperatively
(>12/31/95)

ACB PATIENT DATA:

DISLAD Stenosis > or = 50% in the Left Anterior Descending arterial system (YES/NO)
(LAD, DIA-1, DIA-2, PER)

DISCX Stenosis > or = 50% in the Circumflex arterial system (YES/NO)
(CX, INT, OM-1, OM-2)

DISRCA Stenosis > or = 50% in the Right Coronary Arterial system (YES/NO)
(RCA, PIV, PL, AM)

LMAIN Stenosis > or = 50% in the Left Main Artery (YES/NO)

DISNUM Total number of diseased arterial systems

ARTERIES BYPASSED:

LIMA LIMA graft used
1 - LAD
2 - CX
3 - RCA

RIMA RIMA graft used
1 - LAD
2 - CX
3 - RCA

RADIAL Radial artery graft used
(>12/31/97) 1 - LAD
2 - CX
3 - RCA

GFTLAD Number of grafts to the Left Anterior Descending region
GFTCX Number of grafts to the Circumflex region
GFTRCA Number of grafts to the Right Coronary Artery region

ENDART Endarterectomy
1 - LAD
2 - CX
3 - RCA

CVPA Coronary vein patch angioplasty
1 - LAD
2 - CX
3 - RCA

OTHGFT Other conduit used
1 - Gastro-epiploic artery
2 - Artificial
3 - Cryo-preserved

ACBNUM Total number of distal grafts

GENERAL OPERATIVE DATA:

PUMPCASE Patient went on pump.
0 - No
1 - Yes
2 - Standby only

MININV Minimally invasive surgery (YES/NO)

(>12/31/96)

ORTIME Surgeon time in OR (minutes)

PUMP Cardiopulmonary bypass time (minutes)

CLAMP Cross-clamp time (minutes)

CIR ARR Circulatory arrest (minutes)

BSA Body Surface Area (m2)

HT Height (cm)

WT Weight (kg)

MYOPRO Primary form of myocardial protection
- 0 - None, coronary perfusion
- 1 - Warm blood cardioplegia (35-38°C)
- 2 - Tepid blood cardioplegia (20-35°C)
- 3 - Cold blood cardioplegia (<20°C)
- 4 - Crystalloid cardioplegia

TECHNIQUE Method of cardioplegia
- 1 - Intermittent
- 2 - Continuous
- 3 – Both

DIRECTION Direction of cardioplegia
- 1 - Antegrade
- 2 - Retrograde
- 3 - Both

HYPOTHER Lowest level of systemic hypothermia
- 0 - Normothermia
- 1 - Mild (30-35°C)
- 2 - Moderate (20-29°C)
- 3 - Profound (<20°C)

OFFPUMP Patient came off pump
- 0 - Well
- 1 - With inotropes/balloon
- 2 - Died in OR

COMPLICATIONS (before discharge or within 30 days of DATEOR if discharged):

IABP	Insertion of intra-aortic balloon pump 0 - None 1 - Preop in CCU 2 - Preop in OR 3 - Intra-op in OR 4 - Post-op in ICU NOTE: If IABP=3 or 4, LOS=YES
REOPNUM	Number of times patient returns to the OR on this admission
REOP	Reopening after surgery 1 - Bleeding 2 - Tamponade 3 - Shock/arrest 4 - Infection 5 - Dehiscence 6 - Redo Surgery 7 – Other
REOP2 (>12/31/98)	Reopening after surgery (same as for REOP) Patient is reoperated on twice post-op. This field contains the second most important reason for reopening regardless of date.
REOP3-5 (>12/31/98)	Same as REOP2
IECG	Postoperative electrocardiogram - ischemic changes while in hospital 0 - No changes 1 - New ST and T wave changes 2 - New Q waves 3 - Complete left bundle branch block 4 - Loss of R waves NOTE: If IECG=2, MI=YES
CKMB (<01/01/99)	Highest postop CKMB level (units)
CK (<01/01/99)	Highest corresponding CK level (units)
MI	Perioperative myocardial infarction (YES/NO) NOTE: This includes all patients with new Q waves post-op
INO	Use of dopamine (>3 mg/kg) or dobutamine in ICU for more than 30 minutes to maintain a blood pressure greater than 90 mmHg after adequate reload and after load reduction (YES/NO)
LOS	Low output syndrome (use of inotrope or mechanical devices for more than 30

	min. to maintain a blood pressure greater than 90 mmHg with a C.I. < 2.2 l/m/m2) (YES/NO) NOTE: This includes all patients with IABP inserted intra-op in OR or postop in ICU
POSTRF	Renal failure post-op 0 - No 1 - Yes, required dialysis 2 - Yes, no dialysis but patient died in renal failure
PACE	Insertion of permanent pacemaker (YES/NO)
OCVENDYS	Medically treated ventricular dysrhythmias (YES/NO)
AFIB (>12/31/96)	Presence of atrial fibrillation post-operatively (YES/NO)
OCDVT	Deep vein thrombosis (YES/NO)
OCPULMC	Pulmonary complications (YES/NO)
SEIZURES (>02/04/93)	Seizures post-op (YES/NO)
TIA (>09/08/94)	Transient ischemic attack as diagnosed by cardiologist (YES/NO)
PREHB (>05/31/94)	Hemoglobin pre-op
POSTHB (>05/31/94)	Hemoglobin at time of discharge
PACKCELLS (>07/01/94)	Number of units of red cell concentrate used for transfusion.
STROKE	Evidence of a persistent neurological deficit 0 - No 1 - Yes, intraoperative 2 - Yes, postoperative
INFLEG	Leg infection 0 - None 1 - Superficial - based on the presence of purulent discharge, involving skin & fat (<1 cm deep) 2 - Deep - below the superficial fascia (>1 cm deep)

INFSTERN	Sternal infection 0 - None 1 - Superficial (based on the presence of purulent discharge, involving skin & subcutaneous tissue only) 2 - Deep (requiring surgical debridement)
INFARM	Arm infection - based on the presence of purulent discharge, involving skin & fat <1 cm deep) (YES/NO) >12/31/98
INFSEP	Sepsis - based on positive blood culture (YES/NO)
SURVIVAL	Patient left hospital and survived thirty days post DATEOR (YES/NO)
DCTO (>12/31/96)	Patient discharged to: 0 - Patient died 1 - Home 2 - Convalescence 3 - Other
PROC	Surgery category: to be completed after review of surgical procedure 1 - ACB: isolated bypass surgery (without VALVE or OTHER procedures) 2 - VALVE: any valve replacement or repair (with VALVE procedure as a primary procedure 3 - OTHER: other cardiac surgery including CHD, LVA, MISC and AAS (with VALVE procedure as a secondary procedure)
NOTES	Fifty character space for extra notes

CONVENTIONS:

0 - Negative entry (NO)
1 - Positive entry (YES)

MM/DD/YYYY Date format

PTID:
If last name contains an apostrophe, it should be included in the PTID.
If last name is hyphenated, hyphen should be deleted from the PTID.
If the last name contains only two letters, an underscore (-) should be used in place of the third letter in the PTID.
If the last name contains a space (eg: Le Baron), the space should be deleted from the PTID.
If the patient has a first initial or first name but goes by their middle name, the fourth letter of the PTID should come from their middle name (i.e: the name commonly used).

MISSING DATA:

For TEXT (character field):	BLANK
For Numbers (numeric field):	9 or –9
For Dates (date field):	BLANK If partially complete, use midpoint: eg.: Jan. 1989 = (01/15/1989) 1978 = (06/30/1978)