INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Université d'Ottawa • University of Ottawa

Investigation of Surgical Patient Scheduling at the University of Ottawa Heart Institute

by Ella Belisario

A thesis

submitted to the School of Graduate Studies and Research of the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Science in Systems Science

University of Ottawa
Ottawa, Ontario
Canada

April 21, 1997

Acknowledgments

I would like to express a sincerely gratitude to a number of people who helped me in preparation of this work.

Professor Colin M. Lay, my primary supervisor who introduced me to the health system, led with advice and support, devoted to me his time which were substantial and very much appreciated.

Professor Jeffrey Sidney, my supervisor, whose financial support, advice and encouragement made this project possible.

A number of people from the University of Ottawa Heart Institute: Wilbert J. Keon, M.D., Heather Sherard, J. Earl Wynands, M.D. for support and facilitation of the research, Howard Nathan, M.D., Michael Bourke, M.D., Lorna Bickerton for all their support, advice and consultation, and Alison White and Michael Asselin for the help in preparation of data sources.

Professors Daniel E. Lane and Tony Quon for their time, consideration and advice on the development of the simulation and predictive models.

And the last but not the least my husband, Berzan, and all my family who supported and encouraged me all the time.

Abstract

This thesis presents an attempt to assess how prediction about patients' expected stay in the operating room, intensive care and postoperative surgical units could be used in the process of scheduling of patients for cardiac surgery.

The study was carried out in the setting of the University of Ottawa Heart Institute (OHI) where current scheduling practices and the process of flow of patients through surgery were studied. Cardiac surgery is resource intensive and has moderately high in-hospital morbidity and mortality associated with it. Intensive care units, where all cardiac patients recover after surgery, is found to be the biggest bottleneck for the smooth flow of patients in this specialized setting.

Data were collected from all patients who underwent cardiac surgery at the University of Ottawa Heart Institute during the 1994 and 1995 calendar years. Information of interest is the relation of demographic, clinical and procedural factors for each patient to the time of operation, length of stay in the intensive care unit after surgery, length of postoperative stay until discharge from the hospital and in-hospital mortality after surgery.

As a results of the analysis of the University of Ottawa Heart Institute surgical scheduling system and patients data we (i) understood how the current system is operating, (ii) developed and assessed models for predicting length of stay in ICU after cardiac surgery.

A diagram was developed to show the basic flow of patients from initial registration and triage using the Provincial Adult Cardiac Care Network (PACCN) urgency criteria, through presurgical stay in hospital, surgery, postsurgical stay in the intensive care units and postsurgical stay in the regular patient care units. Review of the literature, and consultation with the medical and nursing staff of the OHI, suggested that stay in the intensive care units (recovery room and cardiac surgical unit) was an important bottleneck worth studying.

Data from 1994 and 1995 were collected from five separate computer files at the OHI and were analyzed to study patterns of length of stay in the intensive care units (ICU LOS). The 1994 data were used as a training set, while the 1995 data were used for a test set. A computer package known as Knowledge Seeker was used to develop tree structures showing the impact of different predictor variables on ICU LOS, and to choose the most important/significant variable at each branch of each tree. ICU LOS was measured as a continuous variable, and also categorized into 2, 3 and 6 classes of LOS. Knowledge Seeker calculates an accuracy rate for its classification trees,

but this is misleading when the dependent variable has one category that predominates. In models with 2 categories of LOS, approximately 70 to 73% of the patients stay less than 2 days in the ICU, but the classification tree model using preoperative predictor variables is only able to add 0 to 1% more to the accuracy of prediction than would be obtained with a naïve model that always predicts the majority class, without taking account of any predictor variables. Models which add intra- and postoperative variables to the predictor set perform slightly better.

In the literature we found an information gain criterion derived from information theory originally developed for electronic communication systems. Information and entropy are complementary concepts, with entropy being a measure of the extra information necessary to gain a perfect prediction for a categorical dependent variable. A prediction model would have 100% relative information gain if it gave a perfect prediction for each patient being classified. It would have 0% gain if there were no improvement over the naïve model. A negative information gain score represents a misleading model! The twelve classification tree models tested gave improvements ranging from -8%, through 0%, to a maximum of 9% relative information gain.

Part of the difficulty of prediction is explained by the fact that the average ICU LOS dropped from 1994 to 1995 at the OHI, and we speculate that there may have been structural, financial, patient care or policy changes that led to the drop. It was particularly difficult because it impacted many patients at the 2 day borderline of "short" stays in the ICU.

Table of contents

Index of Tables.

Index of Figures.

1. Introduction.

The work presented in the following pages is a starting point for conducting more thorough research on possibilities for scheduling of patients for surgery by applying preoperative prediction about patients postoperative length of stay in intensive care.

1.1. Motivation.

Demand for open-heart surgery as one of the treatments for cardiovascular disease and specifically coronary heart disease exceeds the capacity of the Canadian health care system. Open-heart surgery is defined as any surgical procedure performed on a heart or a great vessel in which cardiopulmonary bypass is used. The patient's blood is diverted from the heart and lungs, passing instead through a special pump and oxygenating system outside the body. The heart is then stopped while the operation is in progress. The most common procedures are coronary artery bypass grafting (CABG) and valve replacement. The treatment of coronary heart disease is expensive, with a need for specialized facilities, trained personnel, complex surgical procedures and increasingly sophisticated technology. Because of the nature of open heart surgery, postoperative intencive care is more complex than for general types of surgeries. Each patient after cardiac surgery requires close monitoring and stabilizing in the postoperating intencive care. Given the high cost of cardiac surgery and the fact that cardiac intensive care units (ICU) constitutes a major cost component of costs of heart surgery, simply increasing resources is not a viable option today when amount of health care resources is limited.

One approach to finding a solution to this problem is to investigate if scheduling of patients for heart surgery at the University of Ottawa Heart Institute (OHI) could be made more effective so that more patients could be treated without violating urgency requirements and without increasing the resources allocated to the hospital.

One of the major factors limiting capacity is the constrained number of beds in the intensive care unit. Utilization of ICU beds is very high during working days (Monday to Friday), which leads to blocking of scheduled surgeries when all ICU beds are occupied and patients are not in condition to leave the ICU. A model for predicting length of ICU stay could be useful for deciding which of two or more patients in equal need of surgery should be scheduled first, given potential blockages in the ICU.

Over last ten years several cardiac surgery risk stratification models have been developed for predicting morbidity and mortality during cardiac surgery. Many different variables have been found to be associated with increasing risk in cardiac surgery. In the most recent studies by Jack V. Tu et al. (1993, 1994,1995) [38, 39, 40] risk models were developed for predicting long length of stay in the ICU after cardiac surgery and suggestions were made for using this kind of model for optimizing the scheduling of cardiac surgery patients in times of limited ICU resources.

1.2 The study objectives, setting and population.

Questions that we ultimately want to answer are the following:

(i) To what extent can the prediction of a patient's length of stay in the ICU after cardiac surgery be utilized for improving the scheduling of cardiac surgery patients?

(ii) Is it possible to increase the throughput and lower the average waiting time before surgery by modifying the scheduling

(iii) Is it possible to improve use of the limited resources by modifying procedures for scheduling patients for surgery? If so, what changes to the current system should be made, and at what cost?

In order to approach these questions, it is necessary first to understand how the current scheduling of patients for heart surgery happening in the settings of OHI. After that the idea of applying preoperative predictions of ICU LOS for heart surgery resource use optimization should be studied. There are developed by different researchers models that are supposed to categorize patients preoperatively by different categories of ICU LOS. Most of them categorize patients by two categories: short and long stay in ICU (see section 2.2 for details), which most probably is not sufficient for scheduling purposes. Therefore the next step will be to try to find models that could predict patients ICU LOS more precisely. With assumption that it is possible to predict patients ICU LOS good enough that it could be used for scheduling principles, one could concentrate on developing a model where different scheduling principles could be applied and evaluated. We formulated the following objectives for our work:

- to understand how the current system is operating;
- to develop and assess a model for predicting length of stay in the ICU after cardiac surgery.

With respect to patients, the boundaries of the system under consideration (for open heart surgery performed in the OHI) includes the time from patient acceptance for surgery until discharge from hospital.

There are about 1250 open-heart surgical operations performed annually in the OHI of which about 35-40% urgent and emergency cases. Out of 1250 operations about 900 are CABG operations. However, in developing scheduling strategies it is necessary to take in consideration all other surgical procedures after which patients require use of ICU. The opportunity to do "tactical" scheduling applies only to the urgent and elective cases, not to emergency cases.

The great number of professionals involved in patient care make the surgical system even more complex. Any changes in the scheduling system will affect the patients, the caregivers (surgeons, nurses, physicians, interns), and administrative personnel. In this study it is important to recognize the needs of all of these individuals:

(i) patients: optimal care, safe environment, short waiting time;

(ii) caregivers: needs for well-planned, accurate schedules produced on time, least overtime, balanced workload;

(iii) hospital administrators: lower cost, high operating room utilization, high personnel productivity level, efficient running of the suite, reduction of overall surgical waiting list.

Given the complexity of the system and the problem itself, the current work is regarded as an exploratory study to develop a framework for more detailed research. In particular, we shall attempt to provide a basis for assessing the usefulness of various scheduling approaches, given the nature of the system and the available information.

1.3 Structure of the thesis.

Chapter 2 provides general background for the problem. We start with a general discussion of queuing problems for coronary surgery, including a survey of work that has been done to implement a formal priority system for coronary surgery in Ontario, Canada. This is followed by descriptions of several of the most recent and recognized cardiac surgery risk stratification models for predicting morbidity, mortality and length of stay after cardiac surgery. Further background is given for the problem of surgical scheduling, including discussion of several additional works on the scheduling of patients for surgery.

Chapter 3 documents the information that was gathered in interviews with medical and administrative personnel at the OHI regarding the current system. The concentration is on understanding the current patient flow process through cardiac surgery, and on the current methods (principles, guidelines, etc.) used for scheduling patients for surgery.

Chapter 4 provides a description of the patient sample and the available data describing these patients which provide the fundamental data for our study. The main function of the data is to estimate time spent in the various stages of the patient flow process, most especially, time in the surgical suite (if possible, actual surgical time, which may be less than time in the surgical suite) and time in the ICU. In fact, given that the ICU is in most cases the scarce resource which limits capacity, it is the prediction of ICU time that is most crucial. We discuss approaches to selection of variables to be used as predictors for times in the stages of the patient flow process.

Chapter 5 details the specific approaches to developing predictive models, using the variables selected for analysis as discussed in Chapter 4. We provide general background on the use of a classification and decision trees technique (using KnowledgeSEEKER software), and discuss the results and their interpretation. Various models are explored: models based upon preoperative versus postoperative data, models bases upon continuous versus categorical response variables, and models based upon different numbers of predictive variables. The analysis yields a classification of patients into homogeneous groups, for each of which we provide a distribution of time in the ICU.

2. Literature Review.

2.1 Queuing problems and prioritization for coronary surgery.

This section gives a background on the problem of queuing for coronary surgery, on how patients are prioritized in the queue, and the ethical issue of delaying of service such as heart surgery. Canadian physicians deal with cardiac surgery queues by using a triage process, which assesses patient priorities based on severity of symptoms and likelihood of early ischaemic events. The current criteria for prioritization for cardiac surgery evolved, with regional modifications, from triage guidelines published in 1990 by a consensus panel (the Revascularisation Panel) into a formal priority system whose goal is to monitor waiting lists in the several Ontario cardiac surgery centers in order to reduce or control the lists by moving patients to other centers when surgery is required very urgently.

C. David Naylor *et al.* (1990-1993) [27, 28, 29, 30] intensively studied queuing problems for coronary surgery, waiting list prioritization and the acceptance of formal priority systems by Canadian specialists. They summarized the Revascularisation Panel's consensus methods (1990) [27] and recommendations for assigning priority to patients who wait for revascularisation procedures and set out a simple scoring system to assess patients' urgency for surgery. Clinical factors were found the most relevant for assessing urgency of need for revascularisation and key variations on each factor were defined. The Urgency Rating Score (URS) system based on symptoms status was also developed (Table 1).

It was suggested, that although it is not one of the URS factors (see Table 1), the risk of procedure-related morbidity and mortality constitutes a minor urgency determinant, since it may well affect timing because of practical constraints (e.g., when ICU beds are in such short supply that a prolonged ICU stay could cause cancellation of procedures for another person at the same urgency ranking). Hence, a model for predicting length of ICU stay could be applied here to optimize resource use. When patients of equal urgency compete for an insufficient number of available ICU beds a patient with low risk of procedure-related morbidity may be given priority.

Table 1 : Urgency Rating Scale Accepted by the Expert Panel.[1]

URS score	URS level	Period within which revacularization is expected to be performed	Typical features
1	Emergency	Immediate revascularisation	Right coronary artery damaged during coronary angiography with acute occlusion.
2	Extremely urgent	Within 24 hours	Unstable patient in coronary care unit on a nitroglycerin drip (class IV-C), breakthrough symptoms, known left mainstream stenosis, and impaired left ventricular function.
3	Urgent	24-72 hours	Unstable patient in coronary care unit (class IV-C), symptom-free on moderate parenteral nitroglycerin dose, single-vessel disease with 90% stenosis in proximal left anterior descending artery (LAD).
4	Semi-urgent	72 hr-14 days	Partly stabilized after acute unstable angina; still pain with minimal exertion (IV-B) despite maximum oral therapy. Three vessel coronary artery disease without left main or proximal LAD involvement.
5	Short list	2-6 weeks	Now stable after acute unstable angina (IV-A). Has three vessel coronary disease. Impaired exercise tolerance on oral therapy, and exercise ECG shows early and markedly positive findings.
6	Delayed	6 weeks-3 months	Stable class III angina pectoris. Two-vessel disease not involving the left mainstream or proximal LAD. Markedly positive exercise ECG.
7	Marked delay	3-6 months	Stable class II angina pectoris. Two-vessel disease affecting branches of the circumflex and LAD. Moderately positive exercise ECG.

Based on the guidelines of the Revascularisation Panel, in 1991, the Ontario Ministry of Health established the Provincial Adult Cardiac Care Network (PACCN), a provincewide computerized registry system for monitoring cardiac surgery waiting lists in Ontario. All adult patients who require heart surgery in Ontario are entered into this database at the time of referral for cardiac surgery so that they can be triaged according to the urgency of their surgery.

Studies by C.D. Naylor and associates (1992,1993) [28, 29, 30] indicated remarkable acceptance by cardiologists in Ontario of the reality of a queue-based allocation system.

---

[1] Reproduced with permission from Social Science and Medicine Vol. 37, No. 1, pp. 62.

Physicians are able to arrange patients in the queue according to judgments about the urgency of clinical need. The researchers found that when the priority of patients was assessed by surgeons, they defined urgency as follows: emergent surgery was required to be done within 24 hours; urgent surgery was required to be done within the same admission of the patient to the hospital (maximum recommended waiting time is about two weeks); elective surgery allowed the patient to wait outside of hospital. Assessing the match between the URS recommended maximum waiting times and actual waiting times, the authors found that proportionately more patients in the more urgent categories than in the less urgent categories had waiting times in excess of the maximum acceptable, as defined by the expert panel (see Table 1) [29]. For example, they calculated that only about 12 % of patients with a rounded URS between 1 and 4 underwent the procedure "on time" or "early", as compared with 74% of patients with a rounded URS between 5 and 7. It was shown that throughout the entire period of their study (between Jan. 3, 1989 and June 30, 1991) the proportion of patients who had excessive delays remained significantly higher among those with a low URS than among those with a high URS [30].

Given that the true emergency cases (URS 1) routinely crowd out scheduled procedures, the authors suggested that an insufficient number of openings for urgent cases (URS 2-4) may be available in the operating room schedule. Scheduled elective cases (URS 4-7) represent the regular livelihood and personal practice of each surgeon. Urgent cases can be accommodated expeditiously only if the surgical service creates a pool of openings for that purpose or if a scheduled case is canceled. The authors pointed out a tendency to reduce delays for long waiting elective patients, a practice which may delay servicing more critical patients with unstable symptoms [30].

While these findings are not easily generalizable for our case, since their methodology had several limitations (e.g. limited and selected group of patients), the results are suggestive, and should be considered carefully in designing any new approach to scheduling.

Jafna L. Cox (1994) [6] studied the ethics of queuing for Cardiopulmonary Artery Bypass Surgery (CABS) in Canada. The author concludes that queues provide an imperfect but practical solution whenever fixed resources cannot meet demand. Further, a system of rationing access for coronary surgery as it exists in Canada is a better solution than denying access to surgery altogether in a system of financial rationing [6]. When priority in a queue is based on medical conditions and waiting does not pose unacceptable risk of adverse effect for patients, and when patients understand that, then such rationing is ethical. While physicians are responsible for their own patients, they must also help

preserve a health care system to which all patients have access. Therefore timeliness of care and efficient allocation of scarce resources have to be ensured by any system designing to prioritize service.

2.2 Cardiac surgery risk models.

Significant risk factors for mortality after cardiac surgery have been identified and several studies have developed models for prediction [9,12,14,31]. For example, O'Connor used significant factor in multivariate logistic regression analysis to define five probabilistic risk categories for mortality after cardiac surgery [31]. The Society of Thoracic Surgeons (STS) National Cardiac Surgery Database model by Edwards et. al. (1994) [9] is based on a Bayesian probability model of operative mortality during cardiac surgery.

Two groups of researchers: Higgins et. al. (1992) [14] and Tuman et. al. (1992) [41] have mentioned that mortality alone in many cases does not correlate with the complication rate after cardiac surgery nor with postoperative hospital stay. Authors studied risk factors for predicting postoperative complications (morbidity). They defined morbidity as the presence of one or more cardiac, pulmonary, renal, or neurologic complications or infections post-operatively. Using a multivariate logistic regression model clinical severity score systems were developed. Tuman at. al. defined three risk categories of low, increased and high risk for postoperative morbidity. It was suggested that preoperative prediction of postoperative morbidity is important for assessment of expected duration of intensive care unit stay after cardiac surgery for prospective planning of resource utilization (ICU staff and bed allocation) [41].

Using the findings of a previous study about relation of postoperative morbidity and ICU length of stay after cardiac surgery, Tu. et. al. (1993, 1994, 1995) [38, 39, 40] attempted to develop a predictive instrument for length of stay in the ICU following cardiac surgery. Two methods were used: a neural network was used to stratify patients into three risk groups for prolonged stay (low, intermediate and high) and multivariate logistic regression was used to develop a predictive index. The constructed predictive index was found to predict length of ICU stay greater than 2, 4, 7 and 10 days and patients' death in the validation data set [39]. They also studied methods for predicting mortality risk and total postoperative length of stay after cardiac surgery. A six-variable index (age, sex, left ventricular function, type of surgery, urgency of surgery, and repeat operation) was developed and validated to predict in-hospital mortality, prolonged ICU stay (>6 days), and prolonged postoperative stay (>17 days) after cardiac surgery. Data for index derivation and validation were collected from nine adult cardiac surgery institutions in Ontario [40].

An early study by Jack V. Tu *et al.* (1993) [38] showed that neural networks could also be used as a predictive instrument, and they offer both advantages and disadvantages when compared to more standard statistical techniques. The authors proposed that a neural network could relatively easily be linked to the computer programs currently being used for maintaining cardiac surgery waiting lists and a risk stratification score could be automatically produced, but they questioned whether the neural network model would be accepted by physicians for use, because the methodology and results are unfamiliar.

One result of their studies [39] was the finding that a total of 26.4% of the 1990 patients were classified as long-stay patients, and they accounted for 74.8% of the total use of ICU beds. Within this group 6.2% of the patients had ICU stays over 10 days, and these patients accounted for 47.2% of ICU bed use. Thus a relatively small number of patients consumed a disproportionately large amount of the available resources.

The authors suggest that one potential application of their predictive index is for use in patient and staff scheduling when resources are limited. In the case of contention between two patients, they calculated that successful prediction of ICU stay combined with a rule to schedule first the patient with the shortest expected length of ICU stay would result in the greatest probability (83.7% of the time) that both operations would be completed within 48 hours, whereas operating first on the high risk (longer ICU time) patient would result in a much lower probability (39.2%) of both operations being completed within a 48 hour time period. In other words, by successfully predicting prolonged ICU stay, the two patients could be treated in a 48 hour period 83.7% of the time rather than a single patient, with the same level of resources [38].

Another scheduling option they suggest would be to schedule earlier in the week patients at lower risk and to schedule later in the week those at higher risk whose condition was stable enough for a few days delay so as to maximize ICU bed use over the weekend, when operating rooms (ORs) are closed to elective patients. Either of these approaches would help to improve the use of existing ICU resources, assuming that the risks of a long stay are not greatly increased for patients whose surgery is delayed.

The authors realized that scheduling patients according to their length of ICU stay is complicated and a somewhat controversial ethical issue. They argue that although patient need should be the primary determinant in scheduling processes, there often could be situations in which two or more patients have relatively equivalent need and potential benefit from surgery, but resources are limited. Hence use of the predictive index in the manner described would maximize

the number of operations that could be completed in these situations and would benefit the majority of patients.

Consideration of emergency patients was included in the index (although these patients' operations could not be delayed). The authors suggest that users of the index may want to project how long patients undergoing emergency surgery will stay in the ICU since this may affect their staffing requirements and the ability to schedule other patients.

Although the developed predictive index is simple to use, for scheduling purposes we will need more precise estimates of the date when the patient can be transferred from the ICU to the regular surgical care unit. Information that becomes available about a patient's condition post-operatively might be used for more precise estimation of a patient's stay in the ICU following surgery.

Hoyt and Lay's (1994) [16] analysis of Neural Network use for cost estimates for Heart Bypass Surgery at the OHI found that clinical and diagnostic indicators for individual patients before surgery can be used to estimate the total cost of their heart-bypass surgery. They showed that using this system, exception reporting and analysis for individual patients can be generated on a real-time basis throughout the treatment episode.

2.3 Operating room scheduling.

Recall that the system under consideration is that surrounding the open heart surgery unit at the OHI. The time frame for a patient extends from the time of acceptance for surgery to departure from the OHI. The setting for our study is a specialty hospital in which a substantial percent of admissions are surgical cases of which at least one third is urgent and emergency cases. We note also that the OHI is a teaching hospital, and this may occasionally affect some of the scheduling decisions.

Since the nature of open heart surgery, the length of the intensive care following surgery, and the important role of emergencies are very different from most other types of surgery the results of many studies of scheduling methods for OR's are not easily applicable in our case. We refer to these studies only to indicate the range of work that has been carried out in efforts to improve OR scheduling.

For a complex situation in which experimentation with the actual system is costly, dangerous and impractical, simulation modeling is an appropriate approach. It helps to describe, analyze, and predict the behavior of the system. Some studies have shown that some other modeling techniques such

as queuing models and mathematical programming do not adequately capture important aspects of the OR [7, 17, 21]. The following is the summary of the works that have been done on the issue of operating room scheduling and simulation.

R. Davies (1994) [8] developed simulation models to predict resource use and costs arising from the treatment of patients with coronary artery disease in Guy's Hospital, London. His complex model incorporated all possible medical and surgical treatments for coronary artery disease patients: an angiogram, an angioplasty, CABG and the establishment of long term medical therapy for patients. Results from the simulation indicated that the number of cardiology ward beds was inadequate, causing a bottleneck in the system. The software, which was designed to be of general application to health problems, includes facilities for scheduling of operating theaters and combined modeling of demands made by different groups of acute and chronic patients. Some noteworthy aspects of Davies' study are: (i) A model was developed which not only describes the movement of patients from one type of treatment to another but also describes the use of resources. The model can be used to examine the effects of increasing levels of demand for treatment, different treatment policies and provision of various levels of resources. (ii) The model may be used to budget resources and to solve scheduling problems, providing allowance is made for the uncertainties of patients' progress and survival.

Although the author's simulation programs included the facility to model scheduling of OR theaters, analysis using this facility for improving surgical suite utilization was not conducted. Finally, the author was not able to integrate all of the collected data into one data source in order to link with the simulation system. He pointed out that the simulation could only therefore be made available as a long term decision support system if such an integration were to be completed.

Another recent work on simulation of scheduling of patients for surgery was done by Kathy E. Fitzpatrick *et. al.* (1993) [12] who studied the application of computer simulation to improve scheduling of general hospital operating room facilities in the United States. They showed that computer simulation could be an important and valuable tool in the hospital management decision-making process. Using a simulation model, three alternative block schedules were compared against the scheduling procedure in use, which prioritized cases based on source (waiting list, elective and emergency) and used a first-come/first-served scheduling heuristic for the facility. Performance comparisons of the schedules were made on the basis of throughput, average waiting time, the distribution of waiting times, queue characteristics, and facility utilization.

Their system schedules patients solely on the basis of availability of operating room (OR) times. Alternative approaches are to schedule surgical cases through an admission scheduling system, which

can be used to govern both medical and surgical admissions, through a combined OR/admission scheduling process. This approach would choose an 'optimal' admission date on the basis of the patient's expected length of stay, the expected availability of beds during the stay, the expected OR time, and the availability of OR time.

James M. Magerlein and James B. Martin (1978) [24] reviewed the literature on scheduling of patient demand for surgery and outlined an approach to improving overall performance of hospital surgical suites. Reported scheduling systems are categorized into those that schedule patients in advance of the surgical date and those that schedule available patients on the date of surgery. The authors discussed the failure to implement the majority of reported scheduling schemes. The following reasons for this lack of implementation of apparently superior scheduling methods are suggested: the failure of proposed system to satisfy the medical staff, the failure to consider the surgical scheduling and allocation question within the context of the entire institution, the failure to model the hospital environment adequately, and the failure to consider the total costs associated with model implementation.

Przasnyski (1986) [34] conducted a literature review on operating room scheduling and described it by five different areas: operating room utilization, planning and organization, cost containment, staff and resources scheduling and booking of operating room for surgeries. He concludes that the complexity of OR scheduling is a function of dynamic nature of the scheduling process and the amount of uncertainty in the information involved. Przasnyski says that in his survey of the literature on OR scheduling he found both very technical works, with presentation of solutions to problems, but with many assumptions and restrictions; or completely nontechnical statements of associated problems, often with common sense proposals for local improvements in a few areas. Przasnyski finally concludes that to make any real progress, the best aspects of technical and non-technical approaches should be combined. Specifically, the overly restrictive assumptions of the technical approaches need to be relaxed so that more complete coverage of the problem can be achieved.

J. Blake and M. Carter [3] have been working on development of a proper conceptual framework for operating room scheduling. Reviewing a number of works Blake and Carter discovered that one reason for the lack of practical success of different innovations for operating room scheduling mentioned by other authors (Magerlein and Martin [24], Przasnyski [34]) is the lack of a clear, complete and systematic understanding of the problem and concept of operating room scheduling. Decision making for the surgical process management is described by Blake and Carter as a process distributed over strategic, administrative and operational levels which are linked together by a planning and performance control structure. Scheduling of operations consists of advance scheduling, allocation

scheduling and external resources scheduling, decisions for which involve elements from all three levels of management [3]. Advance scheduling is described as a process of assigning to a patient some future date for surgery and this is a strategically oriented decision. Allocation and external resources scheduling are more operationally centered decisions. Allocation scheduling is a process of assigning to a patient scheduled for a particular date an operating room and start time. And external resources scheduling is a process of allocating to a patient scheduled for surgery all necessary preoperative and postoperative patient care resources [3]. The authors demonstrated how all those elements of operating room scheduling are interrelated on different management levels by functions of planning and control. This framework gives an understanding of the process of surgical scheduling in a picture of organizational interdependencies.

2.4 Summary.

This chapter incorporated background on the several aspects of the problem under consideration. Understanding of how patients are prioritized in the queues for cardiac surgery and the ethical issues accompanying this process is important for trying to design any new scheduling principles. The Provincial Adult Cardiac Care Network (PACCN) has been established in Ontario to monitor cardiac surgery waiting lists. It has established a formal triaging process to identify patients with greater and lesser urgency of undergoing CABG operations and to suggest their maximum recommended waiting times. Physicians' have traditionally had informal processes for rationing scarce resources when scheduling for cardiac surgery. They have adapted well to the formal system of scheduling of patients according to their urgency. Any suggestion to alter priority for surgery in such a system must ensure timeliness of care for each patient and efficient allocation of resources.

Although the PACCN system identifies patients with greater and lesser urgency of undergoing CABG operations, it has been suggested that additional patient related factors might assist in better scheduling of heart operations. Several studies, conducted to develop models for predicting morbidity and mortality after cardiac surgery, proposed that ICU length of stay could be predicted preoperatively and consequently used for scheduling of patients when ICU resources are limited. The argument is that although patient urgency should be the primary determinant in scheduling processes, there often could be situations when resources are limited and there are two or more patients with relatively equivalent need and potential benefit from surgery. In this case priority between those patients could be given based on predicted use of the scarce resource.
Analysis of the factors promising prediction of ICU length of stay for the patients undergoing cardiac surgery in the setting of the OHI should yield insights into the extent to which such prediction could be utilized for improving the scheduling of cardiac surgery patients.

A review of different reports on scheduling of operating rooms, whether simulations of general or specialized operating rooms, or discussion of what constitutes the process of scheduling for surgery, helped us to realize the reasons for the lack of implementation of most of the proposed new scheduling initiatives. In order to represent the system adequately, a model should include all activities from the time of the patient's admission until discharge. Only in this way can one incorporate all of the following constraints: (i) space and time availability of ORs, ICU beds, recovery room beds, and other facilities; (ii) people (staff assignment preferences and patient demand for surgery by urgency category); (iii) time (OR and recovery room bed time availability); (iv) need for patients and surgeons to know in advance the time of operation.

To summarize, the idea of improving patient scheduling for cardiac surgery with the use of predicted ICU length of stay could be utilized if (i) it is possible to develop a model that would predict patient length of stay in the ICU better than physicians do it now, and (ii) it is possible to incorporate such information into real day to day scheduling process for cardiac surgeries.

3. Description of Patient Flow through Cardiac Surgery at the University of Ottawa Heart Institute

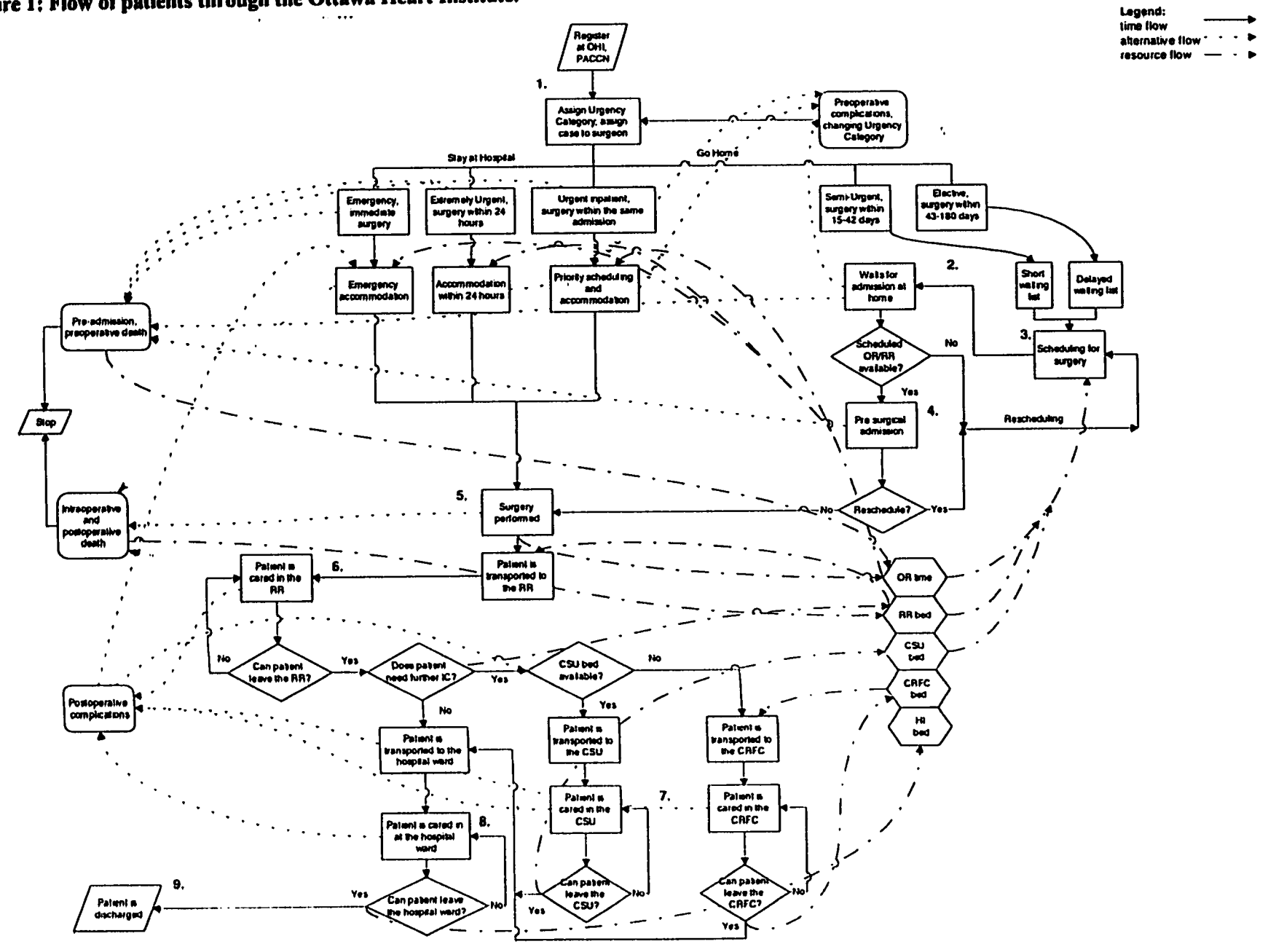
Our study is concerned with the scheduling of open heart surgery, of which the most frequent type is bypass grafting, a treatment for patients with coronary artery disease. The patient flow process through Cardiac Surgery at the OHI consists of a very complex set of activities and events. For purposes of this study we focus on individual patients and study a patient's "processing" by the system from the time of entry into the system until departure. Patient's entry is defined to be the acceptance of a patient for cardiac surgery by Heart Institute surgeons, along with the registration of the patient in the PACCN database. Patient's departure is defined to be normal discharge from the Institute, death of the patient, "signing out", or transfer to another hospital. Figure 1 shows the different possible flow of patients through the Ottawa Heart Institute.

3.1 Acceptance for surgery.

After a positive decision for a cardiac operation, the patient's registry form is completed by the cardiologist and submitted to the Provincial Data Base. The information which is entered includes clinical factors from which an URS (see description in the section 2.1) and maximum recommended waiting time are automatically calculated. According to the PACCN system, patients are categorized into three groups: Urgent with URS up to 4.00 and Maximum Recommended Waiting Time 0 to 14 days, Semi-Urgent with URS 4.01 to 5.00 (15 to 42 days), and Elective with URS 5.01 to 7.00 (43 to 180 days). Within the urgent category there are emergency patients who need to be operated on within 24 hours (they will not be scheduled cases) and urgent patients who need to be kept in hospital until their operation (they are priority cases to be scheduled).

Along with registration in the PACCN data base a patient is assigned to one of the eight OHI surgeons, who will perform the operation. Usually these assignments are done by a triage coordinator in such a manner that all surgeons will have approximately equal workloads.

Figure 1: Flow of patients through the Ottawa Heart Institute.



3.2 Current operational and scheduling policy.

The possible patient flows through the system are shown in Figure 1.

It should be noted that the normal treatment process described below may be interrupted by the patient's departure because of death or other reasons. Because elective patients are able to wait for many days or several months their scheduling has several important differences from urgent/emergent patients' scheduling.

3.2.1 Scheduling of elective patients for surgery.

1) Patient is registered at OHI, and assigned a PACCN urgency category.

   Note: the urgency category may be changed during steps 2 to 4, in which case the patient flow would henceforth be described under the new category.

2) Patient returns to his/her residence to await scheduling.

3) Patient scheduling.

   Patient is scheduled for surgery according to his/her urgency need and available OHI resources; rescheduling may take place, depending on conflicts with more urgent cases.

4) Patient enters hospital for presurgical prep (usually the day before surgery).

   Again there is a possibility of rescheduling, which could result in a return to step 2, or the patient may stay in hospital until his/her surgery can be performed.

5) Surgery performed.

   Note: during steps 6, 7 and 8 of the process described below, the patient may in some cases have to return to surgery for reoperation; in such a case, patient will usually be classified as an urgent case, and reoperation is accommodated as soon as possible. This entails a return to step 5.

6) Patient to Recovery Room (RR).

   After surgery the patient is transported from the OR directly to the RR or Cardiac Surgical Unit (CSU) (if there is no RR bed available) where patients stay until they are extubated, made hemodynamically stable, and arterial and venous lines are removed. During this intensive care cardiovascular, respiratory, renal, neurologic, gastrointestinal functions, bleeding and infections are managed. The level of support is related to preexisting functional disorders, the nature of the patient's response to the cardiac repair, duration of the cardiopulmonary bypass and any coexisting system disorders [1]. Most patients are stabilized within 24-48 hours after surgery. Occurrence of one or several complications in stabilizing any of those functions could prolong

the length of ICU stay over the normal time period. Then a medical decision is made as to whether the patient needs further intensive care, in which case he/she is moved from RR bed (if he/she were previously there) to stay in the CSU bed until he/she could be transferred to the hospital postintensive care nursing unit. If patient does not need post RR intensive care, he/she is cared for in the RR until able to be transferred to the hospital postintensive care nursing unit.

Rehabilitation and discharge planning commence actively once the patient is transferred out of the intensive care unit.

7) Patient to CSU (if necessary) or to the Cardiac Reference Centre (CRFC), if there is no room in the CSU.

8) Patient to postintensive care nursing unit. Sometimes complications could occur and patient could be returned to the intensive care units.

9) Patient discharged.

3.2.2 Scheduling of urgent and emergency patients for surgery.

Note: the PACCN urgent and semi-urgent category of patients is reinterpreted by the OHI surgeons and assigned surgical priority as follows:

- emergency: requiring immediate surgery;
- extremely-urgent, with a maximum wait of 24 hours until surgery;
- urgent: patients are kept in hospital until their operation, with an allowable waiting time of up to two weeks.

The procedure described above for elective patients is modified as follows.

1) Patient arrives and registers at the OHI, and is assigned a priority category.

2) Patient stays at the hospital until he/she could be accommodated for operation.

The rest of the Time Flow of Urgent Patient is the same as for Elective and Semi-Urgent patients starting from step 4.

3.3 OR Staffing Hours.

The regular schedule of ORs is currently:

Monday-Tuesday:

- two ORs staffed from 8 a.m. till 6 p.m., and usually two cases in each OR scheduled each of these days;

- one OR staffed from 8 a.m. till 8:30 p.m., and usually two cases scheduled each of these days;

Wednesday-Friday:

- one OR staffed from 8 a.m. till 3:30 p.m., and usually one case scheduled each of these days;
- one OR staffed from 8 am till 6 p.m., and usually two cases scheduled each of these days;
- one OR staffed from 8 a.m. till 8:30 p.m., and usually two cases scheduled each of these days.

There are usually two slowdown periods during the year: July (the reasons are conferences and vacations) and December-January during which only two ORs instead of three are staffed during regular working hours.

There are no ORs left for emergency cases during regular hours of operation; when emergency cases arrive, they preempt scheduled cases.

Each surgeon who is a staff member at the Heart Institute is assigned a certain period of time in which an operating room will be available for his cases. This regularity in operating room time is a convenient arrangement for the surgeons, who can schedule fixed office hours and regular hospital patients visiting time. Block assignments are convenient for operating room scheduling procedures and for teaching. Currently, 27 time slots are allocated as following: six surgeons have three slots per week each, one surgeon has two time slots, and one surgeon has five slots per week. The last two of predetermined time slots left for on call surgeon. Each week one of the eight surgeons is assigned to perform emergency cases.

Teaching cases are scheduled as usual during normal hours of operation, but since they could take more time than non-teaching ones, teaching cases are scheduled so that they will not tend to disrupt the schedule (e.g. in operating rooms with longer hours of operation).

After normal working hours and on Saturdays, Sundays and Holidays there is usually one team available to perform emergency operations, but if one team is not enough to accommodate all emergency cases, additional teams could be called in.

3.4 OR scheduling principles.

Scheduling for cardiac operations performed at the OHI bases its capacity plans on elective, semi-urgent and urgent inpatient cases. Cases which are not immediately life-threatening but which must be surgically treated by the end of the next day (or several days within the same admission) are designated urgent. Booking of cases is only done for regular operating hours; emergency cases during and after regular working hours are operated on as they arrive or as soon as they can be accommodated.

Based on a planned number of 27 cases per week, the OHI leaves a pool of twelve slots per week for urgent-emergency cases. The remaining fifteen slots per week are distributed between surgeons for scheduling of elective patients.

The PACCN triage coordinator in the Registry Office assigns new patients to surgeons to give them approximately equal workloads. Each surgeon has a predetermined number of slots per week and a secretary who coordinates booking of his cases with the OR and the Registry Office. Currently the OHI does not have a centralized OR booking system. Instead the surgeons make decisions about the scheduling and booking of their individual elective patients. The completed personal schedules of the surgeons are then combined and entered into a computer and printed for each week.

Each week one of the surgeons is responsible for urgent-emergency cases. He has a pool of slots to coordinate accommodation of urgent-emergency patients: all his own slots for this week, one slot from each surgeon and two slots as on-call surgeon. The weekly schedule for the ORs is updated every day. Schedule changes may occur because of emergencies and for other reasons. By 4.00 p.m. each day all changes for the next day should be completed and communicated between the OR and ICU staff, surgeons and the person who does the rescheduling of cases.

When any changes to the weekly schedule occur, the precedence of cases to be operated on is always such that those with more threatening conditions are accommodated first. Other cases are accommodated depending on availability of OR time, and RR and CSU beds.

Emergency cases with life-threatening conditions which must be scheduled for immediate surgery take precedence over all other types of cases. Accommodating these emergency cases may result in delay or cancellation of all other urgency category cases. Scheduled cases could then be completed later on the same day or rescheduled for other days. If there is a time slot available for

performing the preempted case and there are available RR and CSU beds, then the preempted operation could be performed on the same day.

Urgent patients who arrive with conditions which must be surgically treated by the end of the next day could be operated on after normal working hours in order not to alter the schedule.

Emergency and extremely urgent patients may have their surgery performed after regular hours by the on-call or by any other available surgeon. The scheduling of simultaneous emergency cases is based upon medical considerations.

Canceled and postponed cases are rescheduled with priority. There are constraints for rescheduling postponed and canceled cases. For example, if a scheduled patient gave blood for his own operation, then his surgery must be performed within three weeks after last unit of blood was given.

3.5 Patient Admission.

Patients who were not kept in the hospital are admitted to the preoperative unit on the same day or one night before surgery. All necessary tests and laboratory reports should be completed before that time. Exceptions to this are made in case of surgical emergencies where time does not permit. Patients scheduled for surgical procedures should have their chronic medical conditions under optimal control. The decision to admit a patient for cardiac surgery is made by the cardiac surgeon and cardiac anesthesiologist. On occasion surgery may be deferred for medical reasons. If no complications occur during preoperative assessment, on the day of the operation the patient is prepared for surgery.

3.6 Summary.

This chapter has discussed the process which cardiac surgery patients go through and the principles of scheduling of patients for cardiac surgery in the settings of the OHI.

The flow of patients through cardiac surgery, including the admission process, surgery itself, and postoperative intensive care are common to all cardiac surgical institutions. However institutional differences are always encountered in preoperative preparations, evaluation, postoperative monitoring, discharge practices, etc.

Distinctive characteristics of the scheduling process in the OHI are the following: (i) there is no centralized scheduling for elective patients, (ii) there is a predetermined number of cases scheduled each week, (iii) two pools of time slots exist for every week: one for urgent-emergency cases and another for elective cases, (iv) schedule is based primarily on availability of OR time, and scheduled operations are not performed if no ICU bed is available.

Applying a conceptual framework for surgical process management developed by Blake and Carter [3] we can see how patients are granted OR bookings and what controls are imposed on the process. Strategic advance scheduling defines the number of available ORs, the hours of their operation, and the time blocks of surgeons. This timetable of the facility service defines, in aggregate terms, this facility throughput [3]. According to the planned caseload future operations are scheduled, but because of unforeseen circumstances such as elective cancellations or emergency arrivals, allocation of particular ORs and start times is only possible the day before actual surgery. External resources allocation is the third component of surgical scheduling. To be feasible each surgical appointment requires that a set of preoperative and postoperative resources be identified [3]. The allocation of an ICU bed for each scheduled patient is a dynamic operation because of the uncertainty in estimating how long those beds will be occupied.

Optimizing the scheduling of patients in times of limited availability of ICU beds is necessary on operational level when, for example, one ICU bed is available for the next few day and two patients equally require surgery. Scheduling of patients with low probability of postoperative morbidity and prolonged ICU stay is necessary when only one ICU bed is available. Such schedule optimization is not likely to change much aggregate outputs, but benefit all: patients, staff and administration.

4. Description of Data Sources and Variables Selected for the Study.

Our study was retrospective: initially we defined our population as consisting of patients who underwent surgery at the OHI, and proposed to use patients' data collected at the time of surgery as independent variables for estimating times in the various stages of treatment.

After the OHI Research Ethics Committee review was completed (Appendix 1A) and permission was granted for the study to proceed, several physicians and nurses of the OHI staff (they are referred to hereafter as the OHI consultants) volunteered to participate in and guide the study clarifying our understanding of the structure of the OHI, patient flows, scheduling rules and the accuracy and validity of data from the various computer files.

The eventual study population consisted of all patients who underwent open heart surgery at OHI between January 1, 1994 and December 30, 1995. A unique identification number was used in the data bases in order to retain patient confidentiality.

In this chapter we describe how we identified key predictive variables, and the data sources we used to obtain them. We describe the collection, management (e.g., merging, coding, etc.) and verification of data, and the exploratory data analysis approaches used to choose the most promising predictive variables (Section 4.3).

4.1 Selection of potential variables for the analysis.

Following the objectives of this study, our task was to select a set of variables that would be related to prediction of patients' post-surgical ICU length of stay.

Risk factors for morbidity and mortality at cardiac surgery have been described in many studies and several predictive models were described in section 2.3. The studies by Tuman et al. (1991) [41], and Tu et al. (1994, 1995) [39, 40] indicate that increasing clinical risk score is associated with higher mortality, higher morbidity and longer ICU LOS. Tu et al. (1995) [40] summarize the studies by different researchers on risk factors for morbidity, mortality and long ICU stay. Analysis of these factors suggests grouping them in three categories: (i) patient related, (ii) problem related and (iii) comorbidity factors (factors related to associated diseases). We have adapted the results as shown in Table 2.

The columns in Table 2 indicate the principal authors of different studies and the variables that they used for developing their risk models. Factors were rearranged into the three categories. Notice that all factors in the Table 2 indicate preoperative conditions of patients.

Discussions with the OHI consultants helped us to translate the kinds of variables used by other researchers (illustrated in Table 2) and match them up with the appropriate data in OHI's data base. This involved a fair amount of interpretation of data variables, which were often not collected in the forms reported by other researchers. In fact no two research sites can use data with equivalent definitions because the data collection systems are different. For example, 'Special circumstances' as they were specified in other studies (Table 2) do not exist in our data sources. Congestive Heart Failure (CHF), LVEDP, IV inotropic support, cardiogenic shock - all were considered by OHI as similar very severe conditions, and we have data for CHF, cardiogenic shock and also for Heart Failure Class. IV nitrates was considered the same as angina class IVC. PTCA emergency was considered as failed PTCA with emergency CABG. Angina class was considered instead of one variable 'unstable angina'. Renal dysfunction is sometimes defined differently by different specialists (see Table 2). In our data source this variable was recorded as Renal Disease. Table 3 lists all variables that were found in OHI data sources as the closest match to those in table 2.

Since the prediction of ICU length of stay is crucial for developing more effective scheduling procedures and depends on the patient's condition before the operation and on the results of the operation, we also selected intraoperative and postoperative (during the first few hours after surgery) factors that contribute to overall ICU length of stay.

Table 2: Risk Factors Found in Large Studies of the Risk of Cardiac Surgery.[2]

	Studies								
Risk Factors for Morbidity/Mortality and length of stay	**Parsonnet**	**Hannan**	**O'Connor**	**Higgins**	**Tuman**	**Edwards**	**Tu(95)**	**Tu(94)**	**Tu(93)**
I. Patient Related:									
Age	x	x	x	x	x	x	x	x	x
Female gender	x	x	x		x	x	x	x	x
Weight related:									
Morbid obesity	x					x			
Body Surface Area			x						
Weight <=65 kg				x		x			
Smoking history:									
Current tobacco abuse						x			
> 100 pack-years smoking						x			
II. Problem Related:									
Previous history:									
Repeat operation	x	x	x	x	x	x	x		x
Previous Myocardial infraction		x			x	x			x
Cerebrovascular disease				x	x	x			
/prev. cerebrovasc. accident (stroke)									x
Prior vascular surgery				x					
Current problem, initial clinical									
indicators of severity:									
Type of surgery							x	x	x
Valve (aortic/mitral)	x	x		x	x				
CABG plus valve	x				x				
Other		x							
Emergency/urgent surgery	x		x	x	x	x	x	x	x
Left Ventricular function	x	x	x	x	x	x	x	x	x
Aortic stenosis	x			x					
LV aneurysm	x					x			
Left main disease		x				x			

[2] Reproduced with permission Circulation Vol. 91, No. 3, Feb. 1995, pp. 683. Copyright 1995 American Heart Association.

Risk Factors for Morbidity/Mortality and length of stay	Studies								
	Parsonnet	Hannan	O'Connor	Higgins	Tuman	Edwards	Tu(95)	Tu(94)	Tu(93)
Unstable angina		X				X			
Congestive heart failure		X			X				X
Mitral insufficiency				X					
Cardiogenic shock						X			
IV nitrates						X			
IV inotropic support						X			
PTCA emergency						X			
Valvular disease						X			
One-, two, and three-vessel disease						X			
LVEDP			X						
Preoperative IABP	X								
Special circumstances:									
Catastrophic states	X								
Rare circumstances	X								
Disaster		X							
III. Coexisting Problems (Comorbidity factors):									
Renal dysfunction									
Dialysis dependency	X	X				X			X
Elevated creatinine level				X	X				
Diabetes	X	X		X		X			X
Pulmonary hypertension	X				X				
Hypertension	X					X			X
Charison comorbidity score			X						
COPD				X		X			X
Anemia				X					

To avoid confusion, we adopt the following practice: associated with each patient is a measurement/characteristic vector consisting of "predictors". Using these "predictors" (sometimes called independent variables), we attempt to predict accurately the dependent variable (in our case ICU length of stay).

The final list of all variables that will be considered as potential predictors of ICU length of stay is given in Table 3.

Table 3: Variables selected for the analysis.

VARIABLE	MEANING AND/OR USE OF VARIABLE
	Dependent variable:
ICU length of stay including CRR and CSU	Total time patient spent in the ICU area
	Preoperative predictors
Patient related factors	
Age	Patient's age in years
Gender	Patient's Gender: M- Male, F-Female
Weight	Patient's Weight
Height	Patient's Height
Smoking History	Current Smoker: 'Y'- Yes, blank - No/not known
BSA	Patient's Body Surface Area
Problem related factors	
Emergency	1-Emergent, 2-Urgent, 3-Elective
Procedure Group	Aortic Surgery, CABG, CABG+Other, CABG+Valve, Valve, Valve+Other, Transplant, Other
Angina Class (CCS Class)	Presence of Angina: 1-asymptomatic, 2 and 3 - stable, A, B, C -unstable by CCS, blank - No/not known
LV Function	Left Ventricular Class: 0-normal, 1,2,3 or 4-worst,

Heart Failure Class	blank - not known 1,2,3 or 4 - NYHA Class;
REDO	blank - not known 'REDO' indicates previous cardiac surgery;
Acute MI	blank - No/not known Myocardial Infraction of sufficient intensity and duration to produce death of tissue. [1] 'Y'- presence of Acute MI;
Aortic Aneurysm	blank - No/not known 'Y'- presence of Aortic Aneurysm;
Aortic Dissection	blank - No/not known 'Y'- presence of Aortic Dissection;
Arrhythmia	blank - No/not known Disorders of impulse formation or abnormal conduction or both. 'Y'- presence of Arrhythmia;
Cardiogenic Shock	blank - No/not known Pump failure, i.e. failure of the ventricle to generate adequate blood pressure output. 'Y'- presence of Cardiogenic Shock;
Cardiomyopathy	blank - No/not known 'Y'- presence of Cardiomyopathy;
CHF (Congestive Heart Failure)	blank - No/not known Pulmonary venous congestion due to heart failure [1]:
Congenital Defect	'Y'- presence of CHF; blank - No/not known 'Y'- presence of Congenital Defect;
Failed PTCA (Percutaneous Transluminal Coronary Angioplasty)	blank - No/not known 'Y'- failure of PTCA; blank - No/not known
Left Main Stenosis	'Y'- presence of Left Main Stenosis;
	blank - No/not known

LV Aneurysm	'Y'- presence of LV Aneurysm; blank - No/not known
Valvular Disease	'Y'- presence of Valvular Disease; blank - No/not known
Ventricular Rupture	'Y'- presence of Ventricular Rupture; blank - No/not known
VSD (Ventricular Septal Defect)	'Y'- presence of VSD; blank - No/not known
Comorbidity diseases	
Asthma	Presence of Asthma: 1-Mild, 2-Moderate, 3-Severe; blank - No/not known
C.O.P.D. (Chronic Obstructive Pulmonary Disease)	Presence of C.O.P.D.: 1-Mild, 2-Moderate, 3-Severe; blank - No/not known
Diabetes	Presence of Diabetes: 0-Absence of Diabetes, 1-Diet Controlled, 2-Controlled with Oral Med., 3-Controlled with Insulin; blank - No/not known
Hepatic Disease	'Y'- presence of Hepatic Disease; blank - No/not known
Hypertension	'Y'- presence of Hypertension; blank - No/not known
Obesity	'Y'- presence of Obesity; blank - No/not known
Prior CVA (Cerebral Vascular Accident)	Presence of Prior CVA: 1-Embolic, 2-Hemorrage, 3-Carotid Vascular Disease; blank - No/not known
Pulmonary Hypertension	Presence of Pulmonary Hypertension: 1-Mild, 2-Moderate, 3-Severe; blank - No/not known
PVD (Peripheral Vascular Disease)	'Y'- presence of PVD; blank - No/not known

Bleeding Tendency	'Y'- presence of Bleeding Tendency; blank - No/not known
Renal Disease	'Y'- presence of Renal Disease; blank - No/not known
Other Medical Problems	Name of medical associated diseases different from the above.
	Intra-operative and postoperative predictors
IABP	Use of Intra-Aortic Balloon Pump: 'IABP' -Yes, blank - No/not known
Quality of grafts	(good/fair/poor/endarterectomy): 'E' -at least one graft quality is end., blank - No/not known
Pump time	Time on Cardiopulmonary Bypass (CPB) in min.
Medications on transfer from OR to ICU	Medications on Weaning from Bypass and Post-Bypass Medications: Calcium, Dopamine, Dobutamine, Epinephrine, Amrinone or other.

4.2 Data sources.

In order to acquire the information that was specified in the previous section we had to use several data sources available at the OHI. All of the sources are computerized data bases. In total we used four different data sources: (i) the file named MASTERS, which contains demographics, procedures, admission and discharge dates; (ii) the Provincial Cardiac Care Network (PACCN) data base; (iii) the Admission-Discharge-Transfer data base (ADT); and (iv) the Total Medical Record system (TMR) which in itself consists of two parts - anesthesia and surgical. All data sources contain information for two calendar years - 1994 and 1995.

Detailed descriptions of all data fields in all data sources that we used are found in Appendix 1: Description of the Collected Data.

Because the files were created by different groups within the hospital for different purposes, and are not maintained as a single integrated system, some of the fields in the data bases

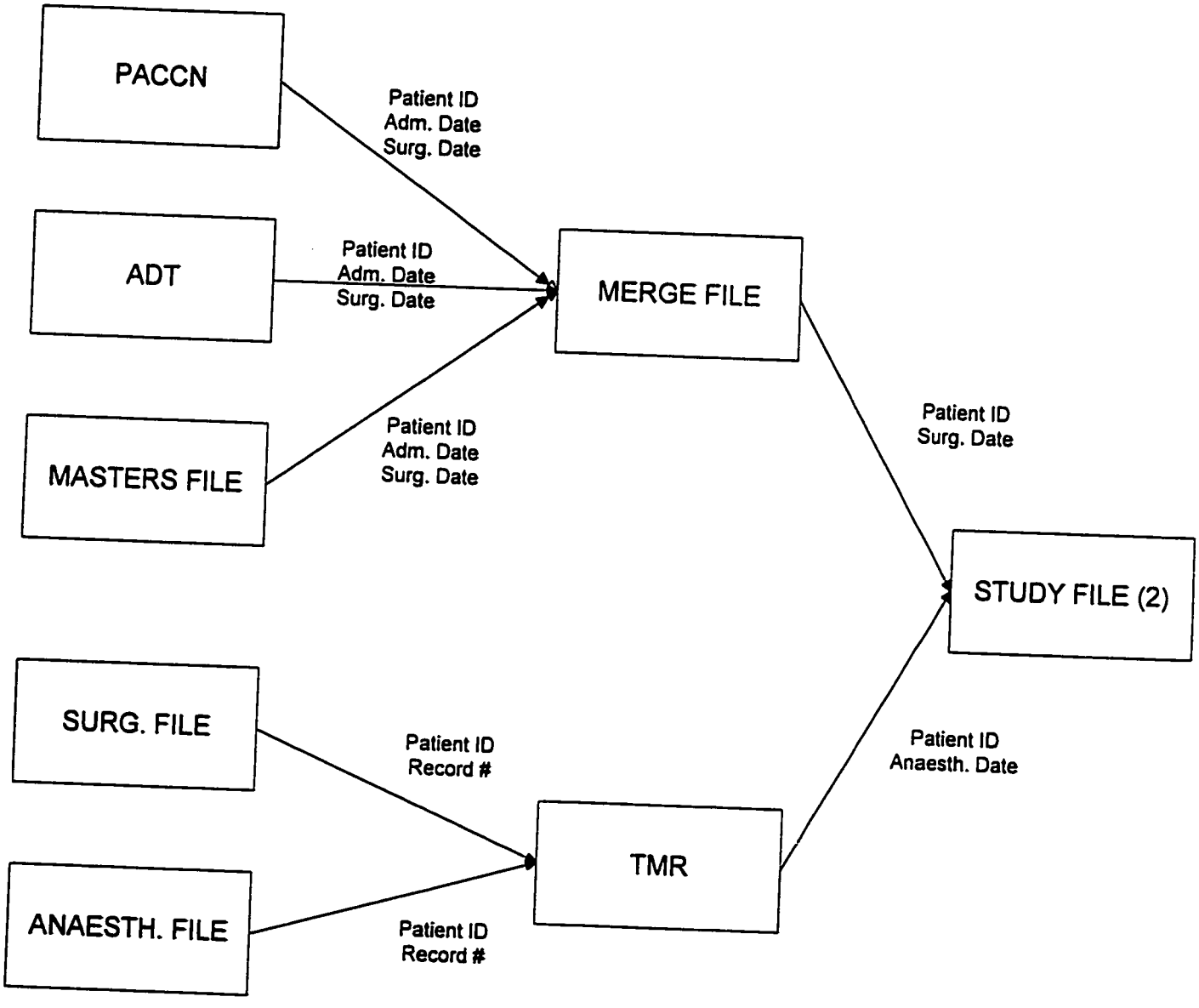
were duplicated. The extent of the problem of data fields that are supposed to be "the same" but actually have not been subjected to rigorous cross checking is seldom appreciated by people who do not have to work matching the details of the field. For each variable we used the source which was considered to be more reliable by the OHI consultants. Extensive validation and correction of the data fields or missing values was not possible within the time limits of the study.

Together four data sources provided all the data for the study. The format of each data source was modified to facilitate merging of the data bases. Most of the required calculation on the fields was done using EXCEL 5.0 (1995, Microsoft Corp.). Assembling the data set used for model development and testing was done using Paradox 4.5 (1993, Borland International Inc.). The linkages necessary for the merging process are presented in Figure 2. The linking keys are shown on the lines joining the source files to the consolidated study file.

All data sources contained the unique patient IDs which provided the key for linking them all together. Verification of linkages was facilitated by the presence of two other common data items in the PACCN, ADT and MASTERS file, namely, patient admission dates and patient surgery date.

Anesthesia and surgical parts of the TMR were merged by patients ID and by record number which corresponded. The date of anesthesia in TMR also had to correspond to the surgery date in the rest of the data sources.

Figure 2: Linking of all data sources together



It was felt that the number of data linkages described above between the data bases provided a strong basis for a reliable merge of the data bases, and the identification of questionable data.

To simplify the study, for a first approximation we decided to exclude from consideration those patients who had more than one operation during the same admission and those who were returned to the ICU after having been transferred to the ward. These cases combined accounted for only about 7% of the total population, but posed difficult analytical problems. In the follow-up work repeat operations could be treated as new, probably emergency cases with REDO variable set to 'yes'. Returns to ICU after patient have been transfered to the ward could be either treated as separate cases with zero OR time or total time in ICU could be added for the returning patient.

After linking all sources together, we obtained complete information on a total of 1400 patients (this includes patients who underwent surgery in calendar years 1994 and 1995). About 600 records from the PACCN, ADT and MASTERS files were not linked to the TMR data file because of incomplete TMR data. The major extent of this problem is understandable when it is noted that the files had never before been required to be matched for an extensive study.

The TMR data base contains mostly procedural data and clinical information. Since we were not able to obtain complete TMR data, we decided to create an additional study file without the TMR data, thus giving us a greater number of cases for analysis but fewer available predictive variables. However, we were able to retain those variables that have been consistently found to be the major risk factors across multiple and very diverse study settings. Those factors include the six risk factors contained in the index developed by Tu et al. (1995) [40]: age, sex, left ventricular function, type of surgery, urgency of surgery and repeat operation. We also had available two out of four of the key postoperative predictors (IABP, and pump time). In our presentation of the data and the results of the analysis we use the code SM to indicate a SMaller set of predictor variables, excluding the TMR records, and LG to indicate LarGer set of predictor variables, including TMR records. More detailed description of the fields in the two newly formed merged files (one with TMR records and the other without TMR records) can be found in the Appendix 2: Description of the Fields in Merged Data Files.

4.3 Description of the sample of patients.

This section provides an overview of the data and rationale for selecting different years of collected data for development and for testing of models for predicting length of stay in the ICU.

Using the "learning set/test set" approach the entire population was divided into two groups. One group - patients who underwent surgery between Jan. 1, 1994 and Dec. 31, 1994 was designated as a learning set and was used for developing of predictive models. The other group - patients who underwent surgery between Jan. 1, 1995 and Dec. 31, 1995 is considered as a test set for validation of the developed predictive models. Another way to select learning and test sets was to randomly divide the entire population into two groups of equal size. This would ensure that the distribution of the dependent variable which is length of stay in the ICU is similar between the two sets. The decision to use two different years of data as learning and test sets was based on the assumption that the predictive model is to be used in forthcoming years and testing it on a separate whole year of data would give a good approximation of its behavior.

The frequency distribution of length of ICU stay in the learning and test data sets is shown in Figure 3.

Figure 3: Distribution of length of stay in ICU in 1994 and 1995.

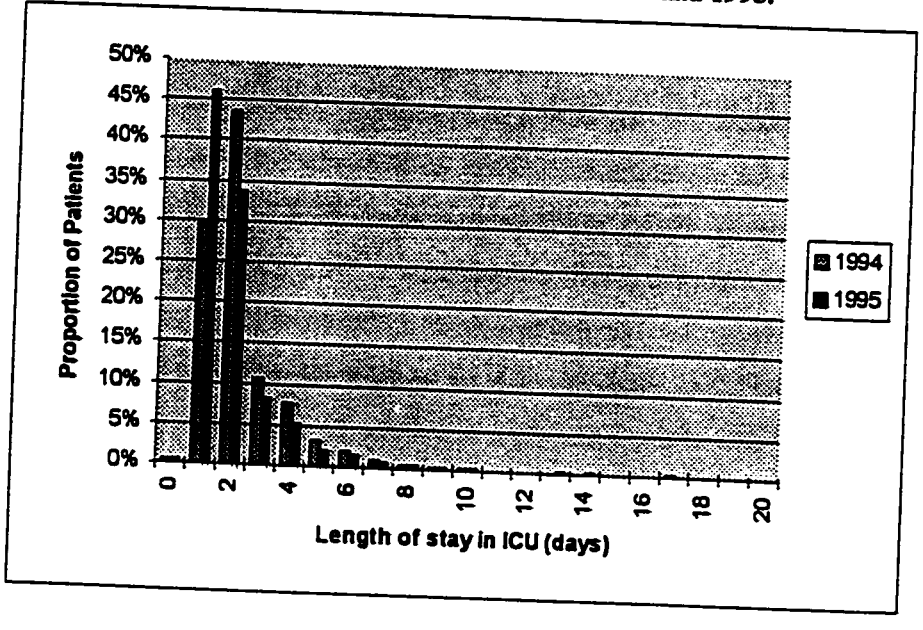


Figure 4: Test for differences in distribution of ICU length of stay between 1994 and 1995.

22 Apr 96 SPSS for MS WINDOWS Release 6.1

----- Kolmogorov - Smirnov 2-Sample Test

ICU_STAY

by YEAR_GR

Cases

1150 YEAR_GR = 1

1092 YEAR_GR = 2

2242 Total

Most extreme differences

Absolute	Positive	Negative	K-S Z	2-Tailed P
.17639	.00706	-.17639	4.1746	.0000

The data are highly skewed to the right and have a long statistical tail. The length of stay varied from 0 to 77.7 days in 1994 and from 0 to 44.5 days in 1995. The median and mean length of stay was 1.87 and 2.6 days respectively in 1994 and 1.69 and 2.4 days in 1995. The Kolmogorov-Smirnoff two sample test for significant differences between distributions of length of stay in ICU in 1994 and 1995 years shows that they are significantly different (Figure 4).

The decrease in average length of stay from 2.6 to 2.4 days suggests that some structural changes in the system took place, such as technological changes, provincial health care policy changes, including stricter financial constraints. In spite of the differences between the 1994 and 1995 data, we shall still use the 1994 data as the training set and the 1995 data as the test set. Whatever results ensue, it may be assumed that a more sophisticated procedure, such as periodic recalculation, "rolling horizon" approach or other that takes into account the non-stationary nature of the data will do better than the results we will derive with respect to predictive ability.

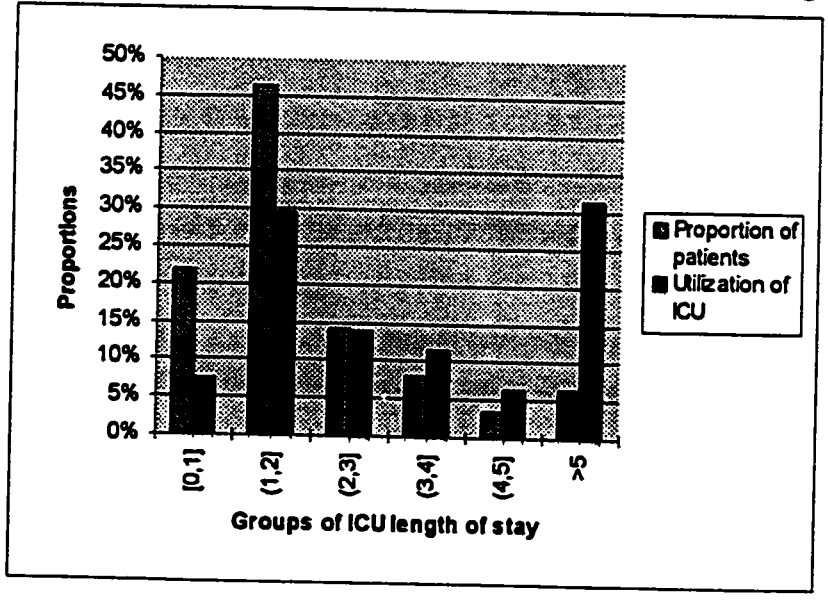
Table 4 and Figure 5 show the use of ICU beds by patients with various length of stay in the 1994 learning set. The majority of patients with uncomplicated stays leave ICU within 2 days, as we can see from distribution of ICU length of stay and as was confirmed by OHI consultants. Therefore defining a stay which is less than or equal to 2 days as a short stay and a long stay as any stay greater than 2 days, we see the use of ICU resources similar to that found by Tu et. al. [39]. For the 1994 data set a total of 32% of patients are classified as long-stay patients, and they accounted for 63% of the total

use of ICU beds. Within that group 6% of the patients had ICU stays over 5 days, and these patients accounted for 31% of ICU bed use. Utilization of ICU was derived by calculating percent of the total number of hours spent by each group of patients. These results correspond to what was found by others and suggest again possibility of using of predictive instrument for wise scheduling of patients when ICU resources are limited.

Table 4: Utilization of ICU resources by patients with different length of stay in ICU.

Groups of length of stay in ICU	Proportion of patients	Utilization of ICU
[0,1]	21.77%	7.53%
(1,2]	46.25%	29.42%
(2,3]	14.18%	13.80%
(3,4]	8.03%	11.54%
(4,5]	3.43%	6.25%
>5	6.33%	31.46%
Grand Total	100.00%	100.00%

Figure 5: Utilization of ICU resources by patients with different length of stay in ICU.



4.4 Summary.

During the data collection phase of the study we identified information required for the analysis and sources at the OHI from which data were collected. There were several different data sources which we used to acquire data. Although we were able to merge different sources together we were not able to link all the records in different data sources. This results in a smaller sample for analysis than it would be if we could find correspondence between all the records in all the data sources. But since the study is retrospective in nature and different information was recorded at the OHI in different data sources we have to constrain analysis to the information in a merged samples. As a result two merged files were created for the analysis of ICU length of stay: one with a smaller number of potential predictors of ICU length of stay, and another with more potential predictors and a smaller number of records.

The problems of matching the data files are not surprising, when it is realized that they were created and maintained by different groups within the hospital, for different purposes, and for target population that overlap significantly but also diverge significantly.

5 Modeling of expected time in the intensive care unit.

5.1 Different tools for developing a predictive instrument

In section 2.2 we described different methods that have been used for developing risk stratification models for predicting mortality and morbidity outcomes after cardiac surgery. In particular multivariate stepwise logistic regression and neural networks were mentioned as tools used by researchers for developing an instrument for predicting ICU length of stay following cardiac surgery (Tu, J. V. et al. 1993-1995) [38, 39, 40]. Another method that has proven to be useful in developing predictive models in different areas and particularly in medical applications is the tree structure technique, which we employ for prediction of ICU length of stay after cardiac surgery. This section explains the advantages of this technique for our study.

In describing what they consider to be some of the major advantages and disadvantages of using neural networks versus logistic regression, Tu and Guerriere [38] point out that

1) neural networks are better than logistic regression for identifying complex, nonlinear relationships between independent and dependent variables;

2) neural networks can be trained even if data are missing; and

3) unlike logistic regression, neural networks have the advantage that they can detect interaction between all input variables implicitly.

One of the major disadvantages of the neural network approach in our context is that clinicians may view this approach as a "black-box" method which does not explicitly explain relationships and interactions between input variables and predicted outputs. It is the norm that methods which are not understood will often be rejected, regardless of their advantages.

Logistic regression is a widely used statistical technique which has been effective in many medical applications. If simple linear relationships exist between independent and dependent variables other methods are unlikely to offer any significant advantages over the regression approach. But in order to produce meaningful results using regression methods it is necessary to explicitly define interaction terms and to ensure that data are complete and accurate. An underlying assumption in logistic regression is that the phenomenon under study is homogeneous, i.e., that the same relationship holds between variables over all of the measurement space.

It should be pointed out that the regression approach is at best difficult to use when dealing with large dimensions, with mixed data types and with non-homogeneous data structures. The basic purpose of the tree structure method is to produce an accurate classifier or predictor and/or to provide insight and understanding of the predictive structure of the data. The latter means that we are trying to get an understanding of what variables or interactions of variables drive the phenomenon. Producing an accurate classifier or predictor means developing an efficient and accurate algorithm to classify or predict the response variable corresponding to future measurement vectors as accurately as possible.

Some of the advantages of the tree structure approach summarized by Breiman et al. (1993) [4] are the following:

1. It provides a simple and intuitive method for classifying subjects, and it effectively handles many varied data structures and both categorical and continuous variables.
2. It is robust with respect to outliers and handles missing data in an appropriate way.
3. Using conditional information it identifies interactions among factors and explicitly explains non-homogeneous relationships.
4. The misclassification probability for the objects is easily estimated at each step of the classification and for the whole tree.
5. The tree procedure output gives information regarding the predictive structure of the data; this information is easy to understand, interpret and use.

Having the above advantages, the tree structure approach becomes an important alternative to other procedures.

Several studies have compared the tree structure technique to other methods such as regression analysis and neural networks on specific problems. A survey of those studies has been done by Murthy (1995) [25]. The main conclusions from most comparisons are that (i) no method seems uniformly superior to the others; (ii) machine learning methods seem to be superior in the presence of noise; and (iii) statistical methods are computationally the most efficient.

Because of its simplicity and power in identifying interaction terms we selected the tree structure approach for developing a predictive instrument for predicting length of stay in the ICU after cardiac surgery.

5.2 Classification and decision tree methodology.

5.2.1 Review of the methodological developments.

The tree structure methodology originated with the development of the Automatic Interaction Detector (AID) method by Sonquist and Morgan in the early 1960s at the Institute of Social Research of the University of Michigan [25]. The objective of this approach is to examine the interrelationships of the variables in the data base and to determine which ones are related to the dependent variable. This is accomplished by recursively subdividing the observations, through binary splits, into mutually exclusive, exhaustive subsets which best describe variance in the dependent (response) variable, based on the least squares criterion. In this way a stepwise optimal tree is constructed which gives the predicted value of the dependent variable as being equal to the mean values of the terminal groups. Development of this technique is traced in Sonquist and Morgan (1964), Sonquist (1970), Sonquist, Banker and Morgan (1973), Fielding (1977), and Van Eck (1980).

One of the many applications of the AID technique to applied problems was the development of Diagnosis-Related Groups by Fetter et al (1980) [10, 41]. The grouping process in this work resulted in the formation of attribute case types (Diagnosis-Related Groups) consisting of sets of patients; each patient in a Diagnosis-Related Group could be expected to receive similar outputs or services from a hospital. Length of stay was used as the output measure.

The Classification and Regression Tree (CART) method and software were developed by Breiman et al (1984) [4]. This method constructs binary prediction and classification trees by recursive partitioning. At each node the tree classification algorithm searches through the variables one by one. For each variable it finds the best split which minimizes the overall tree "impurity function". Then it compares all the best splits for all variables and selects the best of the best. Thus, as the execution of the algorithm proceeds, the population of nodes at each split becomes more and more class homogeneous. The algorithm is computationally intensive, its recursive search over all of the variables to produce a sequence of optimal binary splits results in an extremely large tree.

To validate the classification tree produced from the training set, test or cross-validation estimates of misclassification (or error) rate are calculated. When the error rate is calculated over the test set, each case is counted as either a correct or an incorrect classification depending on whether its observed category is matched with the most likely category of the terminal leaf in which

that case falls. By dividing the total number of incorrect assignments by the total number of cases in the test set we obtain the misclassification rate for the test set. Conversely, by dividing the total number of correct assignments by the total number of cases in the test set, we obtain the accuracy rate for the test set.

It was found [4] that too large a tree tends to result in a higher true misclassification rate than in a tree of more reasonable size. On the other hand, too small a tree does not incorporate some of the classification information available in the data set, resulting in a misclassification rate which is unnecessarily high. A pruning algorithm is then applied which cuts off any branches of the tree that while appearing to improve the tree locally, may in fact not add to the accuracy of the entire tree.

The approach taken by CART to construct a regression tree is similar to that of the AID method. The major difference between AID and CART lies in the pruning and estimation process. The cut-off criterion recommended by Sonquist (1970) [4] in the original AID program was the rule that 0.6 per cent of the "total sum of squares" should be explained in order for a split to take place. Then the re-substitution estimate of mean squared error was used as a measure of the accuracy. CART regression trees use the same criterion as AID: mean squared error is minimized, but in addition test or cross-validation estimates are employed (the tree is developed on one set of data and tested on an independent set of data), and these have in practice been proven to be more effective. To find the best sized tree in regression, CART again employs a pruning algorithm with the same criterion as for splitting.

CART methods have been widely used. Many studies were done in clinical diagnosis using CART techniques and results were found to be competitive with predictions made by physicians (see Breiman et al. (1993)) [4]

Kass (1980) [18] gives a method (known as CHAID) of determining the "best" multiway partitions of the data on the basis of significance tests. Important modifications which are relevant to standard AID include: (i) built-in significance testing resulting in use of the most significant predictor (rather than the most explanatory predictor); (ii) adjustments to calculation of significance using the Bonferroni inequality (which requires a greater level of significance before accepting an individual split); (iii) multiway splits (in contrast to binary); and (iv) a new type of predictor which is useful in handling missing information. Whereas AID operates on an interval scaled dependent variable and chooses from the possible splits on predictor variables that split which minimizes the total within group sum of squares, CHAID operates on a nominal scale

dependent variable and maximizes the significance of a chi-squared statistic at each partition, which need not to be a bisection.

Biggs et al. (1991) [2] made additional improvements to the Kass methodology and developed KnowledgeSEEKER software which incorporates their algorithm. This technique operates on both categorical and continuous types of variables and at each node all predictor variables are considered one by one. The best k-way partition for each variable is found. The chi-squared measure is used for categorical association and the F statistic is used for categorical-continuous associations. The significance of the split for each variable is then used to rank all variables which were found to be significantly associated with the dependent variable at this node. The variable with the highest level of significance is then used to partition the node. The process of tree growing stops when no more significant splits are found.

KnowledgeSEEKER was successfully applied by the Geological Survey of Canada [34]. Using a geographical data set, the presence/absence of base-metal occurrences in each small area of the map represented as polygons was predicted using a tree generated by the program and knowledge based decision rules were generated from the tree as outputs.

Given the ready availability of the KnowledgeSEEKER program and the fact that it integrates many of the effective methods of previous approaches, we conducted an analysis of the ICU length of stay data employing this package. The following section describes in more detail the methodology that we used.

5.2.2 "Growing" a decision tree using KnowledgeSEEKER.

KnowledgeSEEKER finds relationships for any combination of continuous and categorical predictors and both continuous and categorical dependent variables. The tree is structured by repeated splits of the predictor variables space into n subsets (nodes) starting with all observations at the initial node. If c is a number of categories of the predictor variable, then 2 <= n <= c. The "splitting criterion" is choose the split which gives the lowest value of p, which is the probability that the predictor is completely unrelated to the dependent variable. For each relationship that has been identified, KnowledgeSEEKER tests whether the relationship is significant at the pre-defined level of significance using the Bonferroni adjustment [2] for the number of categories of predictor variable which drives acceptable value of p to be much less than .01. Significant relationships are presented as alternative ways of forming the branches of a statistical decision tree, one by one. Relationships are presented in order of statistical significance. Once a desired predictor and its associated n categories are selected, the parent population is physically separated, or partitioned into n subgroups. Then the procedure may be applied iteratively on any one of the descendent groups. This process continues until no more significant splits can be found. The process results in the familiar tree-like structure.

KnowledgeSEEKER incorporates several features that allow the analyst to control the process of tree construction, to achieve both high statistical significance and high explanatory power in the splits obtained. These features include: (i) availability of a manual mode for growing trees; (ii) selection from the most statistically significant splits those with the highest explanatory power; (iii) specification of the desired level of significance; and (iv) option to add and/or ignore different variables during the analysis.

In automatic tree-construction mode, KnowledgeSEEKER may produce an enormous tree; it never drops a statistically significant split from the tree. Unlike CART, this program does not have an automatic pruning algorithm which tries to find optimally-sized trees.

We did not find any comments in the documentation for KnowledgeSEEKER that indicates how the tree produced on automatic mode compares in size with an optimal or near-optimal tree. Since Knowledge-SEEKER does not reject statistically significant splits, we may assume in most cases that the tree produced is at least as big as an optimally sized one. Hence, we would suggest that in practice the analyst should experiment with manually collapsing the tree produced to attempt to improve the automated result.

5.2.3 Classifying cases using the tree.

When the tree is grown the sequence of splits from the root to a given leaf defines a rule for classifying a specific subset of cases (patients). The collection of all the rules is the classification scheme and must account for the total population of cases.

When the dependent variable is categorical the leaf specifies the probability that a case belongs to any of the categories. The sum of the probability values on any leaf must be 1.0. Thus, in this study, the dependent variable ICU length of stay can be categorized as "short", "normal", or "long". The sum of probabilities of short, normal, and long is 1.0 in each leaf. One way to treat the information in any given leaf is to name the leaf by the category with the highest probability, as if it had a probability of 1, and the other categories had a probability of 0 (this can be called the "maximum likelihood rule"). Then the expected ICU length of stay for any patient is given as the average ICU length of stay of that category. Another way is to use the probabilities to obtain a weighted average ICU length of stay for each leaf.

In the case of a continuous dependent variable each leaf specifies the average and standard deviation of the values of the dependent variable for all the cases belonging to that leaf in the data set used for growing the tree.

For the purpose of predicting the ICU length of stay for a new previously unclassified patient, the results of the tree analysis are expressed as a set of rules consisting of IF-THEN statements. The last THEN clause leads to a statement giving either the set of probabilities for a categorical dependent variable or the average value for a continuous dependent variable.

5.2.4 Interpretation of the results.

An important criterion for a good classification predictive procedure is that it does not only produce accurate classifiers and predictors, but that is also provide insight and understanding into the predictive structure of the data. Predictive methods are characterized by their "signals" and their "noise". What we want to do is to get insight from the structure we obtain, to be able to understand the "signal". Even though the tree structure approach generally gives better insights into structure than other methods, careful interpretation of the results must be done to arrive at sound conclusions [4].

One problem associated with tree structure approaches is instability, which basically occurs when there are more than one choice for splits. Instability in the tree reflects correlation between variables which results in alternative prediction rules and noise. A symptom of this in the tree structure is that at any given node, there may be a number of splits on different variables, all

of which are significant. Depending on the choices taken at such nodes, different trees (classification schemes) will result.

Two possibilities exist when selecting alternative variables for a split. If the alternative variable which is not chosen at a split is independent of the variable selected at the split, the rejected variable may appear in subsequent stages with the same power as at the earlier stage. If, however, the rejected variable is correlated with a variable selected at a split, then its explanatory power at subsequent stages would be decreased according to the strength of the correlation. For instance, suppose we are predicting ICU length of stay using predictors age and gender. If gender were strongly correlated with age and age was selected as a partitioning variable, then gender would have lower explanatory power in subsequent stages.

Since data are noisy, the choice between competing splits could be somewhat random. Given a particular problem (characterized by its data set), the decision to accept, to reject or possibly to revise the classification scheme produced by the algorithm could be based both on statistical evidence and on domain knowledge.

Hence, the interpretation of the partition (classification) produced in automatic mode by KnowledgeSEEKER is a complex task, and many factors should be examined and weighed simultaneously. Since KnowledgeSEEKER allows manual control of the tree growing process, the most appropriate variable for a split would be one that satisfies not only the significance criterion imposed by the algorithm, but also results in the formation of groups which are meaningful from the clinical perspective, and which lead to a manageable number of groups which have (as viewed by the analyst) the potential to deal effectively with new cases.

5.2.5 Validation of the results.

To asses the accuracy of the classification (tree) produced by KnowledgeSEEKER prior to application is crucial. The test of the classification is its accuracy in handling new cases. We may randomly split the data set into "training" instances (the "learning set") and testing instances (the "test set"). We generate the classification (tree) using the training set, and validate it using the test set, with some agreed-upon criterion.

Different criteria are used depending on whether the dependent variable is continuous or categorical. We first discuss the case of a continuous dependent variable, and then the case of a categorical dependent variable.

Given a measurement (predictor) vector for a particular case in a test sample and using predictive rules from the learning sample we assign this case to its class (leaf), as previously

described in section 5.2.4, and take as a predicted value the mean value for the leaf of the tree in which that case falls. In a perfect model the predicted value will equal the observed value for that case and the prediction will be perfect. When there is less than a perfect prediction the observed value deviates from the predicted value. The mean squared sum of the deviations represents the error of the model. The value of the mean squared error depends on the scale in which the dependent variable is measured. For this reason, a normalized measure of accuracy which removes scale dependence is often used. The idea is to use as a baseline predictor for the dependent variable its expected value. Then the mean squared error with that value is the estimate of the error when no information about predictors is used. Then we can judge the performance of any predictor based on the test set by comparing the mean squared error over the test set using the predictor to the mean squared error over the test set using the baseline predictor. KnowledgeSEEKER calculates this kind of relative error on the specified test set. The relative error is always non-negative. It is usually, but not always, less than 1. Most predictors are more accurate than the baseline predictor, which does not incorporate any information about predictive variables. But on occasion, some construction procedure may produce a poor predictor with relative error greater than or equal 1.

In a case of a categorical dependent variable, KnowledgeSEEKER uses the misclassification (or error) rate and accuracy rate.

Notice that in using this criterion for judging the performance of the classifier, we assume the maximum likelihood rule. As was described in section 5.2.4 sometimes the analysis does not produce an exact classification and a probabilistic classification is used instead. But the validation criterion used in KnowledgeSEEKER does not account for such a case.

In a problem domain such as the OHI when we use a categorical dependent variable where one category dominates all others (i.e. has a very high prior probability), a "naive" classification rule that assigns all cases to the dominant category will have a very high classification accuracy. The performance evaluation criterion ought to evaluate the improvement achieved over the naive rule. KnowledgeSEEKER does not do this.

The evaluation criterion should therefore take into account the prior probabilities of categories and be able to deal not only with yes-or-no classification but also with probabilistic classification. Kononenko and Bratko [20] suggested an information theory based evaluation criterion which accounts for prior probabilities and deals with various types of imperfect and

probabilistic answers. The following is the description of their criterion with the notation used by the authors.

We used as an example the OHI need to distinguish between elective patients who will have a short stay (2 days or fewer) or a long stay (more than 2 days) in the ICU. Let the short stay be called category C. From Appendix 3B we see that P(C)=.81 (rounded to 2 digits). A naive prediction that each patient is a short stay patient will have an 81% accuracy, but be useless for the scheduling purposes. Following Kononenko and Bratko approach, we assume that a prospective patient will actually have a short stay (known only after the fact), and therefore belongs to category C. A classification rule gives a revised probability P'(C) of having a short stay.

If P'(C)>.81 then we have useful information, but

if P'(C)=.81 we have no information, while

if P'(C) <.81 we have misleading information.

Because of its origin in communication system where data are presented as strings of binary digits (0,1), or bits, information theory measures information in bits.

If P'(C) ≥ P(C) (the useful case) then the information content, I, is given by:

$$I = - \log_2 P(C) + \log_2 P'(C) \text{ bits} \qquad (1)$$

where the first term gives the amount of information that would be necessary for a perfect prediction, and the second gives the remaining amount of information needed after the classification rule has been applied.

Using our example, if P'(C) = .90 for a particular patient then

$$I = - \log_2(.81) + \log_2(.90) = - (-.304006...) + (-.152003...) = .152003... \text{ bits.}$$

However if P'(C) = P(C) there is no information provided so that if P'(C) = .81 then I = 0 bits.

Finally if P'(C) < P(C) the information is misleading and a penalty function is used:

$$I = - [- \log_2(1-P(C)) + \log_2(1-P'(C))] \text{ bits} \qquad (2)$$

So that if the classification rule were to give P'(C) = .72, the information content would be

$$I = - [- \log_2(1-.81) + \log_2(1-.72)] = - [- \log_2(.19) + \log_2(.28)]$$

$$= - [- (-2.3959...) + (-1.03650...)] = - [.559427...] = - .559427... \text{ bits}$$

Note that the penalty for misjudging P'(C) by having it 0.09 less than P(C) is much heavier than the positive value of moving P'(C) the same amount in the correct direction.

In general, the higher the value of P(C), the greater will be the penalty for having the estimate of P'(C) move in the wrong direction.

The information score indicates the amount of information obtained using the classification rule on the case.

The average information score I_a of an instance in a test set of size T is calculated as the mean information score over all instances in a test set [20]:

$$I_a = (\sum_j I_j) / T \quad (3)$$

The relative information score I_r was defined as:

$$I_r = I_a/E * 100\%, \quad (4)$$

where N is the number of categories of the dependent variable, which takes on values of C_i, i=1..N and E is the entropy as defined by

$$E = - \sum_{j=1}^{N} P(C_i)*\log_2 P(C_i) \text{ bits} \quad (5)$$

The entropy is the expected amount of information required for classification of one instance.

Therefore by calculating the relative information score for the test set we find the proportion of the information required for classification of all the cases that was obtained by using the classification tree.

Kononenko and Bratko demonstrated examples in medical diagnosis where a high percentage accuracy score could lead to a low relative information score while a low accuracy could lead to a high relative information score.

5.3 Framework of models used for analysis.

Our goal in the analysis is to find a good predictive model for the ICU length of stay after cardiac surgery. Several models for predicting long length of stay in ICU after cardiac surgery were described in the literature review in section 2.2. One of them was developed using neural networks as a predictive instrument (Tu and Guerriere, 1993 [38]) and other models were developed using multivariate logistic regression and are represented in the form of a risk index based on several additive variables (Tuman (1991) [41], Tu et al. (1994, 1995) [39,40]).

5.3.1 Preoperative vs. postoperative models.

Both of the above mentioned studies deal with predicting a patient's ICU length of stay preoperatively; therefore predictor variables were restricted to those patient characteristics that were readily available preoperatively. Since we are trying to develop a predictive model of ICU length of stay that would be useful for scheduling, predictor variables available preoperatively should be used here.

However, as was mentioned already in section 4.2, the development of intra-operative and postoperative complications is an important contributor to overall ICU length of stay which could have serious consequences for scheduling. Therefore models that predict ICU length of stay using as predictors intraoperative and postoperative (within short period of time after operation) patient conditions could be used as feedback for the scheduler to adjust for unforeseen ICU bed use.

5.3.2 Continuous vs. categorical response variables.

The risk stratification models mentioned at the beginning of this section are stratifying patients into only two categories, namely, long and short length of stay in the ICU. The definition of prolonged length if stay in the ICU varies in these studies: Tu et al. (1993, 1994) [38, 39] define a prolonged ICU length of stay as a stay of greater than two days. In another study (which was a multi-center study), Tu et al (1995) [40] define a prolonged ICU stay as a stay of greater than six days. The justification for the 2 days cut was that "virtually all patients with uncomplicated stays leave the ICU within 2 days". The six day cut-off was chosen because it corresponded to the 90th percentile for ICU length of stay in the Ontario population.

In 90% of the cases, the population in our study experienced a length of stay of less than or equal to four days. Consultation with clinicians from the OHI confirmed the fact that most of the uncomplicated patients leave the ICU within two days. The use of two different criteria (two and four day cut-offs) produce short-stay categories with very unequal prior probabilities between short and long stay categories within their own datasets (74% and 91% respectively). But since the two day cut-off was recommended by clinicians, and since in this case inequality in prior probabilities will be less than with a four day cut-off we decided to construct a classification tree with two categories of ICU length of stay: "short" - less than or equal two days, and "long" - greater than two days. Our hope was that our choice would result in an effective classification scheme, both from the point of view of having low misclassification probabilities and from the point of view of providing an aid to effective scheduling. We also hoped that comparison of our

results with previous studies would validate our choice as being of at least comparable to previous work.

The index developed by Tu et al. (1994) [39] was also found to be useful for predicting stays longer than four, seven and ten days. Clearly, from the point of view of scheduling it makes a dramatic difference whether a patient stays four, or, say, fifteen days.

Another possibility we investigated was to define three categories of ICU length of stay: less than or equal to two days, greater than two and less than or equal to four days, and greater than four days. Thus we are trying to separate long length of stay into more descriptive categories and to find the classification tree for those categories of ICU length of stay. This would provide more precision for scheduling, but the probability of misclassification may be greater.

Finally, we investigated even more detailed subdivision of ICU length of stay: less than or equal to one day, greater than one and less than or equal to two days, greater than two and less than or equal to three days, greater than three and less than or equal to four days, greater than four and less than or equal to five days, and greater than five days. As above, precision for scheduling increases, at a risk of higher misclassification rates.

Each of the three candidate models described above has a categorical dependent variable. Since number of categories and their prior probabilities are different, we shall use as a criterion of comparison the information criterion described in section 5.2.5, which takes into consideration the number of categories and the prior probabilities.

For comparison with previous studies, note that the model with only two categories can be compared to previous studies by comparison of the area under the Receiver Operating Characteristic (ROC) curve [5], which was used in those studies to evaluate the predictive performance of the proposed models. Unfortunately the ROC criterion cannot be naturally extended to more than two categories with probabilistic classification.

Finally, we shall investigate predicting ICU length of stay as a continuous variable. Comparison of the categories produced using the continuous variable approach with the three sets of categories produced using the categorical variable approach may yield useful insights.

5.3.3 Different number of predictive variables.

The index for prediction of ICU length of stay developed by Tu et al. used only six variables as predictors. The performance of this stratification model was found satisfactory based on area under the ROC curve. The neural network models Tu and Guerrier used comorbid diseases in addition to those six variables. The area under the ROC curve, also used for assessing performance of that model, was found to be approximately the same as with the six variables index. This raises the perennial question of predictive models: do additional variables add any statistically useful information? Therefore, in evaluating each one of our analytical models, we compare the results using our models to the corresponding results using the six predictive variables of Tu et al.

5.4 Results of the analysis.

5.4.1 How results are presented.

In order to give the reader an understanding of how the results are presented we will describe briefly the organization of the outputs.

For each model described in section 5.3 the classification tree is presented separately in Appendix 4A. Three sample diagrams are shown in this chapter as Figures 6, 7 and 8. Written at the top of each output is the name of the file which gives information about the model. Thus, the first two letters of the file name tells whether small ("SM") or large ("LG") number of variables were in a set of predictors used in the analysis (refer to section 4.2). For the first set we used nine preoperative variables; in the second set we added associated diseases from the anesthesia file.

Next in the file name for the outputs is '94' which means that learning set for the analysis was based on the collected data from 1994 calendar year.

For the first digit after underscore meanings are the following:

1 - a model using a continuous dependent variable (the actual postoperative length of stay in the ICU – ICU LOS);

2 - a model using a categorical dependent variable with two classes of postoperative length of stay in ICU;

3 - a model using a categorical dependent variable with three classes;

4 - a model using a categorical dependent variable with six classes.

For the second digit:

1 - a model based on the set of preoperative predictor variables only;

2 - a model based on the set of both pre-, intra- and postoperative predictor variables.

Each node of the classification tree for the continuous type of dependent variable shows the average value and standard deviation of ICU length of stay for the group of cases represented by the node and total number of cases in the node. Figures 6 and 7 show the models SM94_11 and SM94_12, which have ICU LOS as the continuous dependent variable. SM94_11 is developed using only preoperative predictor variables, while SM94_12 uses the pre-, intra-, and postoperative predictor variables.

Above each node on the classification trees where split was made on categorical predictor, categories of the predictor on which the split was made and the number of cases for each category are shown. KnowledgeSEEKER uses ??? to designate a missing value. This value is treated just like any other valid value in the partitioning process and it is merged with the category that it most closely resembles (if any).

Figure 6: SM94_11 model

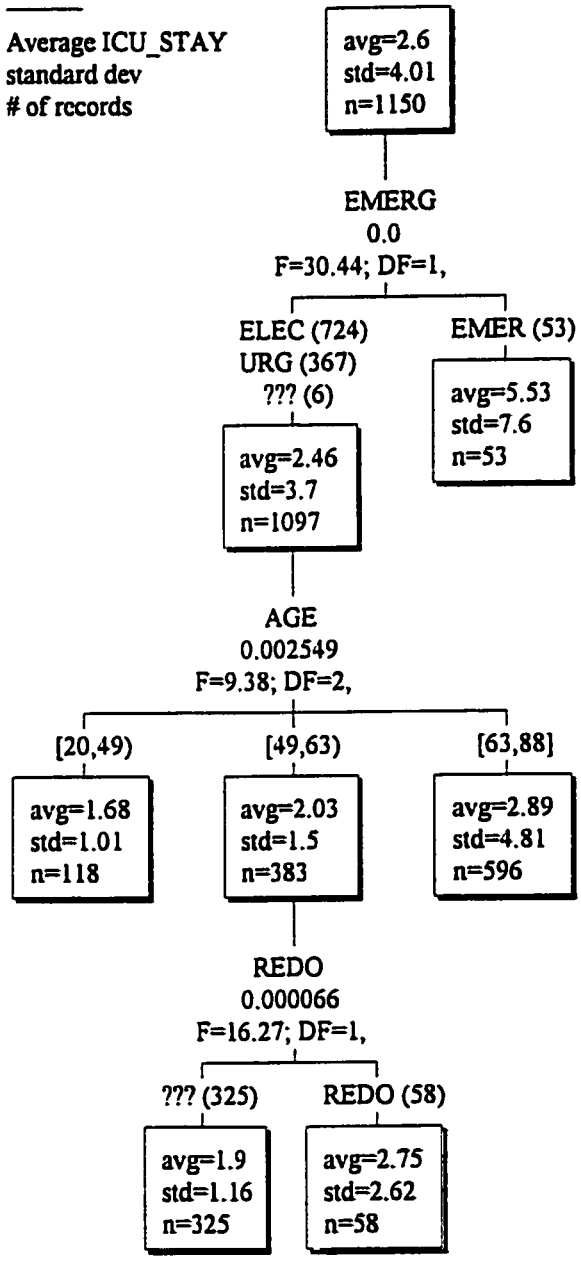


Figure 7: SM94_12 model

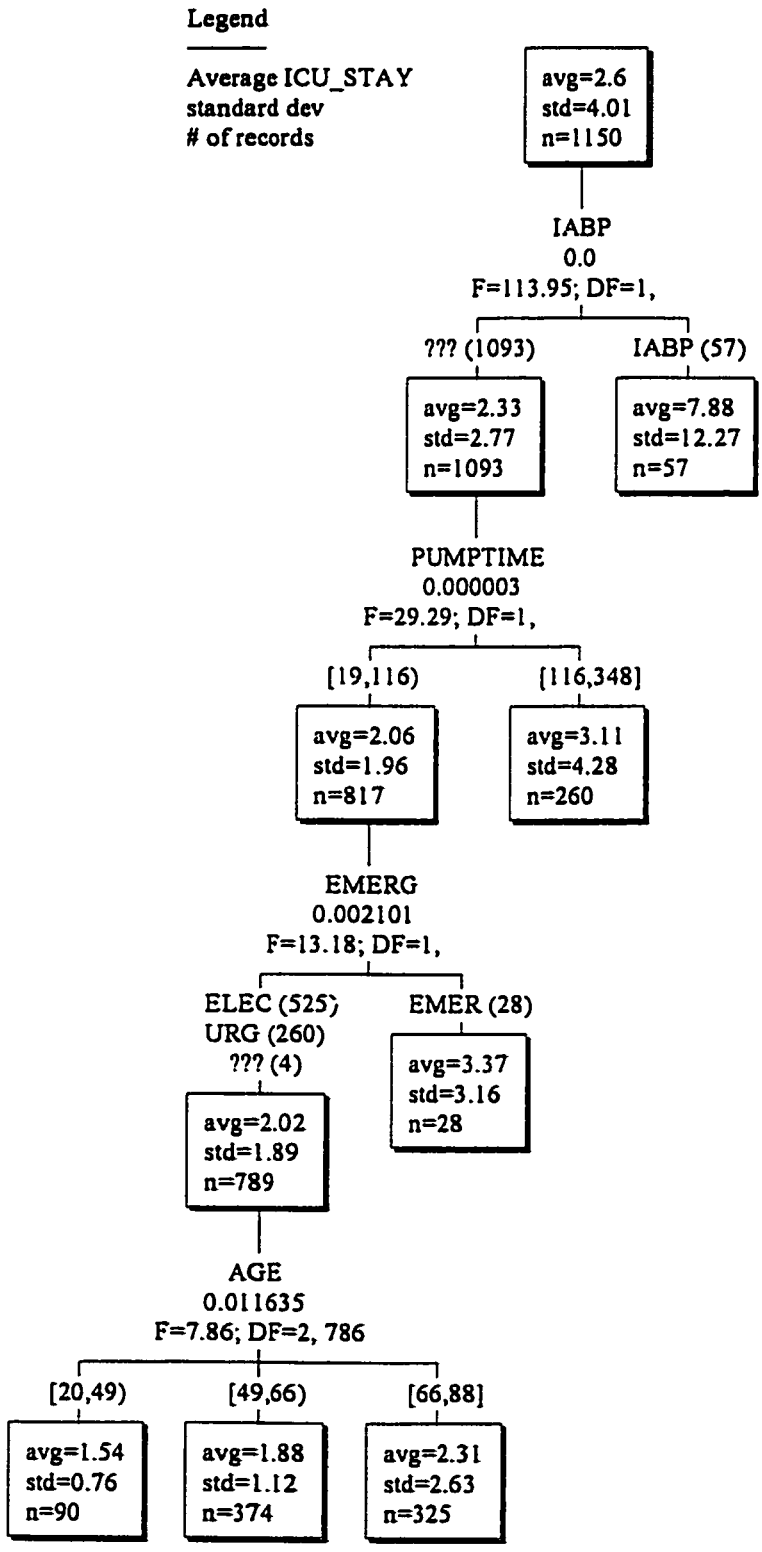


Figure 6 tells us that the first, most important split is for emergency status. The 53 emergency patients had an average ICU LOS of 5.53 days, with a standard deviation of 7.6 days, while the 1097 non-emergency patients had a much shorter ICU LOS on average (2.46 days) and smaller standard deviation. The split is highly significant with a p-value of "0.0" (ie. Less than 0.000000005). Only the non-emergency patients have any further splits, with the 2nd level being based on age. Three age categories are created from the raw age values, with the expectation that these categories are best for the 1097 patients. (Note that in other analyses the age categories are not the same!) Only the middle age group is split any further into those who were having the operation repeated from some previous time.

Figure 7 is based on the same 1150 patients as in figure 6, but has added the predictors for intra- and postoperative variables. IABP (Intra Aortic Balloon Pump) is the most significant, followed by the time the patients were on the heart bypass pump (pumptime). These are then followed by emergency and age.

Each node of the classification tree for the categorical type of dependent variable shows the total number of cases grouped in the node and the distribution of cases between classes of ICU length of stay. Figure 8 shows the model SM94_21 which has ICU LOS categorized into 2 classes, using only preoperative predictor variables.

Figure 8: SM94_21 model

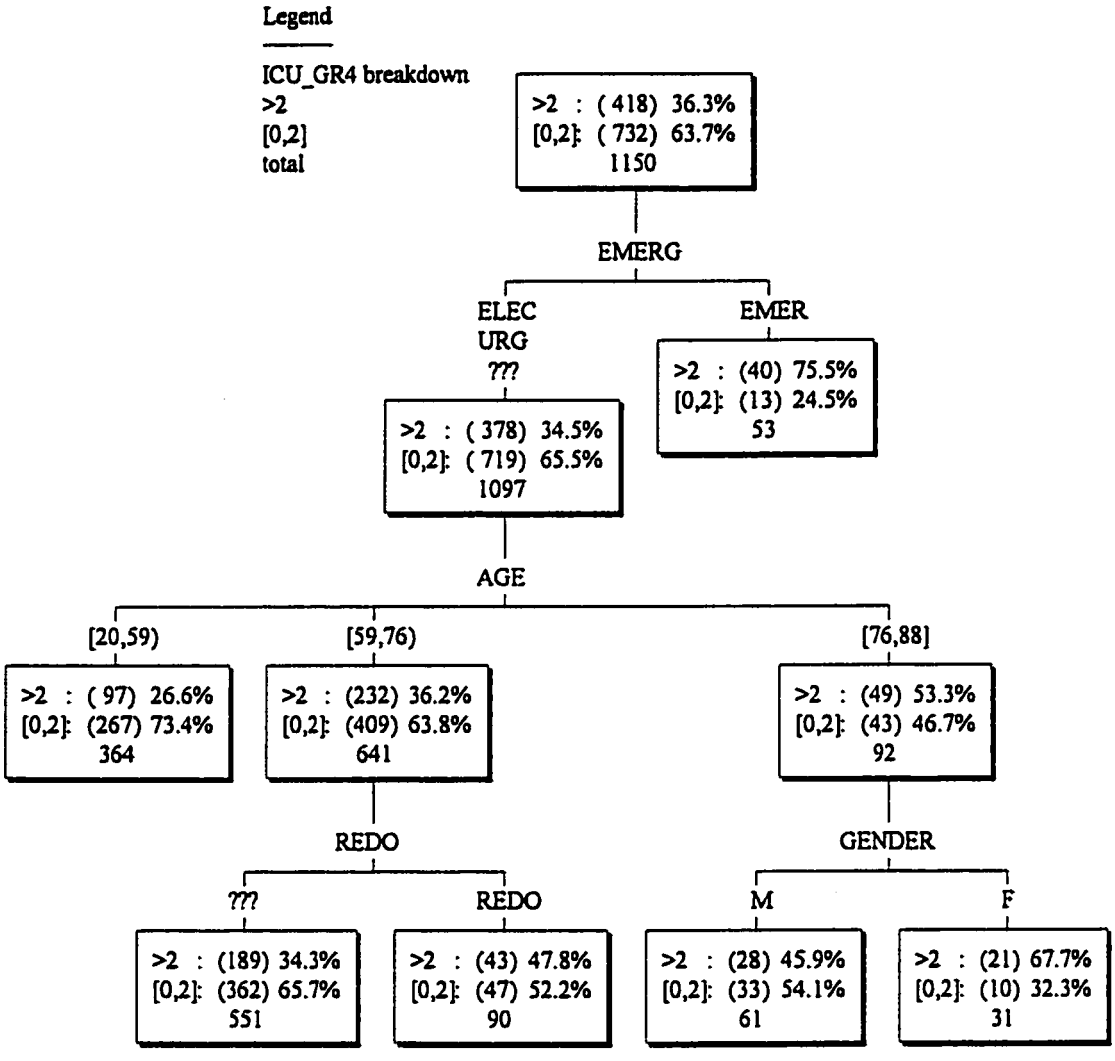


Figure 8 is similar to Figure 6, with the exception that the ICU LOS is shown in the two categories. In the 1150 patients, almost 64% had an ICU LOS between 0 and 2 days. The percentage changes drastically for emergency patients, of whom only 25% had stays of 0 to 2 days. For the non-emergency patients age is the next level of splitting. Two of the three age groups show a further split. The middle age group is split on REDO, similar to figure 6, but the oldest age group is split on gender, which shows very different ICU LOS probabilities for stays of 0 to 2 days, with males having much greater probability of staying shorter times.

On every split KnowledgeSEEKER shows the name of the independent variable on which split was produced, the *p* value (which is the probability that chosen predictor is completely

unrelated to the dependent variable), and also the F statistic with degrees of freedom on the models with continuous type of dependent variable, and the Chi-square statistic on the models with categorical type of dependent variable.

Another part of the results is the estimation of the test error based on 1995 test set, which is presented in Appendix 4B as generated by KnowledgeSEEKER for each model. The "Learning Sample" column gives the distribution of the dependent variable for 1994 year patients. the "Test Sample" column is for the 1995 year patients, and the "Difference" column shows the extent of the shift between 1994 and 1995.

In Table 5 (see page 63) characteristics of each model with categorical dependent variable are presented and the classification accuracy generated by KnowledgeSEEKER is compared with the information score per answer [20] for each model.

5.4.2 How the classification trees were grown, interpretation of the results.

The KnowledgeSEEKER exhaustive partitioning algorithm with significance level of 0.05 (reduced by the Bonferroni adjustment) was used to produce the trees.

The exhaustive method finds the most significant grouping of codes and collapses them together. Other significant groupings are also identified and collapsed together. This produces a range of alternative tabular relationships for each predictor in the data set. The best relationship - the one that produces the highest level of statistical significance - is picked to display the association between the predictor and the dependent variable when this method is selected. The algorithm is as follows [2]:

1. Select the pair of categories of the predictor variable which is most similar on the basis of the smallest F or Chi-square value. Collapse these categories into a group, labelling the result as possible split number i+1 (1 for first iteration). Repeat step 1 until only two groups are left.
2. Calculate the significance (probability) of each of the groups that have been formed in step 1, using the F or Chi-square test. The group with the lowest probability (highest significance) is taken to be the 'best' split of the node for that predictor variable.
3. Determine if this 'best' split passes the significance test threshold (0.05 in our case).

If the 'best' split is not found to be significant then no significant splits using the predictor variable can be found. The procedure is repeated for each predictor variable.

Since the data were noisy, at each step several variables turned out to be significantly associated with ICU length of stay. Therefore the task was to select the variable for the split that

satisfies not only the significance criterion imposed by the KnowledgeSEEKER algorithm but also satisfies the following subjective criteria as much as possible:

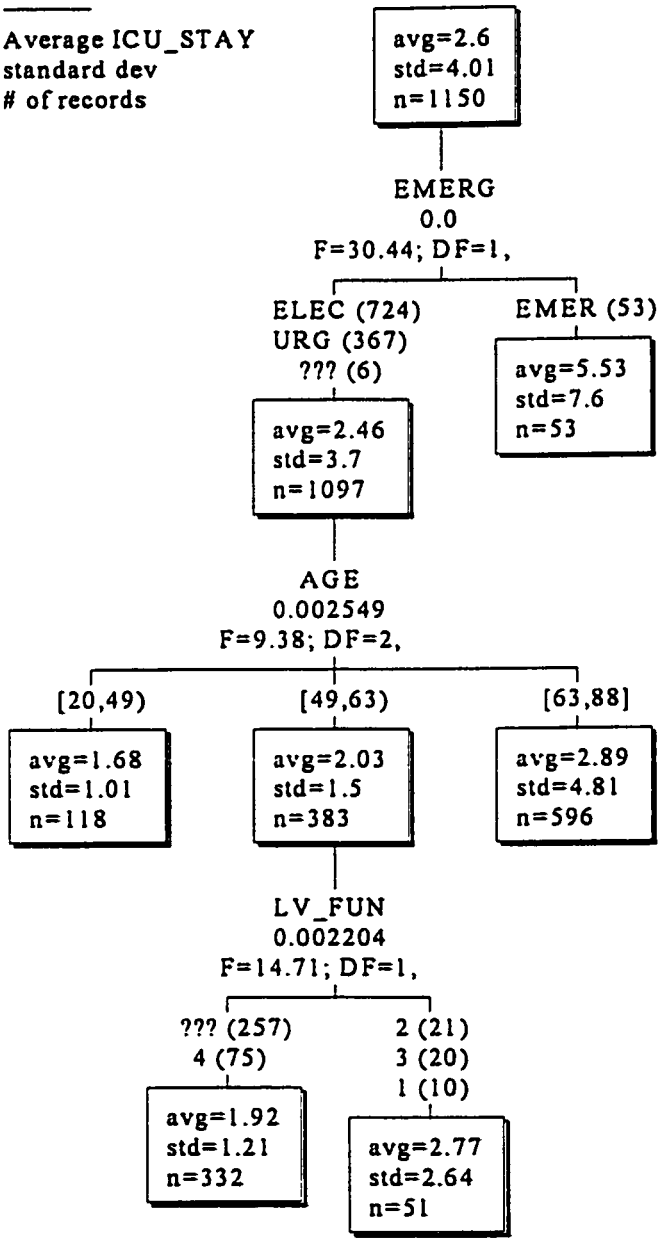
1) split based on variable results in the formation of groups which are meaningful from the clinical and problem prospective;
2) split based on variable leads to a manageable number of groups which have potential to deal effectively with new cases;
3) split introduced by the variable does not increase overall error of the tree as calculated by KnowledgeSEEKER.

Since only significance testing is automated by the KnowledgeSEEKER algorithm, we manually controlled the tree growing process for different scenarios.

For example, for preoperative models on the first split, out of several alternative choices emergency was selected as the variable for the split because it was not only the most significantly associated with ICU length of stay but, also from the point of view of scheduling, urgency categories have to be separated first since the most urgent cases have to be scheduled first despite the predicted ICU length of stay. Therefore on all the classification trees based on the predictive variables available preoperatively the first split was produced on urgency category.

Some variables important from the clinical point of view and known as predictors of ICU length of stay from other studies such as Left Ventricular Function (LVF), were not selected for any splits because groups introduced by KnowledgeSEEKER as significantly different in respect to ICU length of stay were not intuitive and therefore did not have the potential to deal effectively with new cases. For example, during modeling on SM94_11 group of elective and urgent patients of age between 49 and 63 years could be potentially split by variable LV_FUN as seen on Figure 9. The split results seem counterintuitive because from practice is known that the greater the class of LVF the more probability that the patient will stay longer in ICU. The surprising feature of the lowest level split is that KnowledgeSEEKER combines patients with LVF at level 4 with those patients for whom LVF was not recorded (???), but presumably was 0 for most of them. Patients with level 4 LVF would be the ones with the worst function prior to the operation, and would be intuitively expected to die in a short period of time or to stay for a long time in the ICU. This might tend to increase the average LOS, and would increase the standard deviation, neither of which was found in this node in this model. Thus, the decision was made not to accept a split on this variable in this model.

Figure 9: Potential split on variable LV_Function



Other variables, particularly associated diseases, found to be significantly associated with ICU length of stay, were not chosen for splits because of the small numbers of patients having those

diseases in the learning set. These variables therefore had no good potential to deal effectively the test sets where the number of similar patients might be very different (even zero).

For the models based on the set of variables available both preoperatively and postoperatively, postoperative variables were the most significantly associated with the postoperative length of stay in ICU. Thus, in the models when IABP and pumptime were added to the nine variables available preoperatively (see outputs for SM94_12.dat, SM94_22.dat, SM94_32.dat and SM94_42.dat in Appendix 4A) IABP became the variable for the first split. All cases therefore were partitioned into two groups: one group consisting of cases which had postoperative IABP and another group consisting of cases which did not have postoperative IABP. The majority of cases did not have IABP and were divided further into groups with different pattern of postoperative ICU length of stay based on how long they were on pump during operation (see outputs for SM94_12.dat and SM94_42.dat in Appendix 4A).

When medication on transfer from OR to ICU were added to the analysis (see outputs for LG94_12.dat, LG94_22.dat, LG94_32.dat and LG94_42.dat in Appendix 4A) some of them turned out to be the most significantly associated with postoperative length of stay in ICU. For example, B1142 - dopamine on weaning from bypass and B1225 - amrinon, B1223 - dobutamine, B1222 - dopamine on post-bypass were significantly associated with postoperative ICU length of stay and were selected for partitioning on different levels of the classification trees.

After the first split the choice of each new variable for the partitioning was based on the error of the overall tree estimated by KnowledgeSEEKER. When on the new split several potential variables were introduced by KnowledgeSEEKER as significantly associated with postoperative length of stay in ICU in order from most significant to less significant, for each of them estimated error was calculated for the whole tree and the splitting variable was selected which introduced the smallest relative error for the model with continuous type of dependent variable or the smallest misclassification error (biggest accuracy) for the model with categorical type of dependent variable. Therefore at each step classification tree could be considered as optimal. However, we can not say that the overall classification tree is optimal. This criterion of estimated error was necessary because it was found that estimated error of the classification tree growing in the "uncontrolled" mode is lower than if the tree was optimized ("controlled") for this criterion on each split. Consequently, the stopping rule for growing the classification tree was when no more variables were found by KnowledgeSEEKER to be significantly associated with postoperative length of stay in ICU or none of the potential ones decreased the estimated error for the tree.

Most of the classification trees presented as results in Appendix 4A have very few splits (see for example outputs for models with only nine preoperative variables used for the analysis: SM94_11.dat, SM94_21.dat, SM94_31.dat, SM94_41.dat). This results from the fact that either no other variables decreased the estimated error of the trees, or there was a marginal decrease, but the small numbers of patients in the subgroups meant that the same subgroup might not exist in the next year and therefore the splits were not stable enough to be validated in the test set.

5.4.3. Validation of the results.

Estimation of the error for all the classification trees was performed on the equivalent 1995 test data sets. The results are presented in Appendix 4B. For all models estimated error for the test set was worse (higher) than for the learning set (error estimated on 1994 learning set is not shown).

Since the set of predictors chosen for the analysis of postoperative length of stay in ICU consists of variables which were already found by other researchers and physicians as potentially associated with postoperative length of stay in ICU, the task was basically to find a good classification rule based on those variables for predicting postoperative length of stay in ICU. In order to say how good the predictive rule is, criteria must be defined. Thus, for the models with continuous type of dependent variable, the relative error/variance explained could be a base for comparison of different classification trees.

From the Appendix 4B for models with continuous type of dependent variable we can see that all of four classification trees are poor predictors of postoperative length of stay in ICU, because, based on 1995 validation set, at most 20% of variance was explained in the model with postoperative variables (see LG94_12.dat). Models with only preoperative predictor variables explained at most 8% of variance in postoperative length of stay in ICU in validation set (see LG94_11.dat).

In the models with the categorical dependent variable the misclassification error and accuracy were calculated in 1995 test data set for models with different number of classes of postoperative length of stay in ICU. As it could be seen from results in Appendix 4B the biggest accuracy rate among all models was 77% for the classification tree SM94_22.dat. The smallest calculated accuracy rate was 44% for the tree LG94_41.dat. In fact none of them is a good classifier. The misclassification/accuracy rate for categorical type of dependent variable is misleading because it does not take into account prior probabilities of patients belonging to the various classes of postoperative length of stay in ICU. Thus, prior probability of postoperative

length of stay in ICU less than or equal two days in the model with only two classes is 76%. The model results give a total accuracy rate of 77% which is only 1 percentage point higher than the prior probability of majority class. This implies that the predictors did not add much information to the initial naïve model where nothing is known about the predictors.

In search for a better criterion we found an information gain criterion which was previously described theoretically in the section 5.2.5. Notice that the evaluation of the information score of an answer for classification trees is needed because in all models the majority class of ICU length of stay is much bigger than other classes. This criterion was applied to the results of the analysis with the categorical type of dependent variable.

In Table 5 the classification accuracy and information score per answer from [20] compared for all the classification trees. The characteristics of each classification tree are also presented in Table 5. In all models the majority class of postoperative length of stay in ICU is much more probable than the other classes. Therefore as already was mentioned before the classification accuracy that was achieved for all the classification trees can be achieved in all those models with a default classifier which classifies always into the majority class.

Table 5: Comparison of performance of different classification trees.

model	no. pred. var.	no. classes	no. test instances	prob. of majority class	accuracy	entropy [bits]	aver. inf. score [bits]	rel. inf. gain score
Preoperative models (by the name of output file)								
SM94_21	9	2	1092	73%	74%	0.8471	0.0029	0%
SM94_31	9	3	1092	73%	73%	1.0919	0.0113	1%
SM94_41	9	6	1092	44%	45%	2.0436	-0.1003	-5%
LG94_21	31	2	656	70%	70%	0.8801	-0.0320	-4%
LG94_31	31	3	656	70%	70%	1.1291	0.0094	1%
LG94_41	31	6	656	44%	44%	2.0698	-0.1628	-8%
Postoperative models (by the name of output file)								
SM94_22	11	2	1092	73%	77%	0.8471	0.0045	1%
SM94_32	11	3	1092	73%	75%	1.0919	0.0570	5%
SM94_42	11	6	1092	44%	46%	2.0436	-0.0656	-3%
LG94_22	45	2	656	70%	76%	0.8801	0.0761	9%
LG94_32	45	3	656	70%	72%	1.1291	0.1009	9%
LG94_42	45	6	656	44%	45%	2.0698	0.0491	2%

Entropy in Table 5 is the expected amount of information required for classification of one instance and was calculated by formula (5) in section 5.2.5. From the properties of the information gain evaluation criterion [20], if the probability of the class returned by the classifier is 1 for every case (patient), then the *average* information score (see formula (3) in section 5.2.5) of the system's answer is equal to the expected necessary information for the correct classification of one instance (ie. the entropy), and the *relative* information score (see formula (4) in section 5.2.5) is equal to 1. However, if the probability of the class returned by the classifier is 0 for every case (patient), then *relative* information score in general is worse than -1. As Kononenco and Bratko explain in [20]: "The lack of symmetry shows that the information score of correct classification is not equal to the penalty of incorrect classification. This lack of symmetry should not be regarded as anomaly. It merely reflects the asymmetry in the classification problem itself: It is not equally difficult to predict the correct class as to predict an incorrect class." Finally, if the classifier returns correct answers with the same frequency as the prior probabilities of the classes would predict, then the *average* and *relative* information scores are equal to 0.

Entropy can be used to evaluate the difficulty of a decision problem. A problem domain where the dependent variable has more classes is more difficult than a problem domain with fewer classes, and this is shown by a greater entropy. Also the entropy in a problem domain with equal prior probabilities of the dependent variable classes is greater than the entropy in a problem domain with unequal prior probabilities [20]. We have experimented with several classification problems where the dependent variable has different numbers of classes and different distributions of prior probabilities, and therefore different difficulties of classifying cases (patients) into the correct class, as shown by the entropy measure in Table 5. The *relative information* score can be used to compare the performance of the different classification trees.

The results in Table 5 show that the classification trees are not very good classifiers since their relative information scores are not much different from zero. This fact confirms our supposition that the accuracy rate by itself, without accounting for prior probabilities of the dependent variable categories, gives misleading information. The criterion of the *relative information* score for comparison of different classifiers gives a more accurate picture. From Table 5 we see that for the trees with only preoperative predictors, the relative information score is highest (at 1%) for models with three classes of postoperative length of stay in ICU. In the models with six classes of postoperative length of stay in ICU, the information score is negative which means that the classification trees are misleading. On average the probability returned by the classifier for a case being in the correct class is lower than the prior probability for that class. Scenarios with six classes of postoperative length of stay in ICU are particularly difficult ones (see entropy in Table 5), because they have *many classes* of postoperative length of stay in ICU and because there is *less inequality in prior probabilities* of the classes. Since classification trees were not optimized for the information gain criterion, it is difficult to say which particular variables cause the misleading information in the classifiers. The *accuracy rate* calculated for those classification trees does not seem worse than for those classifiers which look more promising by information criterion.

Surprisingly, for classification tree LG94_21.dat (see Appendix 4A) information score is negative although the accuracy is not lower than the prior probability of the major class. Investigation of causes of such a low information score showed that the split on CHF (congestive heart failure) variable was not validated in the 1995 test set: all cases with 'Y' for CHF in the test set had postoperative length of stay in ICU less than 2 days. Since the training set had a probability of 90% that ICU LOS was greater than 2 days, the information score for cases with

CHF became so low that it affected *average* information score per answer. Because of the small number of patients with CHF, the problem with the classification tree could not be detected by comparing accuracy rate with the prior probability of the major class for that classification tree: the difference between them is not bigger than for other classification trees.

Now we turn our attention to those classification trees which score higher on the information gain criterion. From the models with only preoperative predictors, the classification trees SM94_31 and LG94_31 with three classes of postoperative length of stay in ICU get the highest relative information score of 1%. Clearly this is not sufficient to say that we can use those classifiers to prognosticate preoperatively a patient's postoperative length of stay in ICU. Although more preoperative variables were used for the growing of the tree LG94_31, neither of associated diseases became variable for the splits on the tree based on the tree growing criteria that were described above.

In the models where postoperative variables were added, a 5% *relative* information score was achieved by the classification tree with eleven predictors and three classes of postoperative length of stay in ICU (model SM94_32). A 9% *relative* information score was achieved in models with two and three classes of postoperative length of stay in ICU where postoperative medication on transfer were added to the analysis (models LG94_22 and LG94_32). Although postoperative predictors clearly improve prediction of postoperative ICU length of stay, it is still only marginal improvement from the scenario where nothing is known about predictors.

We would conjecture that the lack of predictive power from the models is tied to the non-stationarity of the data reported previously. Clearly, further experimentation with a non-stationary approach should be implemented and tested. For example, an approach similar to the "rolling" schedule approach to planning could be tested. In such an approach, a predictive model such as KnowledgeSeeker could be applied to, as an example, the most resent three months of data, and the results could be used for a one or two month period, after which the process would be repeated. This proposal is only suggestive to illustrate how non-stationarity might be handled.

5.5 Summary and discussion.

From the results of the analysis we can conclude that (i) adding information about associated diseases known preoperatively does not improve prediction of postoperative ICU length of stay because having small number of cases they do not have potential to be validated on new cases; (ii) splitting cases into more classes of postoperative ICU length of stay adds difficulty in classification problem and produces unsatisfactory classifiers; (iii) postoperative variables improve information about postoperative length of stay in ICU.

To assess and compare the performance of different classifiers one should be able to estimate the difficulty of decision problems. Entropy takes into account number of classes and prior probabilities: the greater the entropy of the model problem domain, the more difficult it is.

The amount of available information also depends on the number of available training instances for constructing a classification tree. For example, if there is not sufficient number of instances to make significant statistical choices then perfect classification could not be achieved. Therefore the classification model does not give the information necessary to correctly classify cases (patients), or it may not be stable enough to support validation in the test set, as observed in the model with associated diseases.

Assuming we have perfect classification technique and sufficient number of cases required by this technique, then incompleteness of domain could lead to producing a classification tree which does not distinguish very well between classes of cases. A domain is said to be incomplete if the available predictors do not suffice to completely distinguish between classes of the dependent variable. As a consequence, there could be overlapping between cases - different classes of the dependent variable correspond to the same predictor. Using this classifier on a test sample could not give a large information score, and even could reveal that the classifier is misleading as it was for classification tree LG94_21.dat.

Therefore we see that a low *relative* information score could become such because of several reasons. The fact that information score for a particular classification tree is low is unfortunate but it reassures us, because this measure does not give us misleading information, as for example the *accuracy rate* does. At the same time the *relative* information score gives us an appreciation of how much particular classifier added to the information, given the complexity of the domain.

The question arises as to why the significant degree of non-stationarity of postoperative length of stay in ICU exists. In addition to the usual statistical problems of incomplete and/or inaccurate data, there may very well have been structural changes in the system which relate to public policy in the health care area, including stricter financial constraints. The decrease in average length of stay from 2.6 to 2.4 days is certainly suggestive of these kinds of causes. Moreover, if indeed the OHI is operating in a period of identifiable structural change, a scheduling system should certainly take account of these changes, probably through the provision of regular input from the OHI consultants.

To summarize, it is recommended strongly that as the next step, the predictive approaches presented in this investigation be applied carefully, taking into account data non-stationarities. Looking into the future, it is recommended that any scheduler be provided with ready mechanisms to integrate "user" insights, including those which deal with anticipated structural changes in the institutional system.

6. Conclusions, limitations and recommendations.

This work on the problem of scheduling of patients for heart surgery at the University of Ottawa Heart Institute was guided by the set of questions posed in Chapter 1, section 1.2. In order to find answers to the problems which motivated this work we set several objectives hoping to approach the solutions of the problems. Reaching our objectives by the end of this work we (i) understood how the current system is operating, (ii) developed and assessed models for predicting length of stay in ICU after cardiac surgery.

The risk stratification models developed by other researchers are stratifying patients into only two categories, namely, long and short length of stay in the ICU. Our hope was that more precise prediction of ICU LOS is possible, that is more categories of ICU LOS could be more useful for scheduling purposes. Having considered two categories of ICU LOS, we also investigated models with three and six categories, where long length of stay was further subdivided. Because of the very small probability of the long ICU LOS, last models became very difficult to get good prediction from. Finally, we investigated case with ICU LOS as a continuous variable, which would be preferable for scheduling purposes. However, the more precise subdivision of ICU LOS, is the less informative generally the models are.

As this study was an initial attempt to approach a very complex situation there are several limitations which prevented us from getting final recommendations for solutions to the problem. The major limitation which is common to all studies based on data analysis is the quality and completeness of the available data. The study was retrospective and data was collected from the sources which were not initially intended to be used for this kind of analysis. As a result accuracy and completeness of the data is far from perfect. Although the analysis tool that was chosen for the developing of predictive models can easily deal with missing data and outliers, the results could be clearer if data were specifically collected for the purposes of this analysis. One of the recommendations arising from the study is that definition and collection of the data at the OHI should be standardized.

We were not able to find from the developed predictive models one that would satisfactorily predict length of stay in ICU using data available preoperatively. There are several factors that could make the predictor not informative enough. Those factors were discussed in the conclusion of Chapter 5. Also, there were several assumptions and simplifications along the work,

which could decrease information content of the predictive models. We did not consider patients who had repeat operation within the same admission and those who have been returned to ICU after they have been transferred to the ward. In the follow-up studies this could be fixed by including those cases and counting a second entry as new case.

Our data sets were not homogeneous in terms of types of surgeries as were data sets in studies mentioned at the beginning of this work. Therefore in the future study it would be useful to make a sensitivity analysis on how adding different types of surgeries changes predictive ability of the models.

The exclusion of LVF from predictive models could also decrease their information content, because LVF is one of the variables that were considered by previous studies. It seems that the problem here is with not accurate enough data, therefore this variable needs to have special attention in the future works.

In fact, the extent of incomplete inaccurate data is a major problem for this project. Since the source datasets were developed independently by different groups within the OHI for different purposes, and are not managed as an integrated dataset, there are major areas where similar data fields have extremely different values. One of the OHI consultants indicated which fields/files were most accurate. However, the more appropriate solution for the OHI is to define and manage an integrated set of files. As well as collecting data that are more accurate overall, it would be important to record the date (and perhaps time) of changes in each patient's clinical status. For example, the urgency category in PACCN does not seem to get updated, and frequently disagrees with the urgency category at the time of operation. There is no indication whether the surgeon has well defined criteria to define the urgency category at the time of operation.

Knowing the fact that the average ICU LOS dropped significantly from 1994 to 1995 the use of non-stationary predictive models as mentioned before might greatly enhance the predictive accuracy.

We based development of the predictive models using KnowledgeSeeker on the variables that were found by others to be the significant predictors of ICU LOS. However our goal was to find models that would be useful for scheduling purposes, that is the predictions of ICU LOS based on the models should be much better than physicians can do it.

While developing models with KnowledgeSeeker we found that a very important question is about criteria that would show how useful the model is. In spite of the already discussed problems with data it seems that our validation criteria give a good idea about how informative the

models are. We could not compare our results with results of other researchers because of the difference in criteria. Area under ROC that was used in other studies for validation of results could not be applied in cases with more than two categories and probabilistic answers.

It is suggested for further research to investigate and compare different criteria. Information based criterion used in this work could be further studied and extended to reflect not only the fact that predicted value is not in the right class but also how far it is from the correct class.

It is suggested for the future works to compare models developed on OHI data with models developed by other researchers. One of the way of doing that would be using our approach for developing predictive models with other researcher's (e.g. Tu) data. The other way would be to see if our data would support other researcher's models. Also for scheduling purposes it is very important to evaluate how any models perform when compared with physicians' predictions.

In general, this work aimed at exploring the problem and system, rather than testing hypothesis. It is the author's hope that the reader finds this work useful and worth continuing. Comments and suggestions are always welcome.

Glossary of Abbreviations.

A

ADT. Admission-Discharge-Transfer data base

AID. Automatic Interaction Detector

B

BSA. Body Surface Area

C

CABG. Coronary Artery Bypass Grafting

CABS. Coronary Artery Bypass Surgery

CAD. Coronary Artery Disease

CCS Class. Angina class (by Canadian Cardiovascular Society)

CHF. Congestive Heart Failure

COPD. Chronic Obstructive Pulmonary Disease

CRFC. Cardiac Reference Centre

CRR. Cardiac Recovery Room

CSU. Cardiac Surgical Unit

CVA. Cerebral Vascular Accident

I

IABP. Intra-Aortic Balloon Pump

ICU. Intensive Care Unit(s)

ICU LOS. Length of stay in the Intencive Care Unit(s)

L

LVF. Left Ventricular Function

N

NYHA Class. Heart failure class (by New York Heart Association)

O

OHI. Ottawa Heart Institute

OR. Operating Rooms

P

PACCN. Provincial Adult Cardiac Care Network

PTCA. Percutaneous Transluminal Coronary Angioplasty

PVD. Peripheral Vascular Disease

R

CRR. Cardiac Recovery Room

S

STS. The Society of Thoracic Surgeons

T

TMR. Total Medical Record

U

URS. Urgency Rating Score

V

VSD. Ventricular Septal Defect

REFERENCES

1. Andreoli K.G., Zipes D.P., Wallace A.G., Kinney M.R., Fowkes V.K. "Comprehencive Cardiac Care." 1987
2. Biggs D., B. de Ville and Suen E. "A method of choosing multiway partitions for classification and decision trees", Journal of Applied Statistics18, 1, 1991: 49-62
3. Blake J.T., Carter M.W. "Surgical Process Management: A Conceptual Framework." To appear in Surgical Services Management.
4. Brieman L., Friedman J.H., Olshen R. A. and Stone C.J., "Classification and Regression Trees." (1990)
5. Centor R.M., Keightley G.E. "Receiver Operating Caracteristics (ROC) Curve Area Analysis using the ROC ANALYZER." SCAMC Proceedings, 1989: 222-226
6. Cox, J.L., "Ethics of queuing for coronary artery bypass grafting in Canada." Canadian Medical Association Journal 1994, 151(7): 949-953
7. Davies, R., "An assessment of Models of a Health System." Journal of the Operational Research Society 1985, Vol. 36, No. 8: 679-687
8. Davies, R., "Simulation for planning services for patients with coronary artery disease." European Journal of Operational Research 71 (1994): 323-332
9. Edwards F.H., Clark R.E., Schwartz M., "The Society of Thoracic Surgeons National Database Experience." Ann Thorac Surg. 1994: 57:12-19
10. Fetter, R.B., Shin, Y., Freeman, J.L., Averill, R.F., Thompson, J.D. "Construction of Diagnosis-Related Groups." Medical Care February 1980, Vol. XVIII, No.2, Supplement
11. Fitzpatrick, K.E., Backer, J.R., Dave, D.S., "An application of computer simulation to improve scheduling of hospital operating room facilities in the United States." International Journal of Computer Applications in Technology 1993, Vol.6, No 4: 215-224
12. Hannan, E.L., Kilburn, H.Jr., O'Donnell, J.F., Lucacik, G., Shields, E.P. "Adult Open Heart Surgery in New York State: an analysis of risk factor and hopital mortality rates". JAMA 1990; 264: 2768-2774
13. Hawkins D. M. and Kass G.V., "Automatic interaction detection", in: D.G. Hawkins (ed.), Topics in Applied Multivariate Analysis (Cambridge, Cambridge University Press, 1982)

14. Higgins, T.L., Estafanous, F.G., Loop, F.D., Beck, G.J., Blum, J.M., Paranandi, L."Stratification of Morbidity and Mortality Outcome by Preoperative Risk Factor in Coronary Artery Bypass Patients: a clinical Severity Score." JAMA 1992; 267: 2344-2348

15. Higginson A.J., Cairus J.A., Keon W.J., Smith E.R. "Rates of Cardiac Catheterization, Coronary Angioplasty and Open-heart Surgery in Adults in Canada." Canadian Medical Association Journal 1992, 146(6): 921-925

16. Hoyt, R.E., Lay,C., University of Ottawa, "Neural Network Cost Estimates for Heart Bypass Surgery." Working Paper 94-23.

17. Kao, E.P.C., "Study of patient Admission Policies for Specialized Care Facilities." IEEE Transactions on Systems, Man and Cybernetics 1974, Nov., Vol. SMC-4, No. 6: 505-512

18. Kass G.V., "An exploratory technique for investigating large quantities of categorical data." Applied Statistics 29, 2, 1980: 119-127

19. Kass G.V., "Significance testing in automatic interaction detection." Applied Statistics 24, 2, 1976: 178-189

20. Kononenko I., Bratko I. "Information-Based Eveluation Criterion for Classifier's Performance." Machine Learning 1991, No. 6: 67-80

21. Kropp, D.H., Hershey, J.C., "Recursive Optimization- Simulation Modelling for the Analysis of Ambulatory Health Care Facilities." Simuletter 1979, Vol. 10, No. 4: 43-46

22. Lay C.M. Disease Costing in an Ambulatory Clinic: Disease and Physician Profiles and the Selection of Patients for Review." Ph. D. dissertation, 1978

23. Lay C.M., Broyles R.W. "Statistics in Health Administration." 1980

24. Magerlein, J.M., Martin, J.B., "Surgical demand scheduling: a review." [Review] Health Services Research, 1978,Winter;13(4): 418-33

25. Morgan, J.N., Sonquist, J.A., "Problems in the Analysis of SurveyData, And A Proposal", Journal of the American Statistical Association, 58, June, 1963: 426.

26. Murthy K.V.S. "On Growing Better Decision Trees form Data." Ph.D. dissertation, 1995

27. Naylor, C.D., Baigrie, R.S., Goldman, B.S., Basinski, A. "Revascularisation Panel and Consensus Methods Group: assessment of priority for coronary revascularisation procedures." Lancet 1990; 335: 1070-1073

28. Naylor, C.D., Levinton, C.M., Baigrie, R.S., Clinical Epidemiology Unit, Sunnybrook Health Science Centre, Toronto, Ontario, Canada. "Adapting to waiting lists for coronary revascularisation. Do Canadian specialists agree on which patients come first?" Chest 1992 Mar;101(3):715-22

29. Naylor, C.D., Levinton, C.M., Wheeler, S., Hunter, L., Clinical Epidemiology Program, Sunnybrook Health Science Centre, Toronto, Canada.. "Queuing for coronary surgery during severe supply-demand mismatch in a Canadian referral Centre: a case study of implicit rationing." Social Science & Medicine 1993 Jul;37(1):61-7

30. Naylor, C.D., Morgan,C.D., Levinton, C.M., Wheeler, S., Hunter, L., Klymciw, K., Baigrie, R.S., Goldman, B.S., Clinical Epidemiology Unit, Sunnybrook Health Science Centre, Toronto, ON. "Waiting for coronary revascularisation in Toronto: 2 years' experience with a regional referral office." Canadian Medical Association Journal 1993 Oct 1;149(7):955-62

31. O'Connor, G.T., Plume, S.K., Olmstead, E.M., Coffin, L.H., Morton, J.R., Maloney, C.T., Nowicki, E.R., Levy, D.G., Trzelaar, J.F., Hernandez, F. "Multivariate Prediction of in-Hospital Mortality Associated with Coronary Artery Bypass Graft Surgery". Circulation 1992;85:2110-2118

32. Oren T.I., Birta L.G., Abou-Rabia O. "Requirements for the Documentation of Simulation Studies." Technical report TR-89.52, 1989

33. Pritsker A., Alan B. "Introduction to Simulation and SLAM II." 1986

34. Przasnyski, Z.H., "Operating Room Scheduling: A Literature Review." AORN Journal July 1986, Vol. 44, No1: 67-79

35. Reddy R.K.T. and Bonham-Carter G.F., (1991) "A decision-tree approach to mineral potential mapping in snow lake area, Manitoba." Canadian Journal of Remote Sensing 17, 2, april, p. 191-200

36. Rosner B. "Fundamental of Biostatistics." 1990

37. Thomson, J.D., Fetter, R.B. et.al., "Case Mix and Resource Use" Inquiry, Vol.12 nom. 6, Dec. 1975: 300-312.

38. Tu, J.V., Guerriere, M.R., Information Systems Department, St. Michael's Hospital, University of Toronto, Canada. "Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery." Computers & Biomedical Research 1993 Jun;26(3):220-9

39. Tu, J.V., Mazer, C.D., Levinton, C., Armstrong, P.W., Naylor, C.D., Department of Medicine, St. Michael's Hospital, Toronto, Ont. "A predictive index for length of stay in the intensive care unit following cardiac surgery." Canadian Medical Association Journal 1994 Jul 15;151(2):177-8

40. Tu, J.V., Jaglal, S.B., Naylor, C.D. and 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 Feb.1, Vol. 91, No 3: 677-684

41. Tuman, K.J., McCarthy, R.J., March, R.J., Najafi, H., Ivankovich, A.D. "Morbidity and Duration of ICU Day After Cardiac Surgery: A Model for Preoperative Risk Assessment." Chest 1992;102:36-44

Appendix 1A.

Ottawa Heart Institute Ethics Committee Review and Approval.

OTTAWA CIVIC HOSPITAL

RESEARCH ETHICS APPLICATION FOR PATIENT RECORD (CHART) REVIEW

LOCATION:

PROJECT #:

1. TITLE OF PROJECT

Open-Heart Surgery Patients' Length of Stay: A Two-Year Review

---

2. PRINCIPAL INVESTIGATOR

Name: Professor Colin Lay, PhD
Professor Tony Quon, PhD

Signature: See attached

Tel. No.: 564-7004

Dept./Div.: University of Ottawa,
Masters of Health Administration

---

3. PRINCIPAL INVESTIGATOR AT THIS SITE

Name: Heather Sherrard BScN, MHA

Signature:

Telephone No. 761-4826

Dept/Div. Nursing - Heart Institute

---

4. APPROVAL BY INVESTIGATOR'S DEPARTMENT OR DIVISION HEAD

I have read this application, believe that the benefits of the proposed research outweigh the risks to patients or normal subjects, and support the implementation of this project.

Name: Heather Sherrard Signature: _____

Dept./Div.: Director of Nursing, University of Ottawa Heart Institute

Name: Dr. Wilbert Keon Signature: _____

Dept./Div.: Director General, University of Ottawa Heart Institute

5. CONTACT PERSON (where appropriate)

Name: Heather Sherrard Signature: _____

Tel. No.: 761-4826

RESEARCH ETHICS APPLICATION FOR PATIENT RECORD (CHART) REVIEW

LOCATION:

PROJECT #:

1. TITLE OF PROJECT

---

2. PRINCIPAL INVESTIGATOR

Colin M. Lay
Tony K. Lyon

Name: Signature: N/A

Tel. No.: Dept./Div.:

---

3. PRINCIPAL INVESTIGATOR AT THIS SITE

Name: Signature:

Telephone No. Dept/Div.

---

4. APPROVAL BY INVESTIGATOR'S DEPARTMENT OR DIVISION HEAD

I have read this application, believe that the benefits of the proposed research outweigh the risks to patients or normal subjects, and support the implementation of this project.

Name: _____ Signature: _____

Dept./Div.: _____

5. CONTACT PERSON (where appropriate)

Name: _____ Signature: _____

Tel. No.: _____

6. TYPE OF STUDY

Which of the following best describes the type of investigation proposed? Check more than one if appropriate.

( ) Clinical trial
( ) Randomized, controlled
( ) Phase 1 2 3 4 (circle appropriate one)
( ) Multicentre trial
( ) Number of centres participating 1
( ) Total number of patients to be recruited 2000 charts
( ) Number of patients to be recruited at this centre 2000 charts
( ) Case-Control study
( ) Cohort study
( ) Pilot study
( ) Sequel to previously approved project
( ) First application in humans
( ) Innovative therapy
( ) Evaluation of medical devices
( ) Study of surgical therapy
( ) Interview, survey or questionnaire
(X) Chart Review
( ) Compassionate Use
( ) Other

---

7. STATUS OF THE STUDY

	Month	Year
Anticipated Start Date at this centre	October	1995
Anticipated Duration Of Study	6 months	-----

---

8. FUNDING SOURCE Not Applicable

	Name	Amount
Granting Agency	----------	----------
Hospital	----------	----------
Hospital Foundation	----------	----------
Departmental Research Funds	----------	----------
Drug Company/Industrial	----------	----------
Other	----------	----------

RESEARCH ETHICS APPLICATION FOR PATIENT RECORD (CHART) REVIEW

9. SUMMARY OF RESEARCH PROPOSAL

.1 RATIONALE

The Ottawa Heart Institute has a waiting list of over 300 patients for open-heart surgery. While waiting lists have been examined using provincial data, there is no clear explanation for the increases. The purpose of this chart review is determine whether a relationship exists between patient characteristics and length of stay (LOS), to see if LOS can be better predicted before surgery. Better knowledge on LOS may allow for improved scheduling strategies that will maximize throughput and reduce average waiting time.

.2 CHART REVIEW PLAN

.1 Patient Population

The patient population includes all patients who underwent open heart surgery at Ottawa Heart Institute in 1993 and 1994.

.2 Conditions, demography, source

Transfer data from the admission, discharge, transfer system (ADT) will be used to calculate LOS. Comorbidity data will be used from the existing HMRI abstract database. Other clinical data such as demographics, expected morbidity/mortality, ischemia risk, coronary anatomy, left ventricular function and triage score will be obtained through the triage database.

.3 If list of cases for chart review required please specify ICD-9 terms that best describe patient population.

.3 SAMPLE SIZE

Approximately 1250 open heart surgery cases are done annually. Approximately one half are elective cases, and the remaining cases are emergent. Since emergency cases are a significant factor in scheduling strategies, all open-heart surgery cases will be included.

.4 WILL THERE BE PATIENT CONTACT?

Yes ___ No X

.5 CONFIDENTIALITY

.1 Who will have access to the record?
The coinvestigators and the research assistant will be the only individuals with access to the data.

.2 How will anonymity of subjects be maintained?
The hospital unique number will be the only identifying data collected. No other information which could be used to identify a specific patient will be collected.

NOTE: RESEARCH ETHICS COMMITTEE POLICY PROHIBITS RELEASE OF PATIENT IDENTIFYING INFORMATION FROM THE HOSPITAL.

RESEARCH ETHICS APPLICATION FOR PATIENT RECORD (CHART) REVIEW

10. STATEMENT OF UNDERSTANDING

Reviews will be conducted by members of the Ottawa Civic Hospital Medical/Dental Staff, House Staff or research assistants working for the medical staff. Projects to be conducted by external groups will be sponsored by the Chief of the Clinical Department/Division under review. If patient contact is required for the study, this must be stated in the application. If the review is being conducted in preparation for publication, the permission of the attending physician must be secured.

Due to space limitations and the need for chart control, records can only be out of file for 3 weeks. If review has not begun by this time the principal investigator will be contacted and the charts returned.

CONFIDENTIALITY

The hospital records which you will be reviewing contain information which patients have the right to expect will be held in confidence. Please review the following items concerning confidentiality and research and sign the Statement of Understanding.

1. Charts retrieved for review will be stored and used in the Health Records Department where access can be controlled.
2. Health data which is abstracted for research or educational purposes must be collected in an unidentifiable format. A unique identifier, unlinkable with any other non-health data system is recommended.
3. All working papers generated by the review will be stored in a secure area and will be shredded when they are no longer useful.

STATEMENT OF UNDERSTANDING

I have read the policy on Confidentiality, Security and Release of Information.

I understand that all health information which I may have access to is confidential and is to be dealt with in keeping with the hospital administrative policies on Confidentiality, Security, and release of Information (#2-20; #2-40).

SIGNED: Colin M. Law — Principal Investigator
Tony K. Hsiung

E. Belisario — Research Assistant

QUALITY OF CARE REVIEW

It is possible that study of these records may enable you to make a report to the Medical Audit and Tissue Review Committee regarding the quality of care of this group of patients. If you require assistance in the design of such a report, please contact the Chairman of the Medical Audit and Tissue Review Committee.

11. IMPACT ON HEALTH RECORDS SERVICE

INSTRUCTION: **Please complete the following impact analysis to determine chart review resource requirement.**

** Please note this is a quality management project - therefore no funds are available. **

7.HEALTH RECORD SERVICES				
	Description of Service	Item for Count	Cost Per Unit	Total Cost
7a	Patient Record Retrieval	# of records	First 20 records $3.00/record Next 80 records $2.75/record Next 100 records $2.70/record Excess 200 records $2.65/record	
7b	On-Line Fees (Computer Database Searches)	# of Hours	Hourly rate or part there of = $50.00	
7c	Data Retrieval (Manual) from HMRI Reports	# of Hours 2	Hourly rate or part there of = $50.00	N/A
7d	Extraction of data from patient records	# of Hours	Hourly rate or part there of = $50.00	
7e	Compilation and presentation of data	# of Hours	Hourly rate or part there of = $50.00	
7f	Other		To be determined	
	Total Health Record Services			N/A

Approved by:
Health Data & Information Service: _____

Date Approved: _____

NOTE: **THERE IS A PRE-PAYMENT POLICY FOR PATIENT RECORD RETRIEVAL**

Appendix 1B.

Description of the Collected Data.

I. Original files: **MASTERS files** (MAST93.DBF, MAST94.DBF, MAST95.DBF). These files have the most complete information for demographics, procedures, admission and discharge dates. The clinical information (clinical diagnoses, etc.) should not be used from these files. Much but not all of MASTERS files data is duplicated in the surgical part of TMR.

Data fields in MASTERS files:

ID	- patient's unique ID
BIRTHDAY	- patient's date of birth (for calculation of age at the date of surgery)
SEX	- patient's gender ("M","F")
CARDIOLOGY	- name of the cardiologist
ADDATE	- date of admission
SURGEON	- name of the surgeon
CATH_DATE	- date of catheterisation
HOSP - hospital	
AUTO_BLOOD	- autologous blood program ("yes", "no")
TOTALSCORE	- risk score
DEATH	- ("T" if true)
DATE	- date of surgery
PROCEDURE	- name of procedure
LVCLASS	- left ventricular class (1,2,3,4)
EMERGENCY	- 1-EMERG.; 2-URGENT; 3-ELECT.
IABP	- use of intra-aortic balloon pump
PUMP	- use of cardiopulmonary bypass (CPB) ("T" if true)
WEIGHT	- patient's weight
HEIGHT	- patient's height
BSA	- patient's body surface area (calculated from weight and height)
REDO	- indication of repeat cardiac operation
PROC2	- name of procedure during the second operation
DOS2	- date of second operation
PROC3	- name of procedure during the third operation
DATE3	- date of the third operation
DC	- discharge code
DDATE	- date of discharge
CTB	- "cease to breath" ("T" if true)
CDATE	- date of CTB
INDATE	- date of transfer from CRR/CSU to the ward
TO	- place of transfer from CRR/CSU
INDATE2	- date of the second transfer after intensive care

TO2	- place of the second transfer after intensive care
INDATE3	- date of the third transfer after intensive care
TO3	- place of the third transfer after intensive care
PUMPTIME	- time on cardiopulmonary bypass (CPB) (see PUMP)
ANOXIA1	- time when heart is excluded from circulation
ANOXIA2	- the same as above
CODE	- code of procedure
OR_TIME	- time when operation started
ORSTOP	- time when operation stopped
EXT	- indication of extubation
EXT_TIME	- time when patient was extubated
EXT_DATE	- date when patient was extubated
LOS	- patient's length of stay
LOSICU	- patient's length of stay in ICU
LOS2	-
LOS3	-
LATE_DEATH	- ("T" if true)

II. TMR DATA

The Anaesthesia part of TMR goes into two complementary files a9?.dbf and b9?.dbf, where question mark stands for the year of 1993, 1994 or 1995.

Data fields in the anaesthesia files are coded so that they would correspond to the order and sequence of the fields in the anaesthesia data sheet.

Data fields in a9?.dbf:

ID	- patien's unique ID
DOA	- date of anaesthesia (which should correspond to the date of surgery (SUR_DATE) in mast?.dbf files)
RESPIRATORY SYSTEM	
A521	- asthma mild
A522	- asthma moderate
A523	- asthma severe
A531	- C.O.P.D. (Chronic Obstructive Pulmonary Disease) mild
A532	- C.O.P.D. moderate
A533	- C.O.P.D. severe
A541	- pulmonary hypertension mild
A542	- pulmonary hypertension moderate
A543	- pulmonary hypertension severe

ENDOCRINE SYSTEM

A61 - normal
A621 - diabetes - diet controlled
A622 - diabetes - controlled with oral meds.
A623 - diabetes - controlled with insulin

CENTRAL NERVOUS SYSTEM

A731 - prior CVA (Cerebral Vascular Accident) - embolic
A732 - prior CVA - hemorrage
A733 - prior CVA - carotid vascular disease

OTHER MEDICAL PROBLEMS

A82 - obesity
A84 - renal disease
A85 - bleeding tendency
A86 - hepatic disease
A87 - peripheral vacular disease (PVD)
A88 - other

Data fields in b9?.dbf

MEDICATIONS ON WEANING FROM BYPASS:

B1141 - calcium
B1142 - dopamine
B1143 - dobutamine
B1144 - epinephrine
B1145 - amrinone
B1146 - other

MEDICATIONS ON POST-BYPASS:

B1221 - calcium
B1222 - dopamine
B1223 - dobutamine
B1224 - epinephrine
B1225 - amrinone
B1226 - other

The Surgery part of TMR goes into two complementary files s9?.dbf (preoperative and intraoperative data) and u9?.dbf (operative procedure data), where question mark stands for the year of 1993, 1994 or 1995.

Data fields in the surgical files are coded so that they would correspond to the order and sequence of the fields in the surgical data sheet.

Procedural data is accurate in these files (as with MASTERS.DBF) but clinical information should only be used after consultation.

Fields in s9?.dbf:

ID - patient's unique ID

PREOPERATIVE DATA:

ADMITTING DIAGNOSIS:

S22 - valvular disease
S23 - aortic aneurism
S24 - aortic dissection
S25 - congenital defect
S26 - cardiomyopathy
S27 - arrhythmia

ASSOCIATED DISEASE:

S32 - peripheral vascular disease (PVD)
S33 - cerebral vascular accident (CVA)
S35 - hypertension

INTRAOPERATIVE DATA:

PRIORITY:

S71 - elective
S72 - urgent
S73 - emergency (ASAP)

OPERATIVE DIAGNOSIS (CORONARY):

S81 - asymptomatic
S82 - angina stable
S83 - angina unstable
S84 - left main stenosis
S85 - field PTCA
S86 - acute MI
S87 - CHF (congestive heart failure)
S88 - VSD (Ventricular Septal Defect)
S89 - ventricular rupture
S810 - LV aneurysm
S811 - arrhythmia
S812 - previous graft stenosis/occlusion (repeat operation)
S813 - cardiogenic shock
S814 - other

Fields in u9?.dbf (OPERATIVE PROCEDURE)

ID - patient's unique ID

CORONARY PROCEDURE (In order of increasing complexity and risk to patient)

U41 - CABG
U42 - redo CABG

U43 - endarterectomy
U44 - closure of VSD (see prev. page)
U45 - repair of ventricular rupture
U46 - plication of LV aneurysm (plication = taking a "tuck" in the wall of the LV)
U47 - resection of LV aneurysm

BYPASS GRAFT:

U612-U662 - quality of grafts 1-6 (good/fair/poor/end)

OTHER:

U101 - cardiac transplantation
U104 - insertion of intra-aortic balloon

III. ADT (ADMISSION-DISCHARGE-TRANSFER) SYSTEM DATA.

Original files:

file name	includes patients discharged between
94.one	94/01/01-94/06/30
94.two	94/07/01-94/12/31
95.one	95/01/01-95/06/30
95.two	95/07/01-95/12/31

Data fields in the original files:

client_id - patient's unique ID
admsn_no - patient's unique admission number
adm_date* - date of admission
adm_ti - time of admission
adm_nu - nursing unit of admission
trf_date - date of transfers
trf_ti - time of transfers
nu_from - nursing unit from which patient was transfered
nu_to - nursing unit to which patient was transfered
dis_date - date of discharge
dis_ti - time of discharge
dis_cd - discharge code
los - total length of stay in the hospital

* - patients with adm_date less than 95/11/01 where adm_nu is equal to EMRE, EMCB or EMOB have their adm_date/adm_ti equals the patient's arrival date/time in emergency. Patients with adm_date greater than or equal to 95/11/01 where adm_nu is equal to EMRE, EMCB or EMOB have their adm_date/adm_ti equals the date/time the physican requested the patient's inpatient admission (i.e. can be several hours after the patient's arrival date/time in emergency).

Guide to discharge codes:

dis_cd	designated to
1	care completed
2	CTB < 48 hr after admission
3	CTB > 48 hr after admission
4	self discharged
5	coronary care elsewhere

PACCN (PROVINCIAL CARDIAC CARE NETWORK) SYSTEM DATA

Original file **paccn945.txt** contains data on patients from 1994 and 1995 calendar years. Data fields correspond to the fields in the "PATIENT REGISTRY FORM".

Data fields in paccn945.txt:

CHART NUMBER	- patient's unique ID
PACCN ID #	- patient's ID in PACCN system
PROCEDURE REQUIRED	
ACCEPTANCE DATE	
PATIENT LOCATION	
ADMISSION DATE	
TRANSFER DATE	
RECENT MI DATE	
CCVS CLASS	- angina class (by Canadian CardioVascular Society)
NYHA	- heart failure class (by New York Heart Association)
MORBID/MORTAL	- expected morbidity and mortality (low or high without indication of risk factors)
LV FUNCTION	- left ventricular function (0,1,2,3,4)
URGENCY RATING SCORE	- calculated value for CABG patients, and the estimated triage score for non-CABG
MAX RECOMMMENDED WAIT	- maximum recommended waiting time derived from urgency rating score
DATE OFF LIST	
REASON OFF LIST	

Appendix 2.

Description of the Fields in Merged Data Files.

1. **merg945.dbf** - ADT, PACCN and Master files (both 1994 and 1995) merged by patients unique ID, date of admission (PACCN and Master file) and dates of admission and surgery (ADT and Master file)

Field	Description	Source
CLIENT_ID	patient's unique ID	all files
ADDATE	admission date	all files
AD_WEEK_DA	admission day of the week	calc
AD_MON	admission month	calc
COTIME	time in the operating room	calc from data in ADT files
ICU_STAY	total time in ICU(included both CRR and CS)	calc from data in ADT files
LOS_ICU_GR	group of length of stay in ICU (1-[0,1), 2-[1-2), 3-[2,3), 4-[3,4), 4-grater than 4)	calc from data in ADT files
ICU_STAY_G	another graoupping of length of stay in ICU based on 10-th and 90-th percentiles of this sample: group 1- less than 1 day, group 2 - [1,4], group 3 - grater than 4 days	calc from data in ADT files
POST_ICU_S	stay in the hospital after transfer from ICU till discharge	calc from data in ADT files
POST_OP_ST	total postoperative stay in the hospital	calc from data in ADT files
DISFROM	nursing unit patient was discharged from	ADT
DIS_CD	discharge code (1- care completed, 2 - CTB<48 hours after admission, 3-CTB>48 hours after admission, 4-self discharged, 5- coronary care elsewhere	ADT
AGE		calc from data in Mast. file
EMERG	emergency as in Master file	Master file
GENDER	patient's gender ("M","F")	Master file
PROCEDURE	procedure as in Master file	Master file
PROCED_GRO	groups of procedures in Master file	calc from data in Mast. file
REDO	indication of repeat operation	Master file
WEIGHT	patient's weight	Master file
BSA	patient's body surface area (calculated from weight and height)	Master file
LVCLASSB	left ventricular class (1,2,3,4)	Master file
IABP	use of intra-aortic balloon pump	Master file

PUMPTIME	time on cardiopulmonary bypass (CPB)	Master file
ANOXIA1	time when heart is excluded from circulation	Master file
ANOXIA2	the same as above	Master file
RECMI		PACCN
CCS_CLASS	angina class (by Canadian Cardiovascular Society)	PACCN
NYHA_CLASS	heart failure class (by New York Heart Association)	PACCN
LV_FUN	left ventricular function (0,1,2,3,4)	PACCN
UR_SCORE	calculated value for CABG patients, and the estimated triage score for non-CABG	PACCN
MAX_REC_WA	maximum recommended waiting time derived from urgency rating score	PACCN
WAITED	days pratient wated after admission till operation performed	calc from data in PACCN
URG_CATEGO	urgency category based on PACCN urgency score	calc from data in PACCN
SUR_DATE	date of surgery	Master file and ADT
SUR_WEEK_D	surgery day of the week	calc
SUR_MON	surgery month	calc
EXTTIME	time of extubation	Master file
EXTDATE	date of extubation	Master file
PROCEDURE_	procedure required at the time of registering in PACCN	PACCN
ACCEPDATE	date of acceptance	PACCN
ACC_WEEK_D	acceptance day of the week	calc from data in PACCN
ACC_MO	acceptance month	calc from data in PACCN
TIME_TO_AD	time between acceptance and admission	calc from data in PACCN

2. **all945.dbf** - all Anaesthesia and Surgery parts of TMR (both 1994 and 1995 years) and merg945.dbf merged together by patients ID and surgery and anaesthesia dates

Field	Description	Source
ID	patient's unique ID	all files
DOA	date of anaesthesia	a9?
A521	asthma mild	a9?
A522	asthma moderate	a9?
A523	asthma severe	a9?
ASTHMA	combined field of A521-A523	calc from a9?
A531	C.O.P.D. (Chronic Obstructive Pulmonary Disease) mild	a9?
A532	C.O.P.D. moderate	a9?
A533	C.O.P.D. severe	a9?
COPD	combined field of A531-A533	
A541	pulmonary hypertension mild	a9?
A542	pulmonary hypertension moderate	a9?
A543	pulmonary hypertension severe	a9?
PULMHYPERT	combined field of A541-A543	
FORMSMOK	former scmoker	a9?
A61	normal	a9?
A621	diabetes - diet controlled	a9?
A622	diabetes - controlled with oral meds	a9?
A623	diabetes - controlled with insulin	a9?
DIABETES	combined field of A561-A523	
A731	prior CVA (Cerebral Vascular Accident) - embolic	a9?
A732	prior CVA - hemorrage	a9?
A733	prior CVA - carotid vascular disease	a9?
PRIOR_CVA	combined field of A531-A533	
OBESITY	obesity	a9?/field A82
RENAL_DIS	renal disease	a9?/field A84
BLEED_TEND	bleeding tendency	a9?/field A85
HEPATIC_DI	hepatic disease	a9?/field A86
PVD	peripheral vacular disease (PVD)	a9?/field A87
OTHER_MED	other medical problem	a9?/field A88
B1141	calcium	b9?
B1142	dopamine	b9?
B1143	epinephrine	b9?
B1144	amrinone	b9?
B1145	other	b9?
B1146	calcium	b9?
B1222	dopamine	b9?
B1223	dobutamine	b9?

B1224	epinephrine	b9?
B1225	amrinone	b9?
B1226	other	b9?
VALV_DIS	valvular disease	s9?/field s22
AORT_ANEUR	aortic aneurism	s9?/field s23
AORT_DISSE	aortic dissection	s9?/field s24
CONGEN_DEF	congenital defect	s9?/field s25
CARDIOMYOP	cardiomyopathy	s9?/field s26
ARRHYTHMIA	arrhythmia	s9?/field s27
S32	peripheral vascular disease (PVD)	s9?
S33	cerebral vascular accident (CVA)	s9?
S35	hypertension	s9?
S71	elective	s9?
S72	urgent	s9?
S73	emergency (ASAP)	s9?
EMERG	combined field of S71-S73	
S81	asymptomatic	s9?
S82	angina stable	s9?
S83	angina unstable	s9?
ANGINA	combined field of S71-S73	
LEFT_MAIN	left main stenosis	s9?/field S84
FAIL_PT	field PTCA	s9?/field S85
ACUTE_MI	acute MI	s9?/field S86
CHF	CHF (congestive heart failure)	s9?/field S87
VSD	VSD (Ventricular Septal Defect)	s9?/field S88
VENT_RUPT	ventricular rupture	s9?/field S89
LV_ANEUR	LV aneurysm	s9?/field S810
ARRHYTHMIA	arrhythmia	s9?/field S811
REPEAT_OPE	previous graft stenosis/occlusion (repeat operation)	s9?/field S812
CARDIOG_SH	cardiogenic shock	s9?/field S813
OTHER_COR	other	s9?/field S814
CABG	CABG	u9?/ field U41
REDO_CABG	redo CABG	u9?/ field U42
ENDARTEREC	endarterectomy	u9?/ field U43
CLOS_VSD	closure of VSD	u9?/ field U44
REP_VENT_R	repair of ventricular rupture	u9?/ field U45
PLIC_LVANE	plication of LV aneurysm (plication = taking a "tuck" in the wall of the LV)	u9?/ field U46
REC_LVANEU	resection of LV aneurysm	u9?/ field U47
U612	quality of graft 1 (good/fair/poor/end)	u9?
U622	quality of graft 2 (good/fair/poor/end)	u9?
U632	quality of graft 3 (good/fair/poor/end)	u9?
U642	quality of graft 4 (good/fair/poor/end)	u9?

U652	quality of graft 5 (good/fair/poor/end)	u9?
U662	quality of graft 6 (good/fair/poor/end)	u9?
GRAFT_QUAL	"E" if any of u612-U662 is "E"	calc
GRAFT_COUN	count of grafts	empty
TRANSPL	cardiac transplantation	u9?/ field U101
IABP	insertion of intra-aortic balloon heart transplant	u9?/ field U104
	All the rest of the fields match fields in merg945.dbf file	
ADDATE		
COTIME		
ICU_STAY		
ICU_STAY_G		
POST_ICU_S		
POST_OP_ST		
DISFROM		
DIS_CD		
BIRTHDAY		
AGE		
EMERG_1		
SUR_DATE		
GENDER		
PROCEDURE		
PROCED_GRO		
REDO		
WEIGHT		
BSA		
IABP_1		
LVCLASSB		
PUMPTIME		
ANOXIA1		
ANOXIA2		
EXTTIME		
EXTDATE		
PROCEDURE_		
ACCEPDATE		
TIME_TO_AD		
RECMIDATE		
CCS_CLASS		
NYHA_CLASS		
LV_FUN		
UR_SCORE		
MAX_REC_WA		
WAITED		

Appendix 3.

A. Distributions of time patients spent in the OR by urgency categories.

	EMERGECY		URGENT		ELECTIVE	
Midpoint (days)	COUNT	CUMUL. %	COUNT	CUMUL. %	COUNT	CUMUL. %
0	0	0.00%	2	0.26%	1	0.07%
0.02	1	0.88%	1	0.39%	0	0.07%
0.04	1	1.75%	0	0.39%	1	0.15%
0.06	1	2.63%	1	0.52%	1	0.22%
0.08	0	2.63%	1	0.65%	1	0.30%
0.1	3	5.26%	2	0.91%	4	0.59%
0.12	7	11.40%	12	2.46%	19	2.01%
0.14	18	27.19%	25	5.69%	36	4.68%
0.16	13	38.60%	44	11.38%	68	9.74%
0.18	13	50.00%	87	22.64%	150	20.89%
0.2	15	63.16%	155	42.69%	257	40.00%
0.22	11	72.81%	151	62.23%	280	60.82%
0.24	6	78.07%	81	72.70%	166	73.16%
0.26	6	83.33%	76	82.54%	154	84.61%
0.28	5	87.72%	54	89.52%	87	91.08%
0.3	4	91.23%	35	94.05%	41	94.13%
0.32	4	94.74%	18	96.38%	31	96.43%
0.34	1	95.61%	9	97.54%	12	97.32%
0.36	0	95.61%	5	98.19%	13	98.29%
0.38	1	96.49%	6	98.97%	3	98.51%
0.4	0	96.49%	4	99.48%	2	98.66%
0.42	0	96.49%	1	99.61%	7	99.18%
0.44	1	97.37%	0	99.61%	6	99.63%
0.46	3	100.00%	2	99.87%	3	99.85%
0.48			0	99.87%	1	99.93%
0.5			1	100.00%	0	99.93%
0.52					1	100.00%

B. Distributions of time patients spent in the ICU by urgency categories.

	EMERGENT		URGENT		ELECTIVE	
Midpoint (days)	Count	CUMUL %	COUNT	CUMUL %	COUNT	CUMUL %
0	1	0.88%	6	0.78%	3	0.22%
1	18	16.81%	267	35.27%	548	41.00%
2	30	43.36%	293	73.13%	534	80.73%

3	17	58.41%	77	83.07%	113	89.14%
4	6	63.72%	59	90.70%	78	94.94%
5	10	72.57%	24	93.80%	25	96.80%
6	8	79.65%	17	95.99%	12	97.69%
7	6	84.96%	8	97.03%	5	98.07%
8	2	86.73%	2	97.29%	9	98.74%
9	2	88.50%	4	97.80%	1	98.81%
10	2	90.27%	1	97.93%	2	98.96%
11	2	92.04%	0	97.93%	0	98.96%
12	1	92.92%	0	97.93%	1	99.03%
13	1	93.81%	2	98.19%	2	99.18%
14	0	93.81%	1	98.32%	2	99.33%
15	0	93.81%	1	98.45%	0	99.33%
16	0	93.81%	0	98.45%	2	99.48%
17	2	95.58%	1	98.58%	0	99.48%
18	0	95.58%	0	98.58%	0	99.48%
19	0	95.58%	0	98.58%	1	99.55%
20	0	95.58%	1	98.71%	0	99.55%
21	0	95.58%	3	99.10%	0	99.55%
22	2	97.35%	1	99.22%	1	99.63%
23	0	97.35%	0	99.22%	0	99.63%
24	1	98.23%	1	99.35%	1	99.70%
25	0	98.23%	0	99.35%	0	99.70%
26	0	98.23%	1	99.48%	1	99.78%
27	0	98.23%	1	99.61%	0	99.78%
28	0	98.23%	0	99.61%	2	99.93%
29	0	98.23%	0	99.61%	0	99.93%
30	1	99.12%	1	99.74%	1	100.00%
31	0	99.12%	0	99.74%		
32	0	99.12%	1	99.87%		
33	0	99.12%	0	99.87%		
34	0	99.12%	0	99.87%		
35	0	99.12%	0	99.87%		
36	0	99.12%	0	99.87%		
37	0	99.12%	0	99.87%		
38	0	99.12%	0	99.87%		
39	1	100.00%	0	99.87%		
40	0	100.00%	1	100.00%		

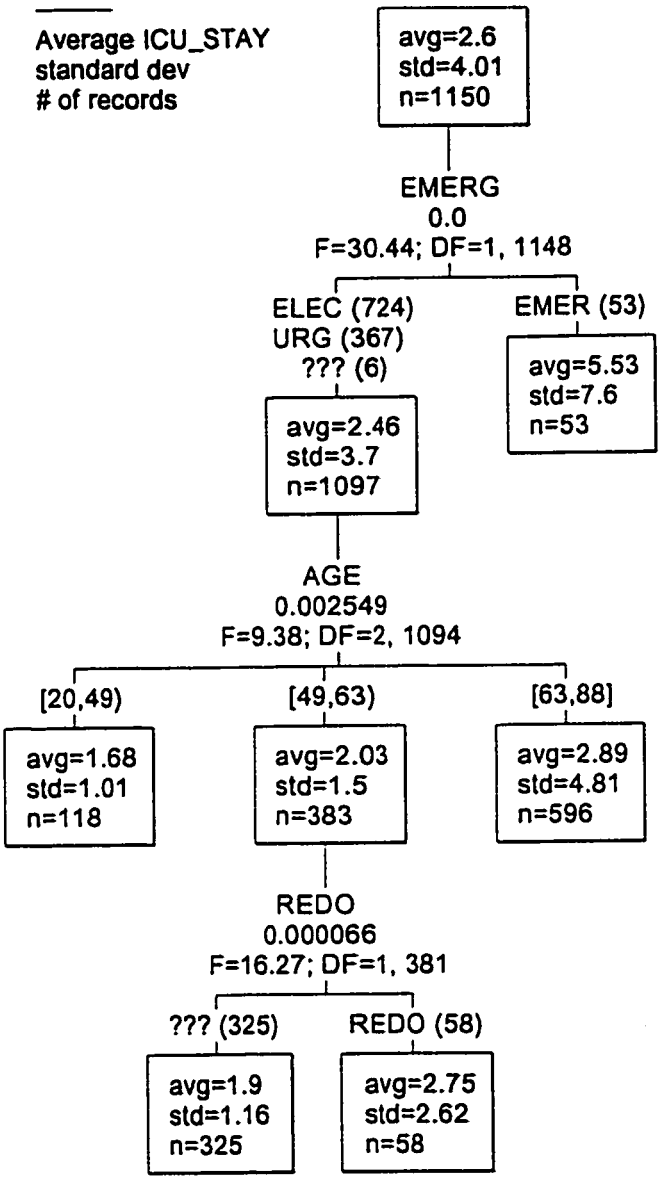
C. Distributions of time patients spent in the POST- ICU by urgency categories.

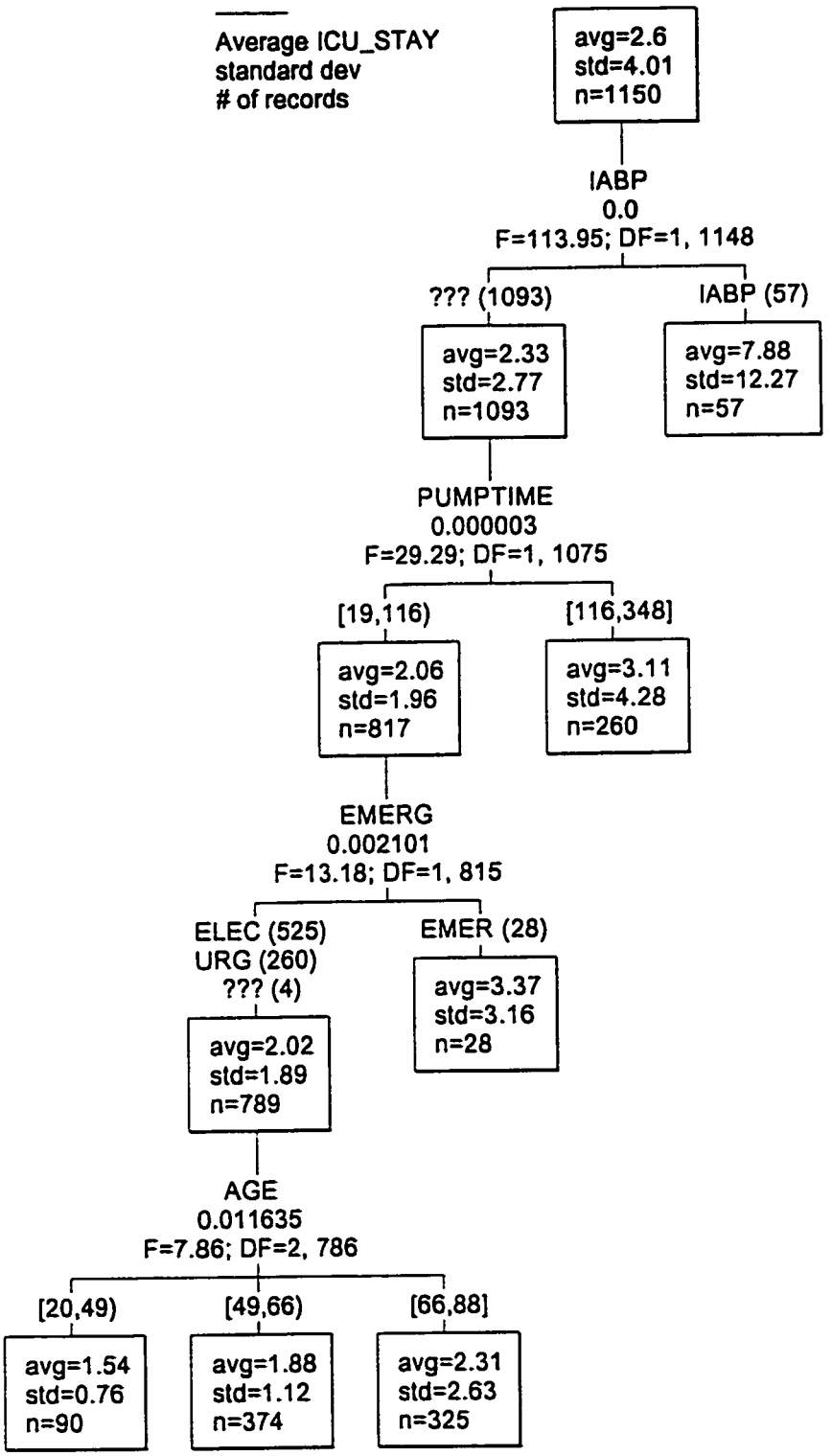
	EMERGENT		URGENT		ELECTIVE	
Midpoint (days)	COUNT	CUMUL %	COUNT	CUMUL %	COUNT	CUMUL %
0	18	16.51%	28	3.64%	29	2.16%
1	0	16.51%	2	3.90%	9	2.83%
2	0	16.51%	3	4.29%	9	3.50%

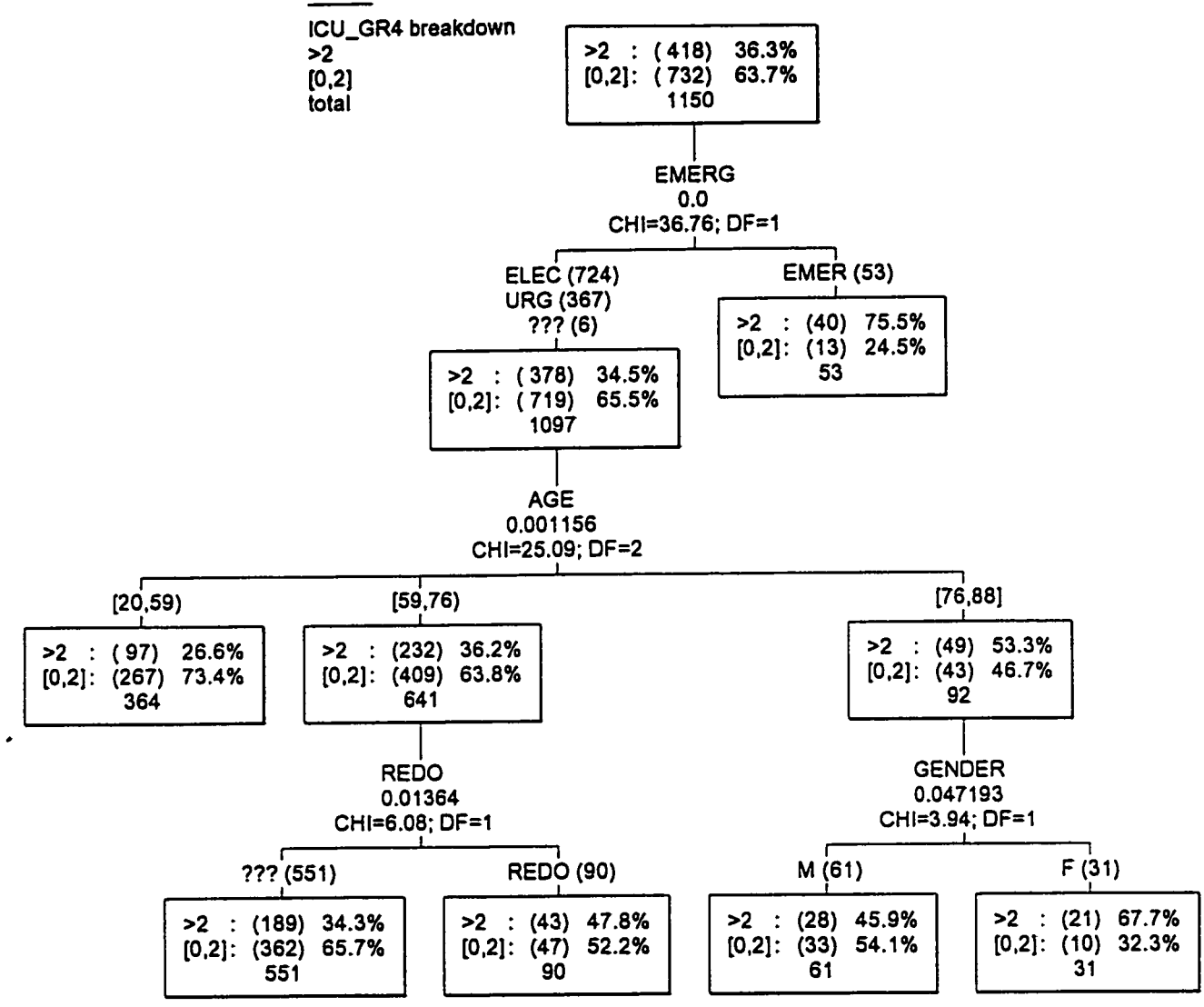
3	5	21.10%	24	7.41%	35	6.11%
4	12	32.11%	76	17.30%	146	17.00%
5	11	42.20%	143	35.89%	269	37.06%
6	16	56.88%	165	57.35%	310	60.18%
7	11	66.97%	100	70.35%	202	75.24%
8	5	71.56%	58	77.89%	105	83.07%
9	9	79.82%	36	82.57%	43	86.28%
10	5	84.40%	20	85.18%	45	89.63%
11	2	86.24%	18	87.52%	27	91.65%
12	0	86.24%	14	89.34%	28	93.74%
13	2	88.07%	14	91.16%	16	94.93%
14	1	88.99%	16	93.24%	6	95.38%
15	2	90.83%	6	94.02%	10	96.12%
16	3	93.58%	2	94.28%	10	96.87%
17	1	94.50%	8	95.32%	3	97.09%
18	0	94.50%	6	96.10%	5	97.46%
19	1	95.41%	6	96.88%	4	97.76%
20	0	95.41%	1	97.01%	2	97.91%
21	1	96.33%	5	97.66%	2	98.06%
22	1	97.25%	0	97.66%	2	98.21%
23	1	98.17%	1	97.79%	5	98.58%
24	0	98.17%	2	98.05%	1	98.66%
25	0	98.17%	2	98.31%	2	98.81%
26	0	98.17%	0	98.31%	2	98.96%
27	0	98.17%	3	98.70%	2	99.11%
28	0	98.17%	2	98.96%	1	99.18%
29	0	98.17%	1	99.09%	1	99.25%
30	2	100.00%	2	99.35%	2	99.40%
31			1	99.48%	1	99.48%
32			1	99.61%	0	99.48%
33			2	99.87%	0	99.48%
34			0	99.87%	2	99.63%
35			1	100.00%	1	99.70%
36					1	99.78%
37					1	99.85%
38					0	99.85%
39					0	99.85%
40					2	100.00%

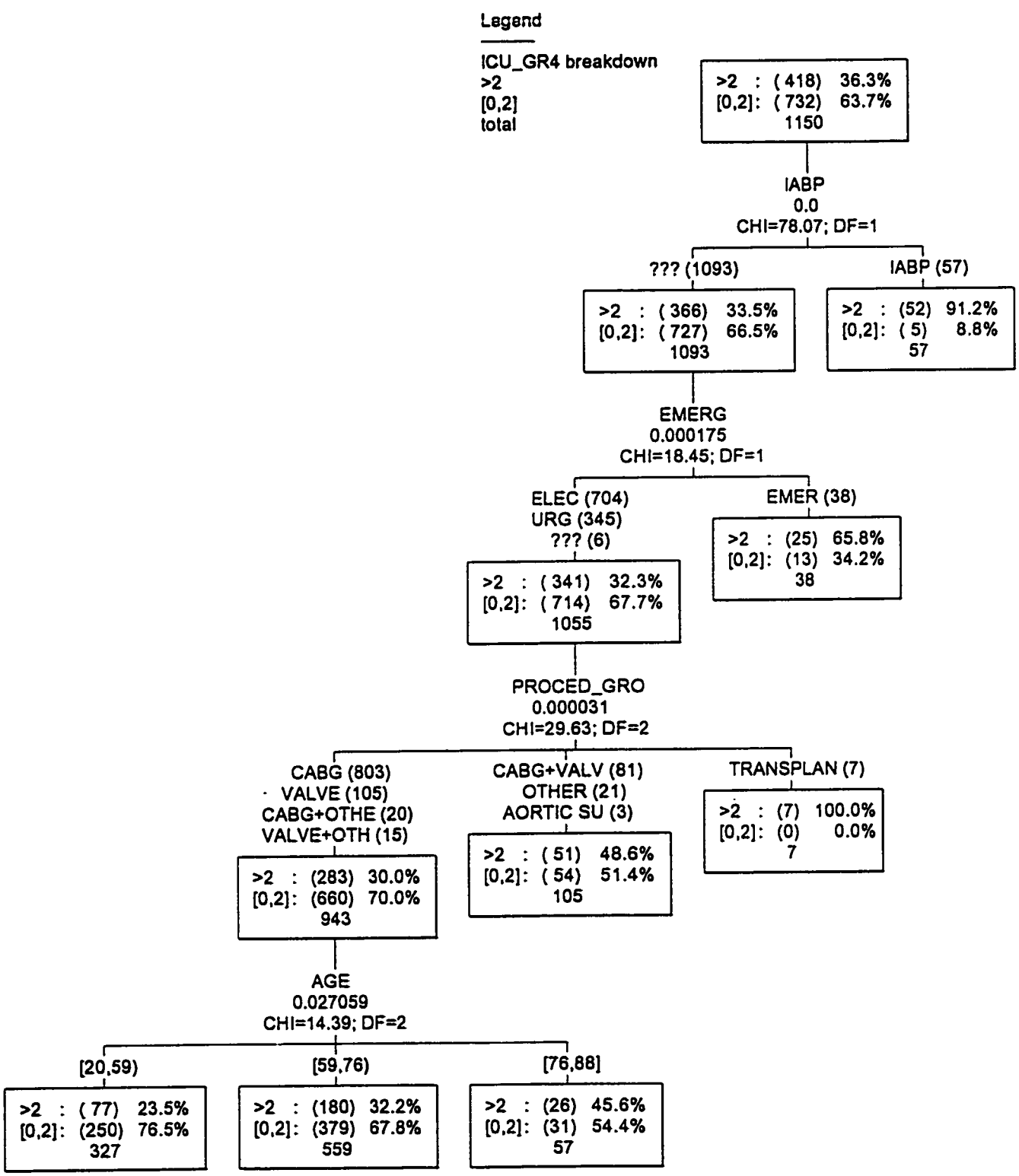
Appendix 4.

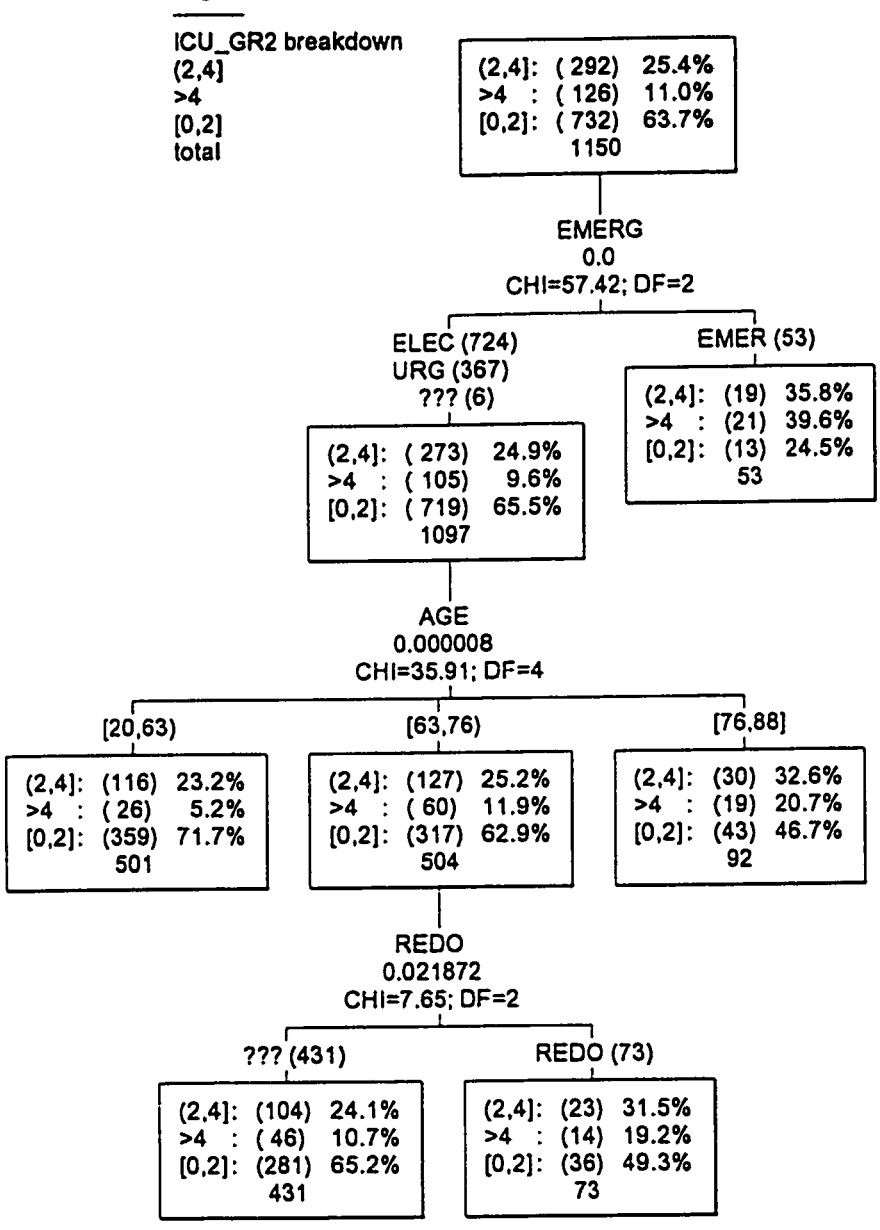
A. Results of the analysis of ICU length of stay by KnowledgeSEEKER .

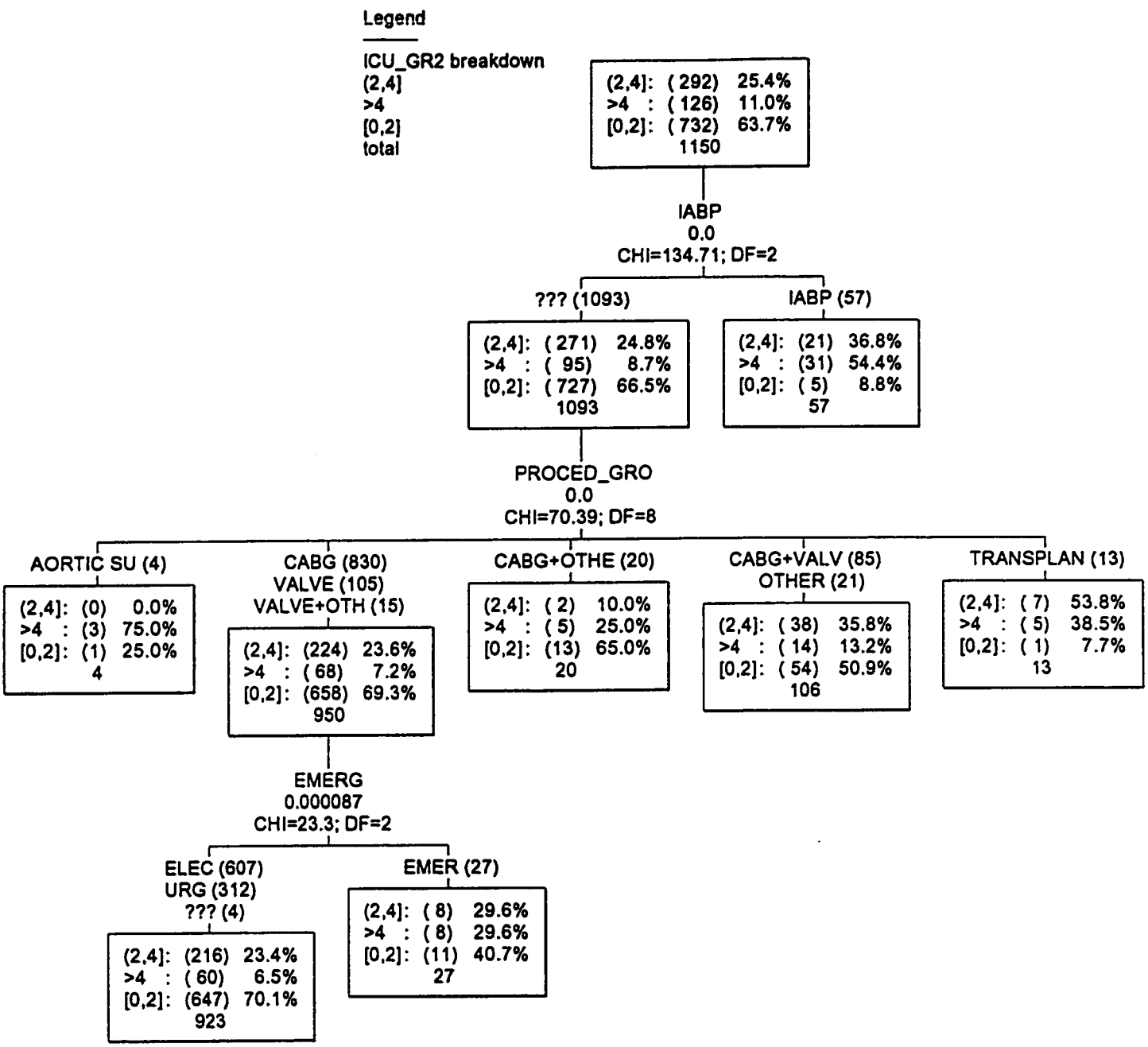

Legend
Average ICU_STAY
standard dev
# of records
avg=2.6
std=4.01
n=1150
EMERG
0.0
F=30.44; DF=1, 1148
ELEC (724)
URG (367)
??? (6)
EMER (53)
avg=5.53
std=7.6
n=53
avg=2.46
std=3.7
n=1097
AGE
0.002549
F=9.38; DF=2, 1094
[20,49)
[49,63)
[63,88]
avg=1.68
std=1.01
n=118
avg=2.03
std=1.5
n=383
avg=2.89
std=4.81
n=596
REDO
0.000066
F=16.27; DF=1, 381
??? (325)
REDO (58)
avg=1.9
std=1.16
n=325
avg=2.75
std=2.62
n=58

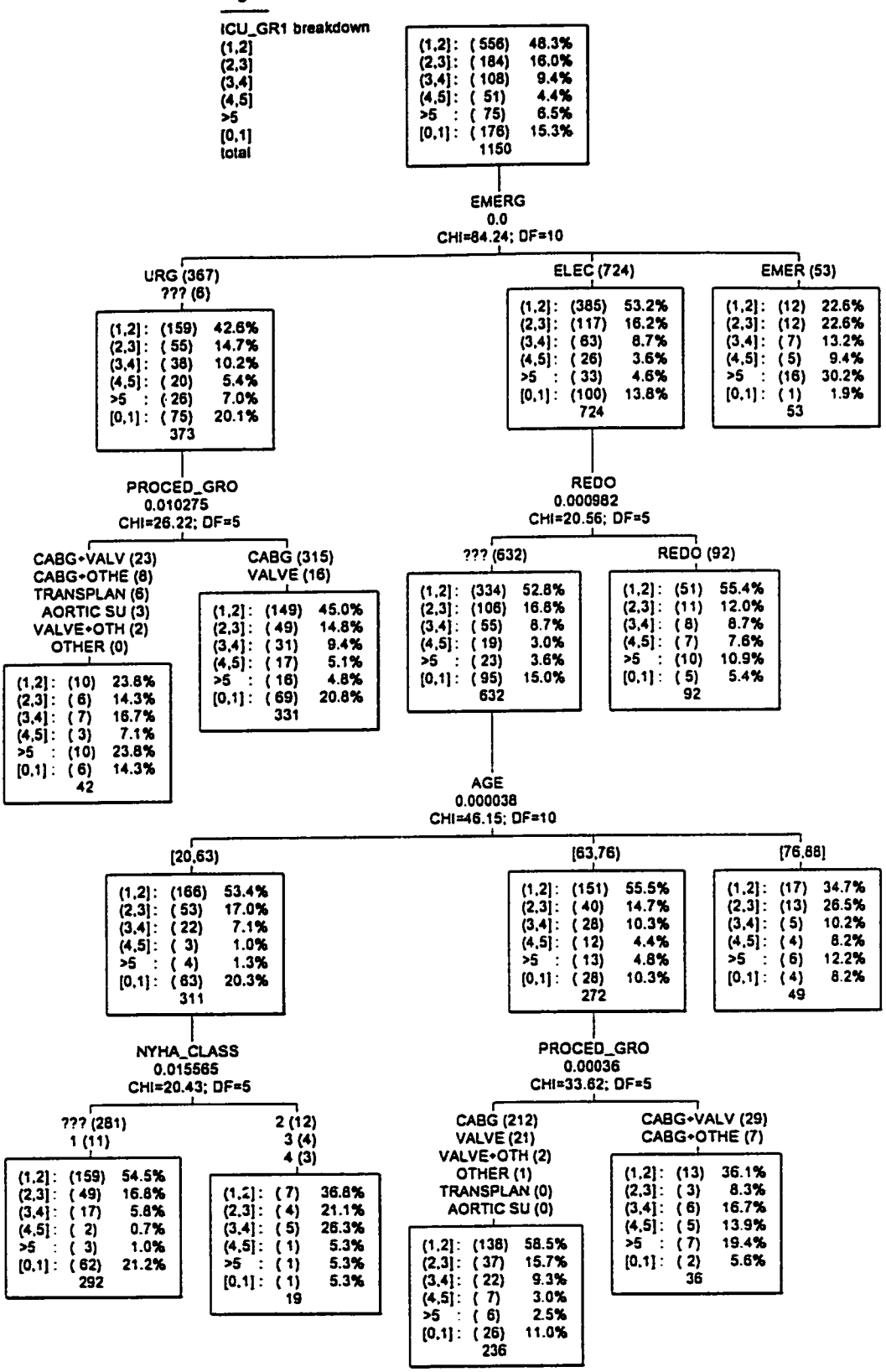

Legend
Average ICU_STAY
standard dev
# of records
avg=2.6
std=4.01
n=1150
IABP
0.0
F=113.95; DF=1, 1148
??? (1093)
IABP (57)
avg=2.33
std=2.77
n=1093
avg=7.88
std=12.27
n=57
PUMPTIME
0.000003
F=29.29; DF=1, 1075
[19,116)
[116,348]
avg=2.06
std=1.96
n=817
avg=3.11
std=4.28
n=260
EMERG
0.002101
F=13.18; DF=1, 815
ELEC (525)
URG (260)
??? (4)
EMER (28)
avg=2.02
std=1.89
n=789
avg=3.37
std=3.16
n=28
AGE
0.011635
F=7.86; DF=2, 786
[20,49)
[49,66)
[66,88]
avg=1.54
std=0.76
n=90
avg=1.88
std=1.12
n=374
avg=2.31
std=2.63
n=325

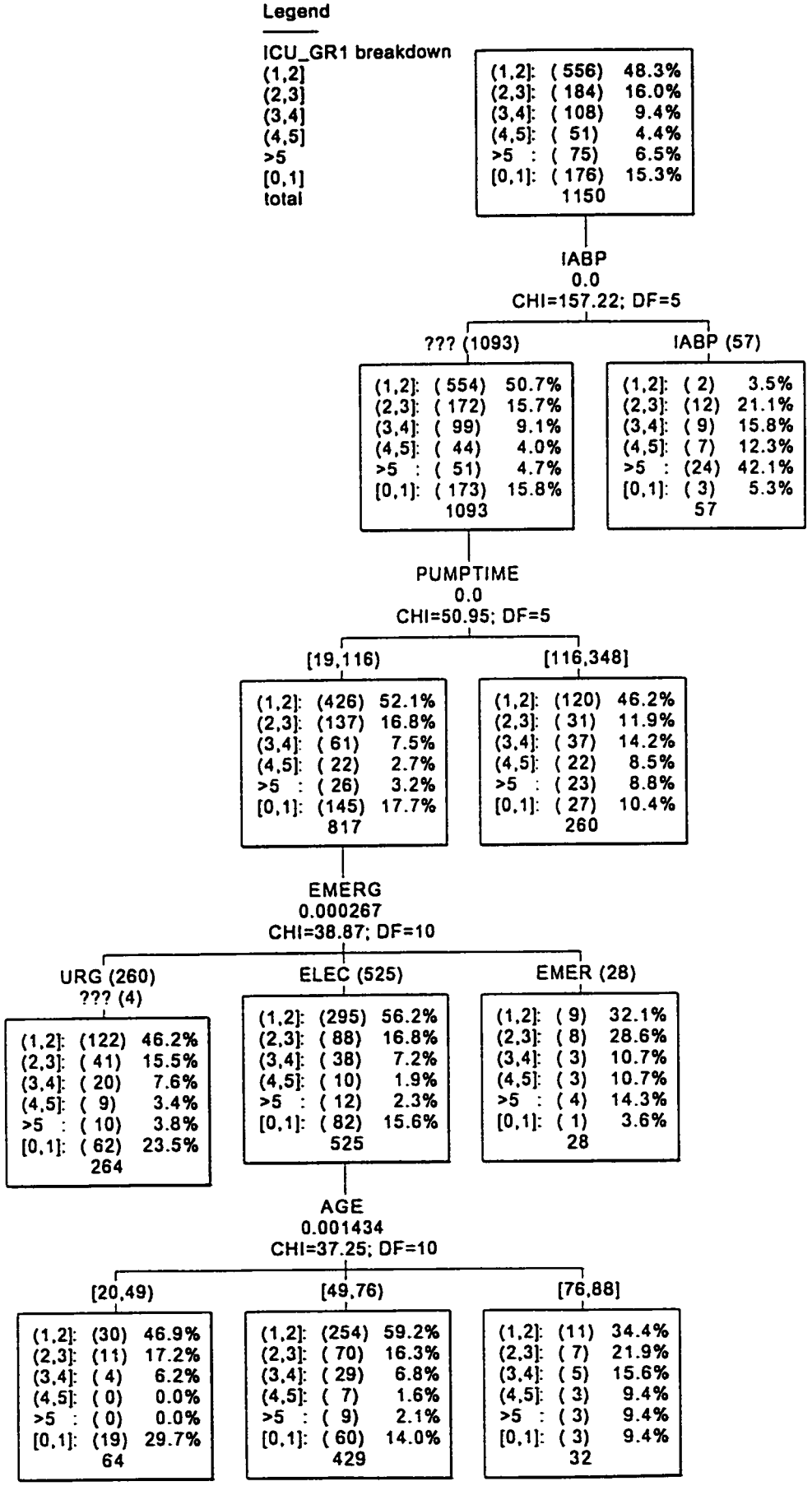

ICU_GR4 breakdown
>2
[0,2]
total
>2 : (418) 36.3%
[0,2]: (732) 63.7%
1150
EMERG
0.0
CHI=36.76; DF=1
ELEC (724)
URG (367)
??? (6)
EMER (53)
>2 : (378) 34.5%
[0,2]: (719) 65.5%
1097
>2 : (40) 75.5%
[0,2]: (13) 24.5%
53
AGE
0.001156
CHI=25.09; DF=2
[20,59)
[59,76)
[76,88]
>2 : (97) 26.6%
[0,2]: (267) 73.4%
364
>2 : (232) 36.2%
[0,2]: (409) 63.8%
641
>2 : (49) 53.3%
[0,2]: (43) 46.7%
92
REDO
0.01364
CHI=6.08; DF=1
GENDER
0.047193
CHI=3.94; DF=1
??? (551)
REDO (90)
M (61)
F (31)
>2 : (189) 34.3%
[0,2]: (362) 65.7%
551
>2 : (43) 47.8%
[0,2]: (47) 52.2%
90
>2 : (28) 45.9%
[0,2]: (33) 54.1%
61
>2 : (21) 67.7%
[0,2]: (10) 32.3%
31

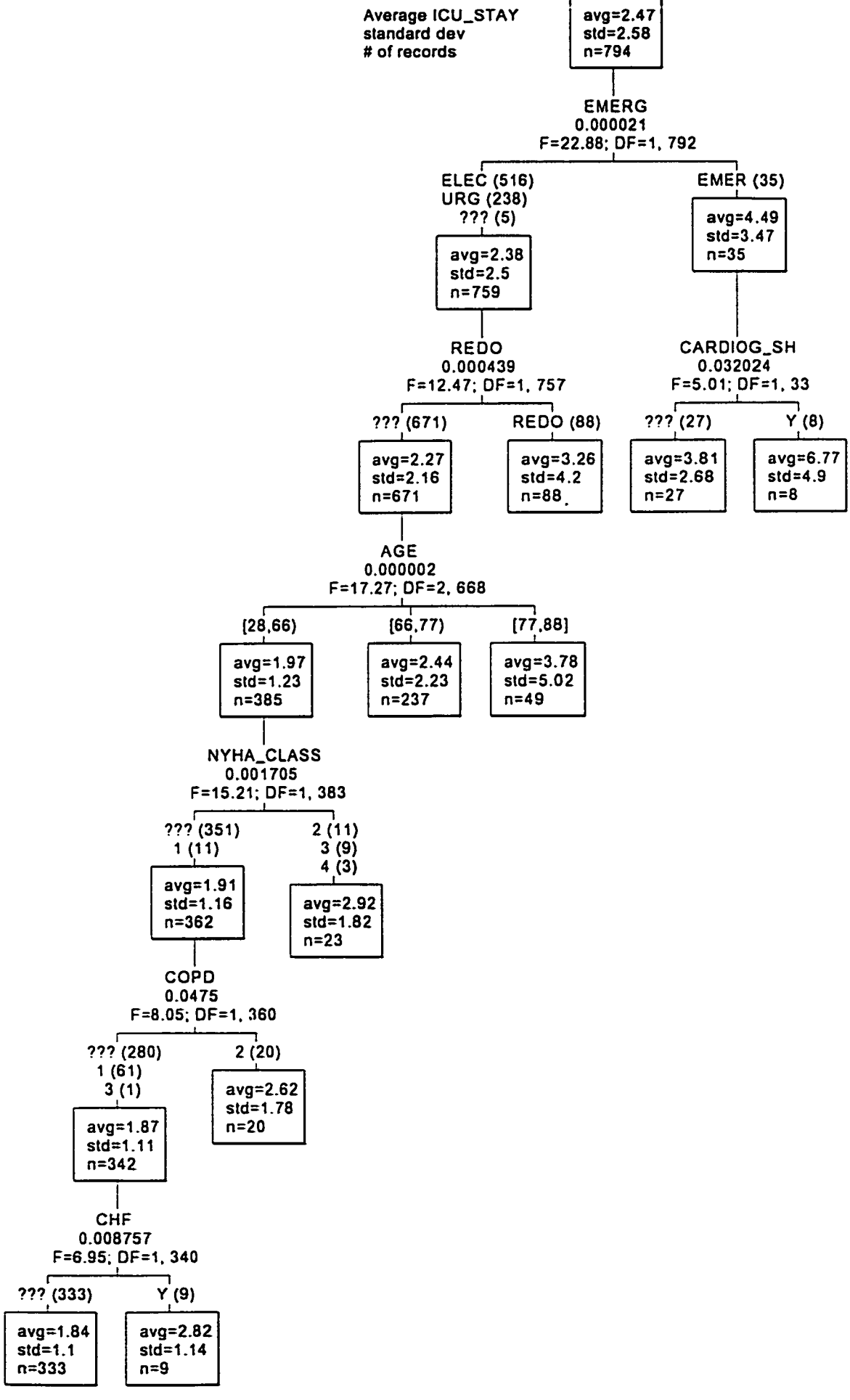

Legend
ICU_GR4 breakdown
>2
[0,2]
total
>2 : (418) 36.3%
[0,2]: (732) 63.7%
1150
IABP
0.0
CHI=78.07; DF=1
??? (1093)
IABP (57)
>2 : (366) 33.5%
[0,2]: (727) 66.5%
1093
>2 : (52) 91.2%
[0,2]: (5) 8.8%
57
EMERG
0.000175
CHI=18.45; DF=1
ELEC (704)
URG (345)
??? (6)
EMER (38)
>2 : (341) 32.3%
[0,2]: (714) 67.7%
1055
>2 : (25) 65.8%
[0,2]: (13) 34.2%
38
PROCED_GRO
0.000031
CHI=29.63; DF=2
CABG (803)
VALVE (105)
CABG+OTHE (20)
VALVE+OTH (15)
CABG+VALV (81)
OTHER (21)
AORTIC SU (3)
TRANSPLAN (7)
>2 : (283) 30.0%
[0,2]: (660) 70.0%
943
>2 : (51) 48.6%
[0,2]: (54) 51.4%
105
>2 : (7) 100.0%
[0,2]: (0) 0.0%
7
AGE
0.027059
CHI=14.39; DF=2
[20,59)
[59,76)
[76,88]
>2 : (77) 23.5%
[0,2]: (250) 76.5%
327
>2 : (180) 32.2%
[0,2]: (379) 67.8%
559
>2 : (26) 45.6%
[0,2]: (31) 54.4%
57

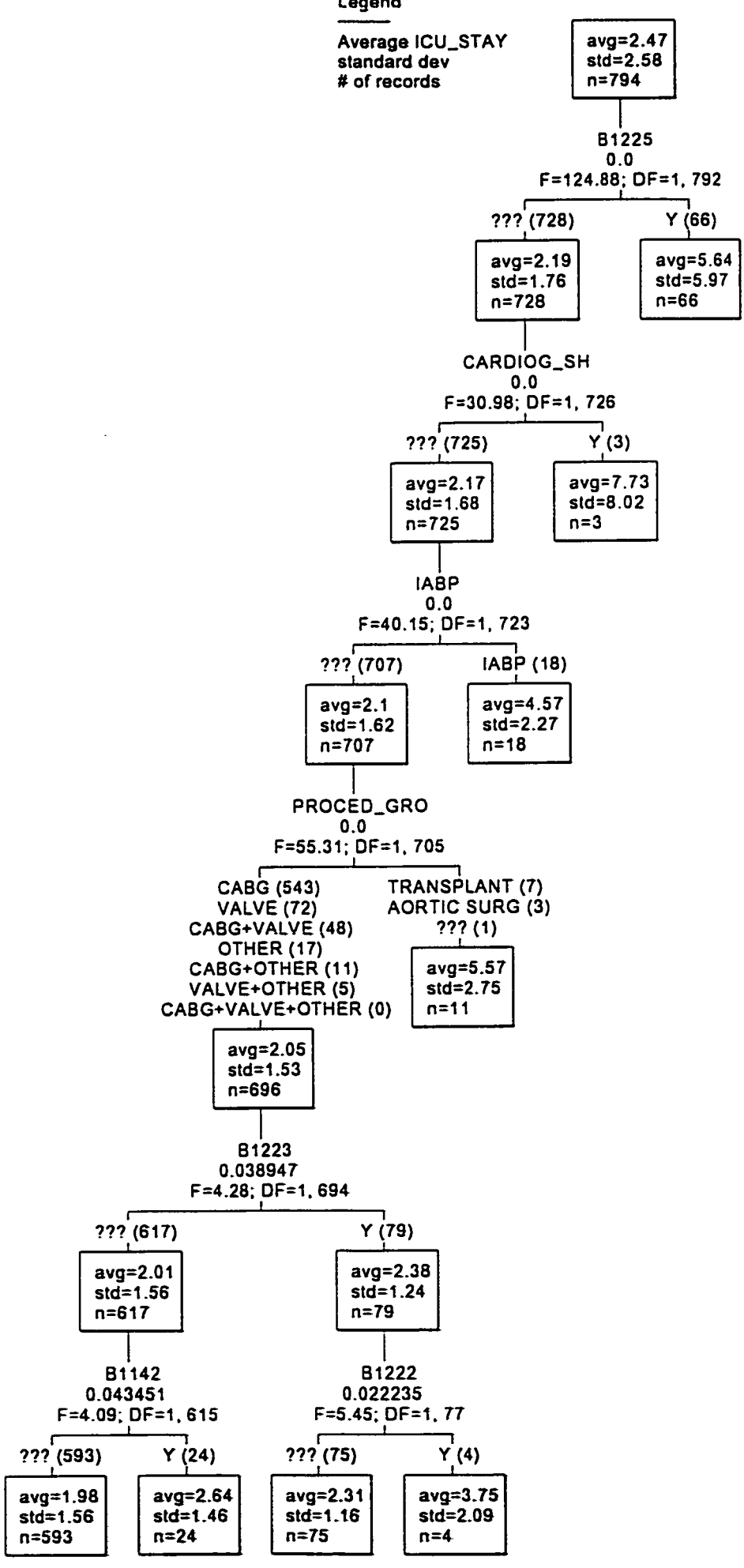

ICU_GR2 breakdown
(2,4]
>4
[0,2]
total
(2,4]: (292) 25.4%
>4 : (126) 11.0%
[0,2]: (732) 63.7%
1150
EMERG
0.0
CHI=57.42; DF=2
ELEC (724)
URG (367)
??? (6)
EMER (53)
(2,4]: (273) 24.9%
>4 : (105) 9.6%
[0,2]: (719) 65.5%
1097
(2,4]: (19) 35.8%
>4 : (21) 39.6%
[0,2]: (13) 24.5%
53
AGE
0.000008
CHI=35.91; DF=4
[20,63)
[63,76)
[76,88]
(2,4]: (116) 23.2%
>4 : (26) 5.2%
[0,2]: (359) 71.7%
501
(2,4]: (127) 25.2%
>4 : (60) 11.9%
[0,2]: (317) 62.9%
504
(2,4]: (30) 32.6%
>4 : (19) 20.7%
[0,2]: (43) 46.7%
92
REDO
0.021872
CHI=7.65; DF=2
??? (431)
REDO (73)
(2,4]: (104) 24.1%
>4 : (46) 10.7%
[0,2]: (281) 65.2%
431
(2,4]: (23) 31.5%
>4 : (14) 19.2%
[0,2]: (36) 49.3%
73

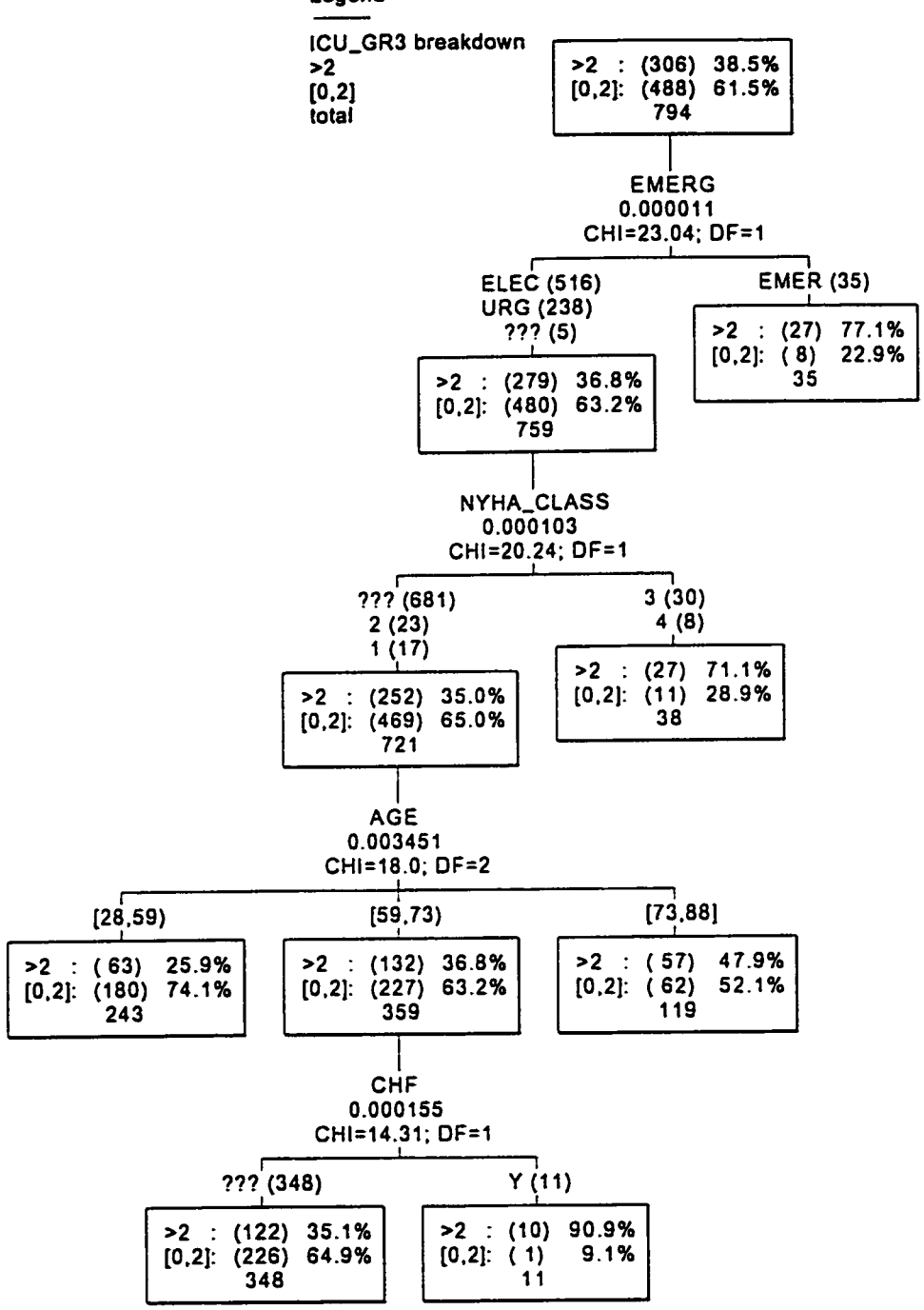

Legend
ICU_GR2 breakdown
(2,4]
>4
[0,2]
total
(2,4]: (292) 25.4%
>4 : (126) 11.0%
[0,2]: (732) 63.7%
1150
IABP
0.0
CHI=134.71; DF=2
??? (1093)
IABP (57)
(2,4]: (271) 24.8%
>4 : (95) 8.7%
[0,2]: (727) 66.5%
1093
(2,4]: (21) 36.8%
>4 : (31) 54.4%
[0,2]: (5) 8.8%
57
PROCED_GRO
0.0
CHI=70.39; DF=8
AORTIC SU (4)
CABG (830)
VALVE (105)
VALVE+OTH (15)
CABG+OTHE (20)
CABG+VALV (85)
OTHER (21)
TRANSPLAN (13)
(2,4]: (0) 0.0%
>4 : (3) 75.0%
[0,2]: (1) 25.0%
4
(2,4]: (224) 23.6%
>4 : (68) 7.2%
[0,2]: (658) 69.3%
950
(2,4]: (2) 10.0%
>4 : (5) 25.0%
[0,2]: (13) 65.0%
20
(2,4]: (38) 35.8%
>4 : (14) 13.2%
[0,2]: (54) 50.9%
106
(2,4]: (7) 53.8%
>4 : (5) 38.5%
[0,2]: (1) 7.7%
13
EMERG
0.000087
CHI=23.3; DF=2
ELEC (607)
URG (312)
??? (4)
EMER (27)
(2,4]: (216) 23.4%
>4 : (60) 6.5%
[0,2]: (647) 70.1%
923
(2,4]: (8) 29.6%
>4 : (8) 29.6%
[0,2]: (11) 40.7%
27

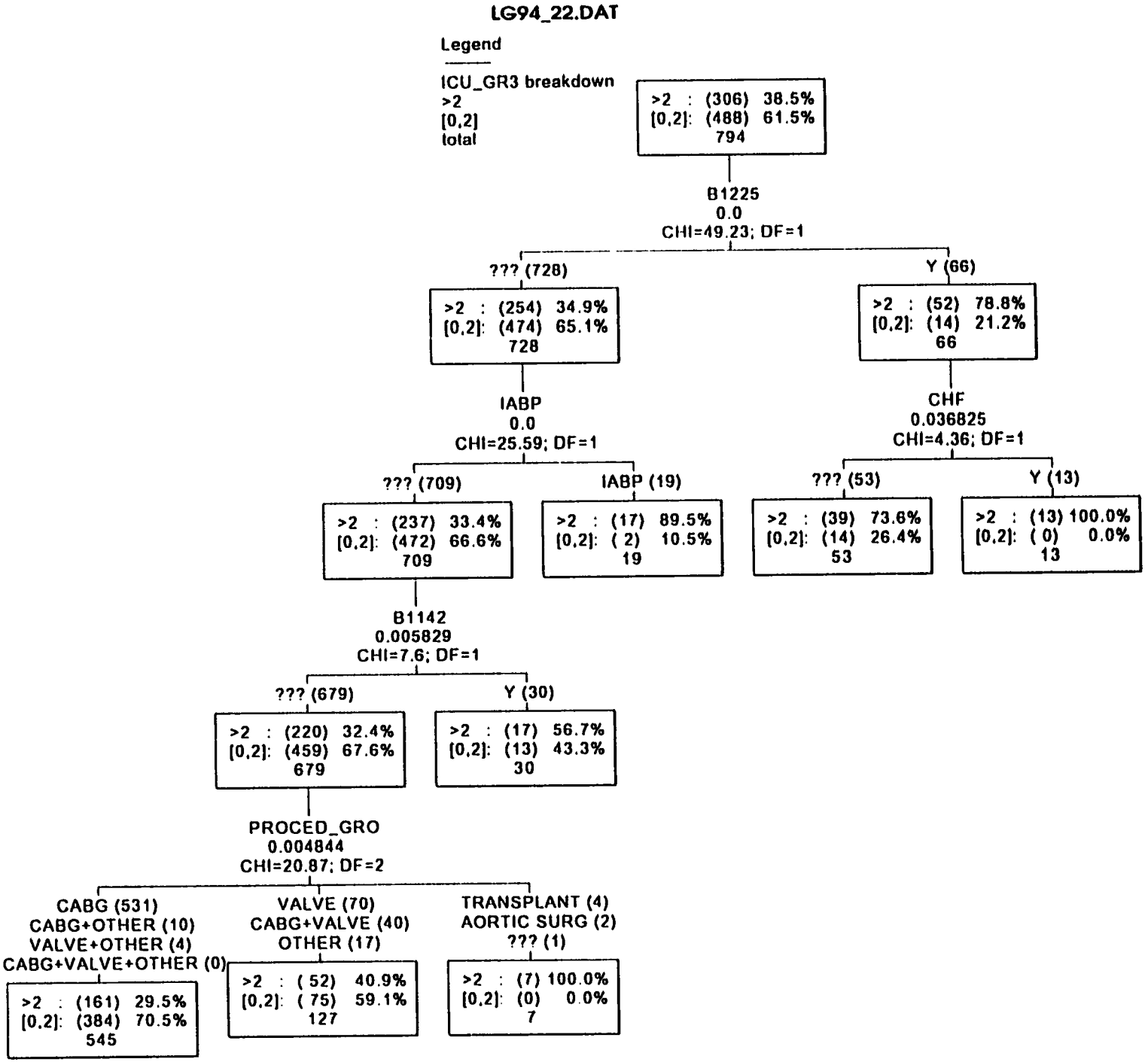

ICU_GR1 breakdown
(1,2]
(2,3]
(3,4]
(4,5]
>5
[0,1]
total
(1,2]: (556) 48.3%
(2,3]: (184) 16.0%
(3,4]: (108) 9.4%
(4,5]: (51) 4.4%
>5 : (75) 6.5%
[0,1]: (176) 15.3%
1150
EMERG
0.0
CHI=84.24; DF=10
URG (367)
??? (6)
(1,2]: (159) 42.6%
(2,3]: (55) 14.7%
(3,4]: (38) 10.2%
(4,5]: (20) 5.4%
>5 : (26) 7.0%
[0,1]: (75) 20.1%
373
ELEC (724)
(1,2]: (385) 53.2%
(2,3]: (117) 16.2%
(3,4]: (63) 8.7%
(4,5]: (26) 3.6%
>5 : (33) 4.6%
[0,1]: (100) 13.8%
724
EMER (53)
(1,2]: (12) 22.6%
(2,3]: (12) 22.6%
(3,4]: (7) 13.2%
(4,5]: (5) 9.4%
>5 : (16) 30.2%
[0,1]: (1) 1.9%
53
PROCED_GRO
0.010275
CHI=26.22; DF=5
CABG+VALV (23)
CABG+OTHE (8)
TRANSPLAN (6)
AORTIC SU (3)
VALVE+OTH (2)
OTHER (0)
(1,2]: (10) 23.8%
(2,3]: (6) 14.3%
(3,4]: (7) 16.7%
(4,5]: (3) 7.1%
>5 : (10) 23.8%
[0,1]: (6) 14.3%
42
CABG (315)
VALVE (16)
(1,2]: (149) 45.0%
(2,3]: (49) 14.8%
(3,4]: (31) 9.4%
(4,5]: (17) 5.1%
>5 : (16) 4.8%
[0,1]: (69) 20.8%
331
REDO
0.000982
CHI=20.56; DF=5
??? (632)
(1,2]: (334) 52.8%
(2,3]: (106) 16.8%
(3,4]: (55) 8.7%
(4,5]: (19) 3.0%
>5 : (23) 3.6%
[0,1]: (95) 15.0%
632
REDO (92)
(1,2]: (51) 55.4%
(2,3]: (11) 12.0%
(3,4]: (8) 8.7%
(4,5]: (7) 7.6%
>5 : (10) 10.9%
[0,1]: (5) 5.4%
92
AGE
0.000038
CHI=46.15; DF=10
[20,63)
(1,2]: (166) 53.4%
(2,3]: (53) 17.0%
(3,4]: (22) 7.1%
(4,5]: (3) 1.0%
>5 : (4) 1.3%
[0,1]: (63) 20.3%
311
[63,76)
(1,2]: (151) 55.5%
(2,3]: (40) 14.7%
(3,4]: (28) 10.3%
(4,5]: (12) 4.4%
>5 : (13) 4.8%
[0,1]: (28) 10.3%
272
[76,88]
(1,2]: (17) 34.7%
(2,3]: (13) 26.5%
(3,4]: (5) 10.2%
(4,5]: (4) 8.2%
>5 : (6) 12.2%
[0,1]: (4) 8.2%
49
NYHA_CLASS
0.015565
CHI=20.43; DF=5
??? (281)
1 (11)
(1,2]: (159) 54.5%
(2,3]: (49) 16.8%
(3,4]: (17) 5.8%
(4,5]: (2) 0.7%
>5 : (3) 1.0%
[0,1]: (62) 21.2%
292
2 (12)
3 (4)
4 (3)
(1,2]: (7) 36.8%
(2,3]: (4) 21.1%
(3,4]: (5) 26.3%
(4,5]: (1) 5.3%
>5 : (1) 5.3%
[0,1]: (1) 5.3%
19
PROCED_GRO
0.00036
CHI=33.62; DF=5
CABG (212)
VALVE (21)
VALVE+OTH (2)
OTHER (1)
TRANSPLAN (0)
AORTIC SU (0)
(1,2]: (138) 58.5%
(2,3]: (37) 15.7%
(3,4]: (22) 9.3%
(4,5]: (7) 3.0%
>5 : (6) 2.5%
[0,1]: (26) 11.0%
236
CABG+VALV (29)
CABG+OTHE (7)
(1,2]: (13) 36.1%
(2,3]: (3) 8.3%
(3,4]: (6) 16.7%
(4,5]: (5) 13.9%
>5 : (7) 19.4%
[0,1]: (2) 5.6%
36

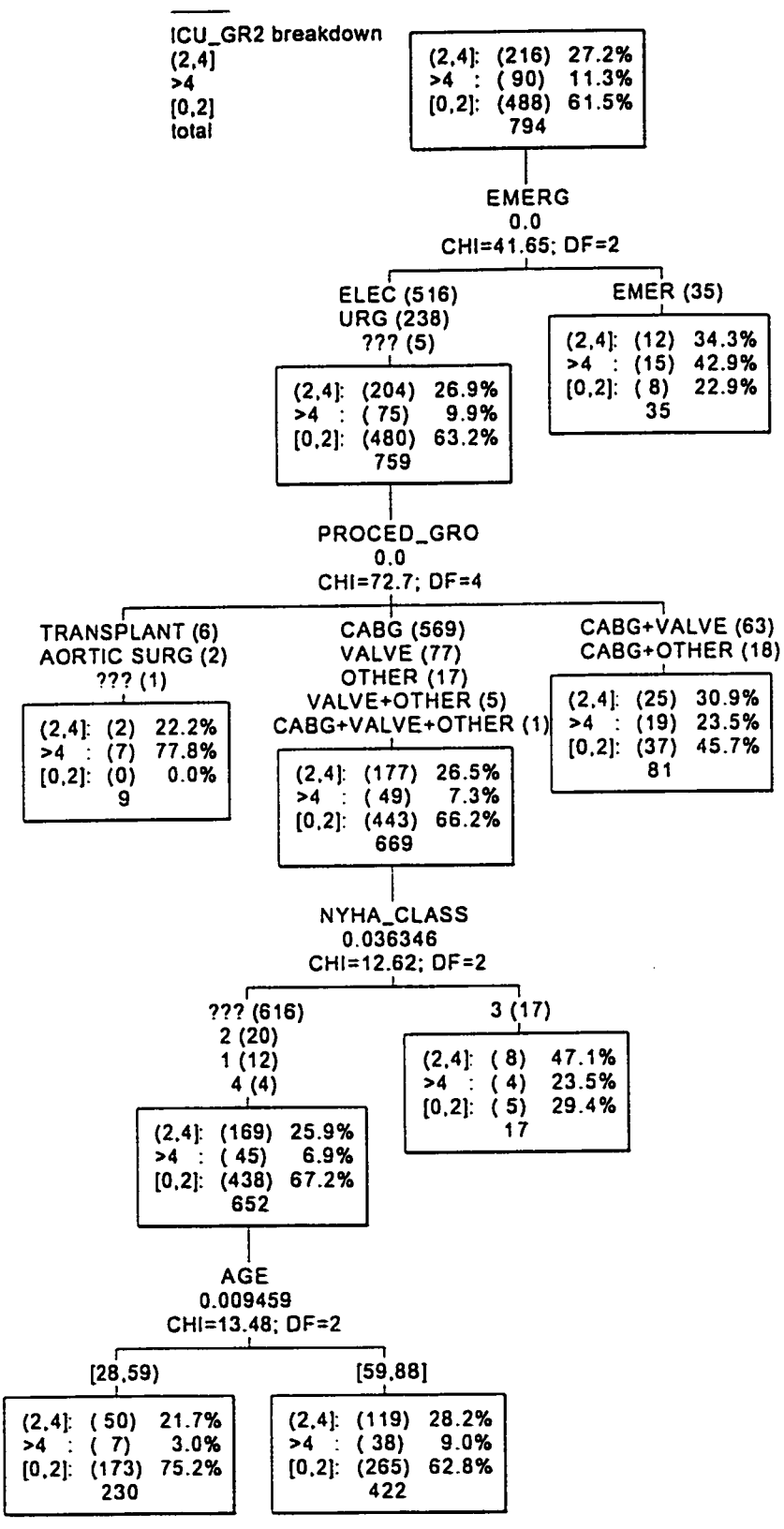

Legend
ICU_GR1 breakdown
(1,2]
(2,3]
(3,4]
(4,5]
>5
[0,1]
total
(1,2]: (556) 48.3%
(2,3]: (184) 16.0%
(3,4]: (108) 9.4%
(4,5]: (51) 4.4%
>5 : (75) 6.5%
[0,1]: (176) 15.3%
1150
IABP
0.0
CHI=157.22; DF=5
??? (1093)
IABP (57)
(1,2]: (554) 50.7%
(2,3]: (172) 15.7%
(3,4]: (99) 9.1%
(4,5]: (44) 4.0%
>5 : (51) 4.7%
[0,1]: (173) 15.8%
1093
(1,2]: (2) 3.5%
(2,3]: (12) 21.1%
(3,4]: (9) 15.8%
(4,5]: (7) 12.3%
>5 : (24) 42.1%
[0,1]: (3) 5.3%
57
PUMPTIME
0.0
CHI=50.95; DF=5
[19,116)
[116,348]
(1,2]: (426) 52.1%
(2,3]: (137) 16.8%
(3,4]: (61) 7.5%
(4,5]: (22) 2.7%
>5 : (26) 3.2%
[0,1]: (145) 17.7%
817
(1,2]: (120) 46.2%
(2,3]: (31) 11.9%
(3,4]: (37) 14.2%
(4,5]: (22) 8.5%
>5 : (23) 8.8%
[0,1]: (27) 10.4%
260
EMERG
0.000267
CHI=38.87; DF=10
URG (260)
??? (4)
ELEC (525)
EMER (28)
(1,2]: (122) 46.2%
(2,3]: (41) 15.5%
(3,4]: (20) 7.6%
(4,5]: (9) 3.4%
>5 : (10) 3.8%
[0,1]: (62) 23.5%
264
(1,2]: (295) 56.2%
(2,3]: (88) 16.8%
(3,4]: (38) 7.2%
(4,5]: (10) 1.9%
>5 : (12) 2.3%
[0,1]: (82) 15.6%
525
(1,2]: (9) 32.1%
(2,3]: (8) 28.6%
(3,4]: (3) 10.7%
(4,5]: (3) 10.7%
>5 : (4) 14.3%
[0,1]: (1) 3.6%
28
AGE
0.001434
CHI=37.25; DF=10
[20,49)
[49,76)
[76,88]
(1,2]: (30) 46.9%
(2,3]: (11) 17.2%
(3,4]: (4) 6.2%
(4,5]: (0) 0.0%
>5 : (0) 0.0%
[0,1]: (19) 29.7%
64
(1,2]: (254) 59.2%
(2,3]: (70) 16.3%
(3,4]: (29) 6.8%
(4,5]: (7) 1.6%
>5 : (9) 2.1%
[0,1]: (60) 14.0%
429
(1,2]: (11) 34.4%
(2,3]: (7) 21.9%
(3,4]: (5) 15.6%
(4,5]: (3) 9.4%
>5 : (3) 9.4%
[0,1]: (3) 9.4%
32

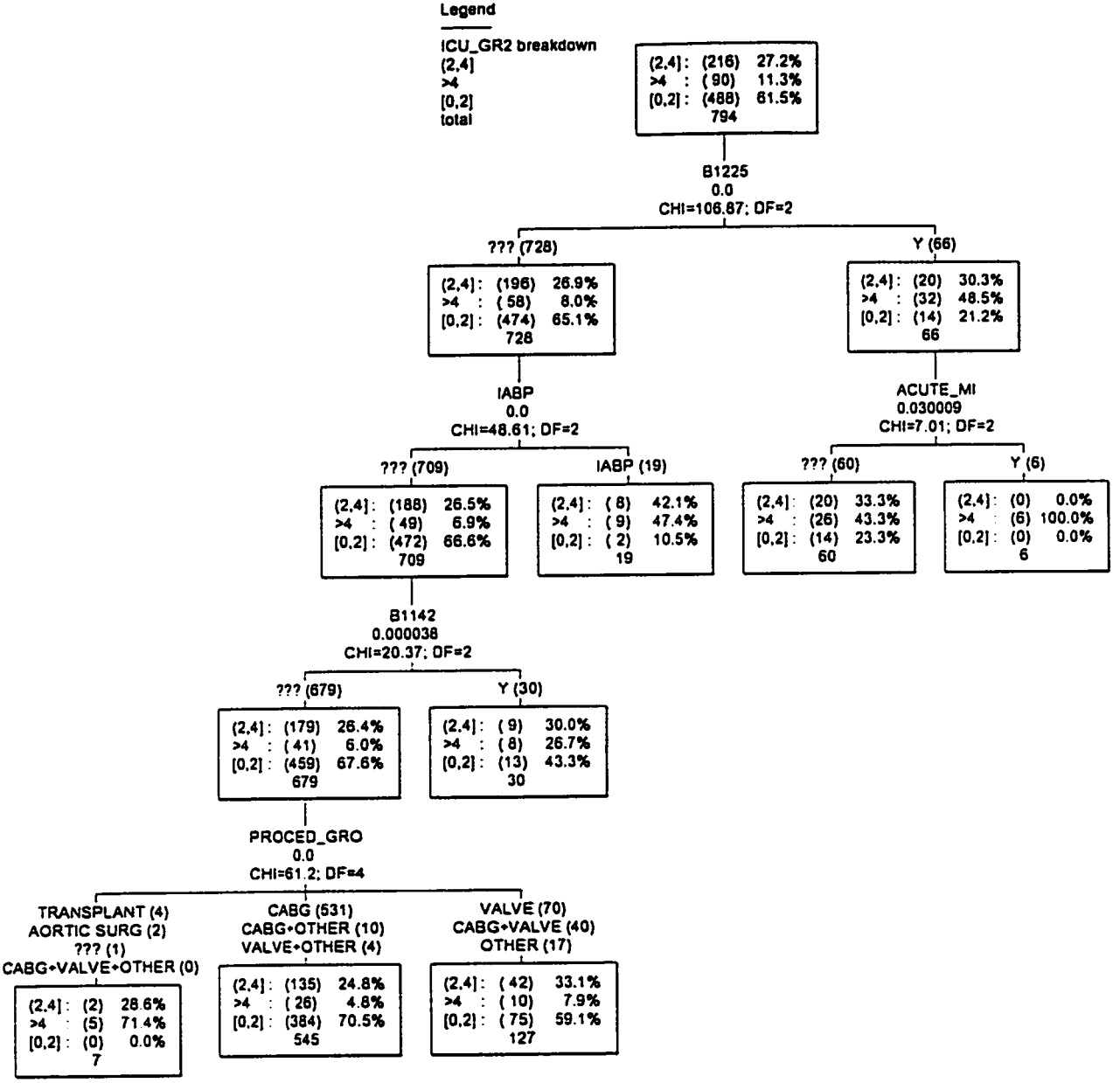

Average ICU_STAY
standard dev
# of records
avg=2.47
std=2.58
n=794
EMERG
0.000021
F=22.88; DF=1, 792
ELEC (516)
URG (238)
??? (5)
EMER (35)
avg=2.38
std=2.5
n=759
avg=4.49
std=3.47
n=35
REDO
0.000439
F=12.47; DF=1, 757
CARDIOG_SH
0.032024
F=5.01; DF=1, 33
??? (671)
REDO (88)
??? (27)
Y (8)
avg=2.27
std=2.16
n=671
avg=3.26
std=4.2
n=88
avg=3.81
std=2.68
n=27
avg=6.77
std=4.9
n=8
AGE
0.000002
F=17.27; DF=2, 668
[28,66)
[66,77)
[77,88]
avg=1.97
std=1.23
n=385
avg=2.44
std=2.23
n=237
avg=3.78
std=5.02
n=49
NYHA_CLASS
0.001705
F=15.21; DF=1, 383
??? (351)
1 (11)
2 (11)
3 (9)
4 (3)
avg=1.91
std=1.16
n=362
avg=2.92
std=1.82
n=23
COPD
0.0475
F=8.05; DF=1, 360
??? (280)
1 (61)
3 (1)
2 (20)
avg=1.87
std=1.11
n=342
avg=2.62
std=1.78
n=20
CHF
0.008757
F=6.95; DF=1, 340
??? (333)
Y (9)
avg=1.84
std=1.1
n=333
avg=2.82
std=1.14
n=9


Legend
Average ICU_STAY
standard dev
# of records
avg=2.47
std=2.58
n=794
B1225
0.0
F=124.88; DF=1, 792
??? (728)
Y (66)
avg=2.19
std=1.76
n=728
avg=5.64
std=5.97
n=66
CARDIOG_SH
0.0
F=30.98; DF=1, 726
??? (725)
Y (3)
avg=2.17
std=1.68
n=725
avg=7.73
std=8.02
n=3
IABP
0.0
F=40.15; DF=1, 723
??? (707)
IABP (18)
avg=2.1
std=1.62
n=707
avg=4.57
std=2.27
n=18
PROCED_GRO
0.0
F=55.31; DF=1, 705
CABG (543)
VALVE (72)
CABG+VALVE (48)
OTHER (17)
CABG+OTHER (11)
VALVE+OTHER (5)
CABG+VALVE+OTHER (0)
TRANSPLANT (7)
AORTIC SURG (3)
??? (1)
avg=5.57
std=2.75
n=11
avg=2.05
std=1.53
n=696
B1223
0.038947
F=4.28; DF=1, 694
??? (617)
Y (79)
avg=2.01
std=1.56
n=617
avg=2.38
std=1.24
n=79
B1142
0.043451
F=4.09; DF=1, 615
B1222
0.022235
F=5.45; DF=1, 77
??? (593)
Y (24)
??? (75)
Y (4)
avg=1.98
std=1.56
n=593
avg=2.64
std=1.46
n=24
avg=2.31
std=1.16
n=75
avg=3.75
std=2.09
n=4

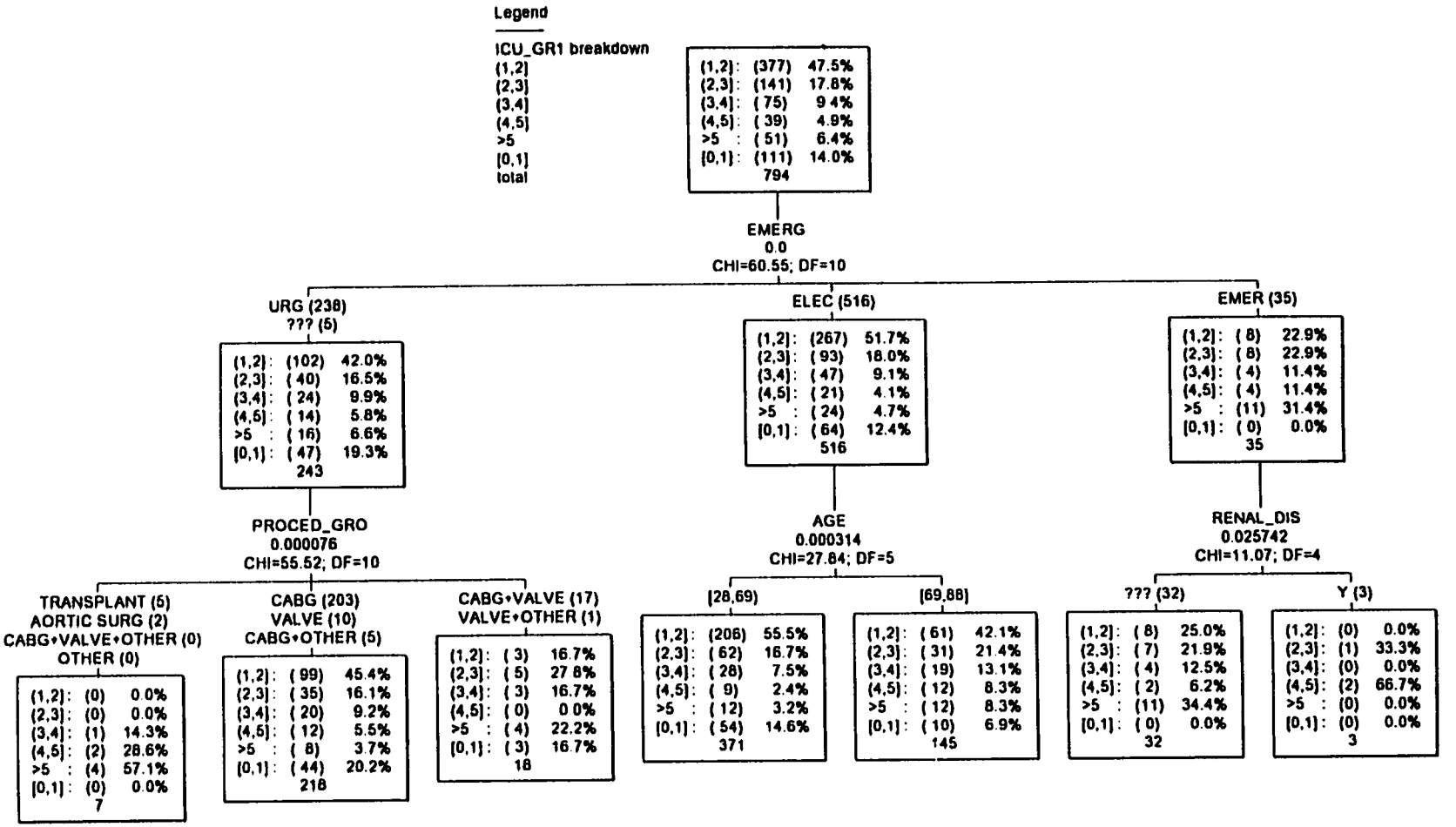

ICU_GR3 breakdown
>2
[0,2]
total
>2 : (306) 38.5%
[0,2]: (488) 61.5%
794
EMERG
0.000011
CHI=23.04; DF=1
ELEC (516)
URG (238)
??? (5)
EMER (35)
>2 : (279) 36.8%
[0,2]: (480) 63.2%
759
>2 : (27) 77.1%
[0,2]: (8) 22.9%
35
NYHA_CLASS
0.000103
CHI=20.24; DF=1
??? (681)
2 (23)
1 (17)
3 (30)
4 (8)
>2 : (252) 35.0%
[0,2]: (469) 65.0%
721
>2 : (27) 71.1%
[0,2]: (11) 28.9%
38
AGE
0.003451
CHI=18.0; DF=2
[28,59)
[59,73)
[73,88]
>2 : (63) 25.9%
[0,2]: (180) 74.1%
243
>2 : (132) 36.8%
[0,2]: (227) 63.2%
359
>2 : (57) 47.9%
[0,2]: (62) 52.1%
119
CHF
0.000155
CHI=14.31; DF=1
??? (348)
Y (11)
>2 : (122) 35.1%
[0,2]: (226) 64.9%
348
>2 : (10) 90.9%
[0,2]: (1) 9.1%
11

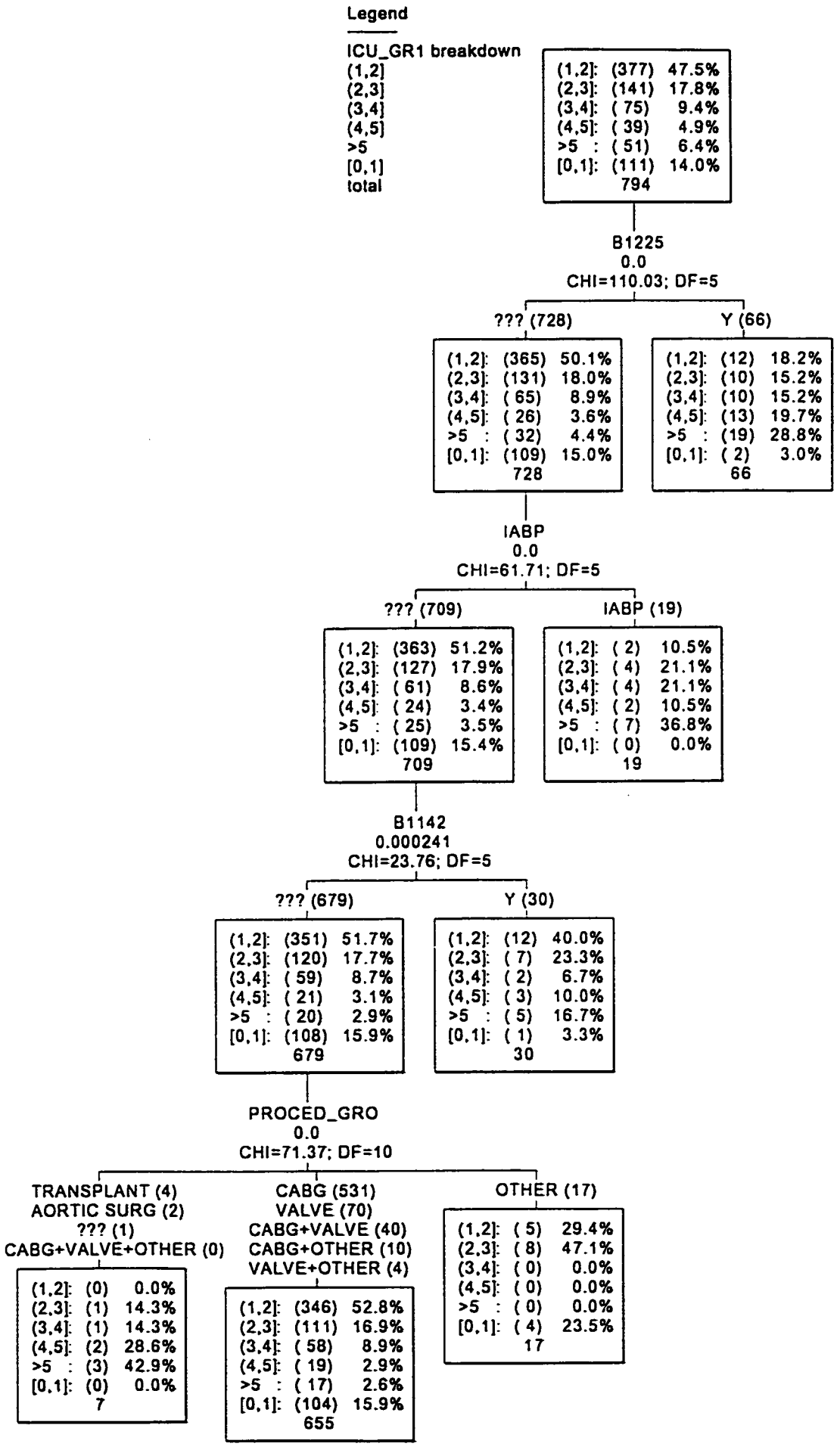

LG94_22.DAT
Legend
ICU_GR3 breakdown
>2
[0,2]
total
>2 : (306) 38.5%
[0,2]: (488) 61.5%
794
B1225
0.0
CHI=49.23; DF=1
??? (728)
>2 : (254) 34.9%
[0,2]: (474) 65.1%
728
Y (66)
>2 : (52) 78.8%
[0,2]: (14) 21.2%
66
IABP
0.0
CHI=25.59; DF=1
CHF
0.036825
CHI=4.36; DF=1
??? (709)
>2 : (237) 33.4%
[0,2]: (472) 66.6%
709
IABP (19)
>2 : (17) 89.5%
[0,2]: (2) 10.5%
19
??? (53)
>2 : (39) 73.6%
[0,2]: (14) 26.4%
53
Y (13)
>2 : (13) 100.0%
[0,2]: (0) 0.0%
13
B1142
0.005829
CHI=7.6; DF=1
??? (679)
>2 : (220) 32.4%
[0,2]: (459) 67.6%
679
Y (30)
>2 : (17) 56.7%
[0,2]: (13) 43.3%
30
PROCED_GRO
0.004844
CHI=20.87; DF=2
CABG (531)
CABG+OTHER (10)
VALVE+OTHER (4)
CABG+VALVE+OTHER (0)
>2 : (161) 29.5%
[0,2]: (384) 70.5%
545
VALVE (70)
CABG+VALVE (40)
OTHER (17)
>2 : (52) 40.9%
[0,2]: (75) 59.1%
127
TRANSPLANT (4)
AORTIC SURG (2)
??? (1)
>2 : (7) 100.0%
[0,2]: (0) 0.0%
7


ICU_GR2 breakdown
(2,4]
>4
[0,2]
total
(2,4]: (216) 27.2%
>4 : (90) 11.3%
[0,2]: (488) 61.5%
794
EMERG
0.0
CHI=41.65; DF=2
ELEC (516)
URG (238)
??? (5)
EMER (35)
(2,4]: (204) 26.9%
>4 : (75) 9.9%
[0,2]: (480) 63.2%
759
(2,4]: (12) 34.3%
>4 : (15) 42.9%
[0,2]: (8) 22.9%
35
PROCED_GRO
0.0
CHI=72.7; DF=4
TRANSPLANT (6)
AORTIC SURG (2)
??? (1)
CABG (569)
VALVE (77)
OTHER (17)
VALVE+OTHER (5)
CABG+VALVE+OTHER (1)
CABG+VALVE (63)
CABG+OTHER (18)
(2,4]: (2) 22.2%
>4 : (7) 77.8%
[0,2]: (0) 0.0%
9
(2,4]: (177) 26.5%
>4 : (49) 7.3%
[0,2]: (443) 66.2%
669
(2,4]: (25) 30.9%
>4 : (19) 23.5%
[0,2]: (37) 45.7%
81
NYHA_CLASS
0.036346
CHI=12.62; DF=2
??? (616)
2 (20)
1 (12)
4 (4)
3 (17)
(2,4]: (169) 25.9%
>4 : (45) 6.9%
[0,2]: (438) 67.2%
652
(2,4]: (8) 47.1%
>4 : (4) 23.5%
[0,2]: (5) 29.4%
17
AGE
0.009459
CHI=13.48; DF=2
[28,59)
[59,88]
(2,4]: (50) 21.7%
>4 : (7) 3.0%
[0,2]: (173) 75.2%
230
(2,4]: (119) 28.2%
>4 : (38) 9.0%
[0,2]: (265) 62.8%
422


Legend
ICU_GR2 breakdown
(2,4]
>4
[0,2]
total
(2,4]: (216) 27.2%
>4 : (90) 11.3%
[0,2]: (488) 61.5%
794
B1225
0.0
CHI=106.87; DF=2
??? (728)
Y (66)
(2,4]: (196) 26.9%
>4 : (58) 8.0%
[0,2]: (474) 65.1%
728
(2,4]: (20) 30.3%
>4 : (32) 48.5%
[0,2]: (14) 21.2%
66
IABP
0.0
CHI=48.61; DF=2
ACUTE_MI
0.030009
CHI=7.01; DF=2
??? (709)
IABP (19)
??? (60)
Y (6)
(2,4]: (188) 26.5%
>4 : (49) 6.9%
[0,2]: (472) 66.6%
709
(2,4]: (8) 42.1%
>4 : (9) 47.4%
[0,2]: (2) 10.5%
19
(2,4]: (20) 33.3%
>4 : (26) 43.3%
[0,2]: (14) 23.3%
60
(2,4]: (0) 0.0%
>4 : (6) 100.0%
[0,2]: (0) 0.0%
6
B1142
0.000038
CHI=20.37; DF=2
??? (679)
Y (30)
(2,4]: (179) 26.4%
>4 : (41) 6.0%
[0,2]: (459) 67.6%
679
(2,4]: (9) 30.0%
>4 : (8) 26.7%
[0,2]: (13) 43.3%
30
PROCED_GRO
0.0
CHI=61.2; DF=4
TRANSPLANT (4)
AORTIC SURG (2)
??? (1)
CABG+VALVE+OTHER (0)
CABG (531)
CABG+OTHER (10)
VALVE+OTHER (4)
VALVE (70)
CABG+VALVE (40)
OTHER (17)
(2,4]: (2) 28.6%
>4 : (5) 71.4%
[0,2]: (0) 0.0%
7
(2,4]: (135) 24.8%
>4 : (26) 4.8%
[0,2]: (384) 70.5%
545
(2,4]: (42) 33.1%
>4 : (10) 7.9%
[0,2]: (75) 59.1%
127

LG94_41.DAT


Legend
ICU_GR1 breakdown
(1,2]
(2,3]
(3,4]
(4,5]
>5
[0,1]
total
(1,2]: (377) 47.5%
(2,3]: (141) 17.8%
(3,4]: (75) 9.4%
(4,5]: (39) 4.9%
>5 : (51) 6.4%
[0,1]: (111) 14.0%
794
EMERG
0.0
CHI=60.55; DF=10
URG (238)
??? (5)
(1,2]: (102) 42.0%
(2,3]: (40) 16.5%
(3,4]: (24) 9.9%
(4,5]: (14) 5.8%
>5 : (16) 6.6%
[0,1]: (47) 19.3%
243
ELEC (516)
(1,2]: (267) 51.7%
(2,3]: (93) 18.0%
(3,4]: (47) 9.1%
(4,5]: (21) 4.1%
>5 : (24) 4.7%
[0,1]: (64) 12.4%
516
EMER (35)
(1,2]: (8) 22.9%
(2,3]: (8) 22.9%
(3,4]: (4) 11.4%
(4,5]: (4) 11.4%
>5 : (11) 31.4%
[0,1]: (0) 0.0%
35
PROCED_GRO
0.000076
CHI=55.52; DF=10
AGE
0.000314
CHI=27.84; DF=5
RENAL_DIS
0.025742
CHI=11.07; DF=4
TRANSPLANT (5)
AORTIC SURG (2)
CABG+VALVE+OTHER (0)
OTHER (0)
(1,2]: (0) 0.0%
(2,3]: (0) 0.0%
(3,4]: (1) 14.3%
(4,5]: (2) 28.6%
>5 : (4) 57.1%
[0,1]: (0) 0.0%
7
CABG (203)
VALVE (10)
CABG+OTHER (5)
(1,2]: (99) 45.4%
(2,3]: (35) 16.1%
(3,4]: (20) 9.2%
(4,5]: (12) 5.5%
>5 : (8) 3.7%
[0,1]: (44) 20.2%
218
CABG+VALVE (17)
VALVE+OTHER (1)
(1,2]: (3) 16.7%
(2,3]: (5) 27.8%
(3,4]: (3) 16.7%
(4,5]: (0) 0.0%
>5 : (4) 22.2%
[0,1]: (3) 16.7%
18
[28,69)
(1,2]: (206) 55.5%
(2,3]: (62) 16.7%
(3,4]: (28) 7.5%
(4,5]: (9) 2.4%
>5 : (12) 3.2%
[0,1]: (54) 14.6%
371
[69,88]
(1,2]: (61) 42.1%
(2,3]: (31) 21.4%
(3,4]: (19) 13.1%
(4,5]: (12) 8.3%
>5 : (12) 8.3%
[0,1]: (10) 6.9%
145
??? (32)
(1,2]: (8) 25.0%
(2,3]: (7) 21.9%
(3,4]: (4) 12.5%
(4,5]: (2) 6.2%
>5 : (11) 34.4%
[0,1]: (0) 0.0%
32
Y (3)
(1,2]: (0) 0.0%
(2,3]: (1) 33.3%
(3,4]: (0) 0.0%
(4,5]: (2) 66.7%
>5 : (0) 0.0%
[0,1]: (0) 0.0%
3


Legend
ICU_GR1 breakdown
(1,2]
(2,3]
(3,4]
(4,5]
>5
[0,1]
total
(1,2]: (377) 47.5%
(2,3]: (141) 17.8%
(3,4]: (75) 9.4%
(4,5]: (39) 4.9%
>5 : (51) 6.4%
[0,1]: (111) 14.0%
794
B1225
0.0
CHI=110.03; DF=5
??? (728)
Y (66)
(1,2]: (365) 50.1%
(2,3]: (131) 18.0%
(3,4]: (65) 8.9%
(4,5]: (26) 3.6%
>5 : (32) 4.4%
[0,1]: (109) 15.0%
728
(1,2]: (12) 18.2%
(2,3]: (10) 15.2%
(3,4]: (10) 15.2%
(4,5]: (13) 19.7%
>5 : (19) 28.8%
[0,1]: (2) 3.0%
66
IABP
0.0
CHI=61.71; DF=5
??? (709)
IABP (19)
(1,2]: (363) 51.2%
(2,3]: (127) 17.9%
(3,4]: (61) 8.6%
(4,5]: (24) 3.4%
>5 : (25) 3.5%
[0,1]: (109) 15.4%
709
(1,2]: (2) 10.5%
(2,3]: (4) 21.1%
(3,4]: (4) 21.1%
(4,5]: (2) 10.5%
>5 : (7) 36.8%
[0,1]: (0) 0.0%
19
B1142
0.000241
CHI=23.76; DF=5
??? (679)
Y (30)
(1,2]: (351) 51.7%
(2,3]: (120) 17.7%
(3,4]: (59) 8.7%
(4,5]: (21) 3.1%
>5 : (20) 2.9%
[0,1]: (108) 15.9%
679
(1,2]: (12) 40.0%
(2,3]: (7) 23.3%
(3,4]: (2) 6.7%
(4,5]: (3) 10.0%
>5 : (5) 16.7%
[0,1]: (1) 3.3%
30
PROCED_GRO
0.0
CHI=71.37; DF=10
TRANSPLANT (4)
AORTIC SURG (2)
??? (1)
CABG+VALVE+OTHER (0)
CABG (531)
VALVE (70)
CABG+VALVE (40)
CABG+OTHER (10)
VALVE+OTHER (4)
OTHER (17)
(1,2]: (0) 0.0%
(2,3]: (1) 14.3%
(3,4]: (1) 14.3%
(4,5]: (2) 28.6%
>5 : (3) 42.9%
[0,1]: (0) 0.0%
7
(1,2]: (346) 52.8%
(2,3]: (111) 16.9%
(3,4]: (58) 8.9%
(4,5]: (19) 2.9%
>5 : (17) 2.6%
[0,1]: (104) 15.9%
655
(1,2]: (5) 29.4%
(2,3]: (8) 47.1%
(3,4]: (0) 0.0%
(4,5]: (0) 0.0%
>5 : (0) 0.0%
[0,1]: (4) 23.5%
17

B. Estimation of test error.

Estimation of test error on 1995 data set.

I. Continuous dependent variable

	Relative Error	Percentage of variance explained
Preoper. models		
sm94_11	0.9401	5.99
lg94_11	0.9161	8.39
Postoper. models		
sm94_12	0.9108	8.92
lg94_12	0.7984	20.2

II. Categorical dependent variable

Test error for SM94_21

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
>2 :(418) 36.3%	>2 :(299) 27.4%	(119) 9.0%
[0,2]:(732) 63.7%	[0,2]:(793) 72.6%	(61) 9.0%
1150	1092	58

Misclassification rate: 0.261
Accuracy: 73.901%

Test error for SM94_22

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
>2 :(418) 36.3%	>2 :(299) 27.4%	(119) 9.0%
[0,2]:(732) 63.7%	[0,2]:(793) 72.6%	(61) 9.0%
1150	1092	58

Misclassification rate: 0.2308
Accuracy: 76.923%

Test error for SM94_31

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
(2,4]:(292) 25.4%	(2,4]:(206) 18.9%	(86) 6.5%
>4 :(126) 11.0%	>4 :(93) 8.5 %	(33) 2.4%
[0,2]:(732) 63.7%	[0,2]:(793) 72.6%	(61) 9.0%
1150	1092	58

Misclassification rate: 0.2738
Accuracy: 72.619%

Test error for SM94_32

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
(2,4]:(292) 25.4%	(2,4]:(206) 18.9%	(86) 6.5%
>4 :(126) 11.0%	>4 :(93) 8.5 %	(33) 2.4%
[0,2]:(732) 63.7%	[0,2]:(793) 72.6%	(61) 9.0%
1150	1092	58

Misclassification rate: 0.2546
Accuracy: 74.542%

Test error for SM94_41

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
(1,2]:(556) 48.3%	(1,2]:(481) 44.0%	(75) 4.3 %
(2,3]:(184) 16 %	(2,3]:(134) 12.3%	(50) 3.7 %
(3,4]:(108) 9.4 %	(3,4]:(72) 6.6 %	(36) 2.8 %
(4,5]:(51) 4.4 %	(4,5]:(26) 2.4 %	(25) 2.1 %
>5 :(75) 6.5 %	>5 :(67) 6.1 %	(8) 0.4 %
[0,1]:(176) 15.3%	[0,1]:(312) 28.6%	(136) 13.3%
1150	1092	58

Misclassification rate: 0.5549
Accuracy: 44.505%

Test error for SM94_42

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
(1,2]:(556) 48.3%	(1,2]:(481) 44.0%	(75) 4.3 %
(2,3]:(184) 16 %	(2,3]:(134) 12.3%	(50) 3.7 %
(3,4]:(108) 9.4 %	(3,4]:(72) 6.6 %	(36) 2.8 %
(4,5]:(51) 4.4 %	(4,5]:(26) 2.4 %	(25) 2.1 %
>5 :(75) 6.5 %	>5 :(67) 6.1 %	(8) 0.4 %
[0,1]:(176) 15.3%	[0,1]:(312) 28.6%	(136) 13.3%
1150	1092	58

Misclassification rate: 0.544
Accuracy: 45.604%

Test error for lg94_21

Error rate profile table

Learning sampl frequency distribution	Test sample frequency distribution	Difference
>2 :(306) 38.5%	>2 :(196) 29.9%	(110) 8.7%
[0,2]:(488) 61.5%	[0,2]:(460) 70.1%	(28) 8.7%
794	656	138

Misclassification rate: 0.3018
Accuracy: 69.817%

Test error for lg94_22

Error rate profile table

Learning sampl frequency distribution		Test sample frequency distribution		Difference	
>2	(306) 38.5%	>2	(196) 29.9%	(110)	8.7%
[0,2]	(488) 61.5%	[0,2]	(460) 70.1%	(28)	8.7%
	794		656		138

Misclassification rate: 0.2378
Accuracy: 76.22%

Test error for lg94_31

Error rate profile table

Learning sampl frequency distribution		Test sample frequency distribution		Difference	
(2,4]	(216) 27.2%	(2,4]	(144) 22.0%	(72)	5.3%
>4	(90) 11.3%	>4	(52) 7.9 %	(38)	3.4%
[0,2]	(488) 61.5%	[0,2]	(460) 70.1%	(28)	8.7%
	794		656		138

Misclassification rate: 0.3034
Accuracy: 69.665%

Test error for lg94_32

Error rate profile table

Learning sampl frequency distribution		Test sample frequency distribution		Difference	
(2,4]	(216) 27.2%	(2,4]	(144) 22.0%	(72)	5.3%
>4	(90) 11.3%	>4	(52) 7.9 %	(38)	3.4%
[0,2]	(488) 61.5%	[0,2]	(460) 70.1%	(28)	8.7%
	794		656		138

Misclassification rate: 0.2805
Accuracy: 71.951%

Visual Basic program for calculation information based score for the LG94_** models.

```
Function Entropy(ParamArray PC())
  For Each X In PC
    LOG2 = Log(X) / Log(2)
    A = X * LOG2
    S = S + A
  Next X
  Entropy = -S
End Function

Function CALC_INF_SCORE(PRIOR_PROB, POST_PROB)
  If POST_PROB >= PRIOR_PROB Then
    CALC_INF_SCORE = -(Log(PRIOR_PROB) / Log(2)) + (Log(POST_PROB) / Log(2))
  Else
    CALC_INF_SCORE = -(-(Log(1 - PRIOR_PROB) / Log(2)) + (Log(1 - POST_PROB) / Log(2)))
  End If
End Function


'File lg94_21

Function LG94_21(ICU_GR3, EMERG, NYHA_CLASS, AGE, CHF)

  Select Case ICU_GR3
    Case ">2"
      PRIOR_PROB = 0.299
    Case "[0,2]"
      PRIOR_PROB = 0.701
  End Select


' RULE_1
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (NYHA_CLASS = "" Or
NYHA_CLASS = "2" Or NYHA_CLASS = "1") And _
   AGE < 59 Then

       Select Case ICU_GR3
    Case ">2"
      POST_PROB = 0.259   ' ICU_GR3 = >2    25.9%
    Case "[0,2]"
      POST_PROB = 0.741   ' ICU_GR3 = [0,2]  74.1%

  End Select
End If



' RULE_2
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (NYHA_CLASS = "" Or
NYHA_CLASS = "2" Or NYHA_CLASS = "1") And _
```

Test error for lg94_41

Error rate profile table

Learning sampl frequency distribution		Test sample frequency distribution		Difference	
(1,2] :	(377) 47.5%	(1,2] :	(289) 44.1%	(88)	3.4 %
(2,3] :	(141) 17.8%	(2,3] :	(94) 14.3%	(47)	3.4 %
(3,4] :	(75) 9.4 %	(3,4] :	(50) 7.6 %	(25)	1.8 %
(4,5] :	(39) 4.9 %	(4,5] :	(16) 2.4 %	(23)	2.5 %
>5 :	(51) 6.4 %	>5 :	(36) 5.5 %	(15)	0.9 %
[0,1] :	(111) 14.0%	[0,1] :	(171) 26.1%	(60)	12.1%
	794		656		138

Misclassification rate: 0.564
Accuracy: 43.598%

Test error for lg94_42

Error rate profile table

Learning sampl frequency distribution		Test sample frequency distribution		Difference	
(1,2] :	(377) 47.5%	(1,2] :	(289) 44.1%	(88)	3.4 %
(2,3] :	(141) 17.8%	(2,3] :	(94) 14.3%	(47)	3.4 %
(3,4] :	(75) 9.4 %	(3,4] :	(50) 7.6 %	(25)	1.8 %
(4,5] :	(39) 4.9 %	(4,5] :	(16) 2.4 %	(23)	2.5 %
>5 :	(51) 6.4 %	>5 :	(36) 5.5 %	(15)	0.9 %
[0,1] :	(111) 14.0%	[0,1] :	(171) 26.1%	(60)	12.1%
	794		656		138

Misclassification rate: 0.5457
Accuracy: 45.427%

```
  If B1225 = "" And IABP = "" And B1142 = "" And _
  (PROCED_GRO = "TRANSPLANT" Or PROCED_GRO = "AORTIC SURG" Or PROCED_GRO =
"") Then
        Select Case ICU_GR3
     Case ">2"
       POST_PROB = 1   ' ICU_GR3 = >2      100.0%
     Case "[0,2]"
       POST_PROB = 0   ' ICU_GR3 = [0,2]    0%
  End Select
End If



' RULE_4
   If B1225 = "" And IABP = "" And B1142 = "Y" Then
      Select Case ICU_GR3
     Case ">2"
       POST_PROB = 0.567   ' ICU_GR3 = >2      56.7%
     Case "[0,2]"
       POST_PROB = 0.433   ' ICU_GR3 = [0,2]   43.3%
  End Select

End If

' RULE_5
   If B1225 = "" And IABP = "IABP" Then
      Select Case ICU_GR3
     Case ">2"
       POST_PROB = 0.895   ' ICU_GR3 = >2      89.5%
     Case "[0,2]"
       POST_PROB = 0.105   ' ICU_GR3 = [0,2]   10.5%
  End Select

End If



' RULE_6
   If B1225 = "Y" And CHF = "" Then
       Select Case ICU_GR3
     Case ">2"
       POST_PROB = 0.736   ' ICU_GR3 = >2      73.6%
     Case "[0,2]"
       POST_PROB = 0.264   ' ICU_GR3 = [0,2]   26.4%
  End Select

End If



' RULE_7
   If B1225 = "Y" And CHF = "Y" Then

     Select Case ICU_GR3
```

```
    Case ">2"
      POST_PROB = 1   ' ICU_GR3 = >2       100.0%
    Case "[0,2]"
      POST_PROB = 0   ' ICU_GR3 = [0,2]    0%
  End Select
  End If

lg94_22 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function


'File lg94_31
  Function lg94_31(ICU_GR2, EMERG, NYHA_CLASS, AGE, CHF, PROCED_GRO)

  Select Case ICU_GR2
    Case "(2,4]"
      PRIOR_PROB = 0.22
    Case ">4"
      PRIOR_PROB = 0.079
    Case "[0,2]"
      PRIOR_PROB = 0.701
  End Select

' RULE_1
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "TRANSPLANT" Or PROCED_GRO = "AORTIC SURG" Or PROCED_GRO =
"") Then
      Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.222   'ICU_GR2 = (2,4]   22.2%
    Case ">4"
      POST_PROB = 0.0778  ' ICU_GR2 = >4     77.8%
    Case "[0,2]"
      POST_PROB = 0   ' ICU_GR2 = [0,2]   0%
  End Select
End If

' RULE_2
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG" Or PROCED_GRO = "CABG+VALVE+OTHER" Or PROCED_GRO =
"VALVE+OTHER" Or PROCED_GRO = "OTHER" Or PROCED_GRO = "VALVE") And _
    (NYHA_CLASS = "" Or NYHA_CLASS = "2" Or NYHA_CLASS = "1" Or NYHA_CLASS = "4")
And _
    AGE < 59 Then
      Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.217   'ICU_GR2 = (2,4]   21.7%
    Case ">4"
      POST_PROB = 0.03    ' ICU_GR2 = >4     3.0%
    Case "[0,2]"
      POST_PROB = 0.752   ' ICU_GR2 = [0,2]   75.2%
  End Select
End If
```

```
'RULE_3
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG" Or PROCED_GRO = "CABG+VALVE+OTHER" Or PROCED_GRO =
"VALVE+OTHER" Or PROCED_GRO = "OTHER" Or PROCED_GRO = "VALVE") And _
     (NYHA_CLASS = "" Or NYHA_CLASS = "2" Or NYHA_CLASS = "1" Or NYHA_CLASS = "4")
And _
     AGE >= 59 Then
        Select Case ICU_GR2
     Case "(2,4]"
       POST_PROB = 0.282  'ICU_GR2 = (2,4]  28.2%
     Case ">4"
       POST_PROB = 0.09   ' ICU_GR2 = >4     9.0%
     Case "[0,2]"
       POST_PROB = 0.628  ' ICU_GR2 = [0,2]  62.8%
  End Select
End If


'RULE_4
   If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG" Or PROCED_GRO = "CABG+VALVE+OTHER" Or PROCED_GRO =
"VALVE+OTHER" Or PROCED_GRO = "OTHER" Or PROCED_GRO = "VALVE") And _
     NYHA_CLASS = "3" Then
      Select Case ICU_GR2
     Case "(2,4]"
       POST_PROB = 0.471  'ICU_GR2 = (2,4]  47.1%
     Case ">4"
       POST_PROB = 0.235  ' ICU_GR2 = >4     23.5%
     Case "[0,2]"
       POST_PROB = 0.294  ' ICU_GR2 = [0,2]  29.4%
  End Select
End If


'RULE_5
 If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG+OTHER" Or PROCED_GRO = "CABG+VALVE") Then
       Select Case ICU_GR2
     Case "(2,4]"
       POST_PROB = 0.309  'ICU_GR2 = (2,4]  30.9%
     Case ">4"
       POST_PROB = 0.235  ' ICU_GR2 = >4     23.5%
     Case "[0,2]"
       POST_PROB = 0.457  ' ICU_GR2 = [0,2]  45.7%
  End Select
 End If


'RULE_6
```

```
    If EMERG = "EMER" Then
      Select Case ICU_GR2
      Case "(2,4]"
        POST_PROB = 0.343   'ICU_GR2 = (2,4]   34.3%
      Case ">4"
        POST_PROB = 0.429   ' ICU_GR2 = >4     42.9%
      Case "[0,2]"
        POST_PROB = 0.229   ' ICU_GR2 = [0,2]  22.9%
    End Select
  End If

  lg94_31 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function

'File lg94_32
  Function lg94_32(ICU_GR2, B1225, IABP, B1142, ACUTE_MI, PROCED_GRO)

  Select Case ICU_GR2
    Case "(2,4]"
      PRIOR_PROB = 0.22
    Case ">4"
      PRIOR_PROB = 0.079
    Case "[0,2]"
      PRIOR_PROB = 0.701
  End Select


' RULE_1
  If B1225 = "" And IABP = "" And B1142 = "" And _
  (PROCED_GRO = "TRANSPLANT" Or PROCED_GRO = "AORTIC SURG" Or PROCED_GRO = ""
Or PROCED_GRO = "CABG+VALVE+OTHER") Then
       Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.286   'ICU_GR2 = (2,4]   28.6%
    Case ">4"
      POST_PROB = 0.714   ' ICU_GR2 = >4     71.4%
    Case "[0,2]"
      POST_PROB = 0   ' ICU_GR2 = [0,2]   0%
  End Select
End If



' RULE_2
   If B1225 = "" And IABP = "" And B1142 = "" And _
  (PROCED_GRO = "CABG" Or PROCED_GRO = "CABG+OTHER" Or PROCED_GRO =
"VALVE+OTHER") Then
       Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.248   'ICU_GR2 = (2,4]   24.8%
    Case ">4"
      POST_PROB = 0.048   ' ICU_GR2 = >4      4.8%
    Case "[0,2]"
```

```
        POST_PROB = 0.705   ' ICU_GR2 = [0,2]   70.5%

    End Select
End If



' RULE_3
    If B1225 = "" And IABP = "" And B1142 = "" And _
    (PROCED_GRO = "VALVE" Or PROCED_GRO = "CABG+VALVE" Or PROCED_GRO =
"OTHER") Then
        Select Case ICU_GR2
        Case "(2,4]"
            POST_PROB = 0.331   'ICU_GR2 = (2,4]   33.1%
        Case ">4"
            POST_PROB = 0.079   ' ICU_GR2 = >4      7.9%
        Case "[0,2]"
            POST_PROB = 0.591   ' ICU_GR2 = [0,2]   59.1%

    End Select
End If



' RULE_4
    If B1225 = "" And IABP = "" And B1142 = "Y" Then
        Select Case ICU_GR2
        Case "(2,4]"
            POST_PROB = 0.3     'ICU_GR2 = (2,4]   30.0%
        Case ">4"
            POST_PROB = 0.267   ' ICU_GR2 = >4      26.7%
        Case "[0,2]"
            POST_PROB = 0.433   ' ICU_GR2 = [0,2]   43.3%

    End Select
End If

' RULE_5
    If B1225 = "" And IABP = "IABP" Then
        Select Case ICU_GR2
        Case "(2,4]"
            POST_PROB = 0.421   'ICU_GR2 = (2,4]   42.1%
        Case ">4"
            POST_PROB = 0.474   ' ICU_GR2 = >4      47.4%
        Case "[0,2]"
            POST_PROB = 0.105   ' ICU_GR2 = [0,2]   10.5%

    End Select
End If
```

```
' RULE_6
  If B1225 = "Y" And ACUTE_MI = "" Then
      Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.333  'ICU_GR2 = (2,4]   33.3%
    Case ">4"
      POST_PROB = 0.433  ' ICU_GR2 = >4     43.3%
    Case "[0,2]"
      POST_PROB = 0.233  ' ICU_GR2 = [0,2]  23.3%

  End Select
End If



' RULE_7
  If B1225 = "Y" And ACUTE_MI = "Y" Then

    Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.333  ' ICU_GR2 = (2,4]   0%
    Case ">4"
      POST_PROB = 0.433  ' ICU_GR2 = >4     100.0%
    Case "[0,2]"
      POST_PROB = 0.233  ' ICU_GR2 = [0,2]  0%

  End Select
End If

lg94_32 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function

'File lg94_41
  Function lg94_41(ICU_GR1, RENAL_DIS, EMERG, AGE, PROCED_GRO)

  Select Case ICU_GR1
    Case "(1,2]"
      PRIOR_PROB = 0.441
    Case "(2,3]"
      PRIOR_PROB = 0.143
    Case "(3,4]"
      PRIOR_PROB = 0.076
    Case "(4,5]"
      PRIOR_PROB = 0.024
    Case ">5"
      PRIOR_PROB = 0.055
    Case "[0,1]"
      PRIOR_PROB = 0.261
  End Select


 ' RULE_1
```

```
  If (EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "TRANSPLANT" Or PROCED_GRO = "CABG+VALVE+OTHER" Or
PROCED_GRO = "AORTIC SURG" Or PROCED_GRO = "OTHER") Then
    Select Case ICU_GR1
     Case "(1,2]"
       POST_PROB = 0  ' ICU_GR1 = (1,2]  0%
     Case "(2,3]"
       POST_PROB = 0  ' ICU_GR1 = (2,3]  0%
     Case "(3,4]"
       POST_PROB = 0.143  ' ICU_GR1 = (3,4]  14.3%
     Case "(4,5]"
       POST_PROB = 0.286  ' ICU_GR1 = (4,5]  28.6%
     Case ">5"
       POST_PROB = 0.571  ' ICU_GR1 = >5     57.1%
     Case "[0,1]"
       POST_PROB = 0  ' ICU_GR1 = [0,1]  0%
  End Select
End If

' RULE_2
   If (EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO = "CABG+OTHER")
Then
    Select Case ICU_GR1
     Case "(1,2]"
       POST_PROB = 0.454  ' ICU_GR1 = (1,2]  45.4%
     Case "(2,3]"
       POST_PROB = 0.161  ' ICU_GR1 = (2,3]  16.1%
     Case "(3,4]"
       POST_PROB = 0.092  ' ICU_GR1 = (3,4]  9.2%
     Case "(4,5]"
       POST_PROB = 0.055  ' ICU_GR1 = (4,5]  5.5%
     Case ">5"
       POST_PROB = 0.037  ' ICU_GR1 = >5     3.7%
     Case "[0,1]"
       POST_PROB = 0.202  ' ICU_GR1 = [0,1]  20.2%
  End Select
End If



' RULE_3
   If (EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG+VALVE" Or PROCED_GRO = "VALVE+OTHER") Then
    Select Case ICU_GR1
     Case "(1,2]"
       POST_PROB = 0.167  ' ICU_GR1 = (1,2]  16.7%
     Case "(2,3]"
       POST_PROB = 0.278  ' ICU_GR1 = (2,3]  27.8%
     Case "(3,4]"
       POST_PROB = 0.167  ' ICU_GR1 = (3,4]  16.7%
     Case "(4,5]"
       POST_PROB = 0      ' ICU_GR1 = (4,5]  0%
     Case ">5"
```

```
        POST_PROB = 0.222  ' ICU_GR1 = >5     22.2%
      Case "[0,1]"
        POST_PROB = 0.167  ' ICU_GR1 = [0,1]  16.7%
    End Select
  End If



' RULE_4
    If EMERG = "ELEC" And AGE < 69 Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.555  ' ICU_GR1 = (1,2]  55.5%
      Case "(2,3]"
        POST_PROB = 0.167  ' ICU_GR1 = (2,3]  16.7%
      Case "(3,4]"
        POST_PROB = 0.075  ' ICU_GR1 = (3,4]  7.5%
      Case "(4,5]"
        POST_PROB = 0.024  ' ICU_GR1 = (4,5]  2.4%
      Case ">5"
        POST_PROB = 0.032  ' ICU_GR1 = >5     3.2%
      Case "[0,1]"
        POST_PROB = 0.146  ' ICU_GR1 = [0,1]  14.6%
    End Select
  End If



' RULE_5
    If EMERG = "ELEC" And AGE >= 69 Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.421  ' ICU_GR1 = (1,2]  42.1%
      Case "(2,3]"
        POST_PROB = 0.214  ' ICU_GR1 = (2,3]  21.4%
      Case "(3,4]"
        POST_PROB = 0.131  ' ICU_GR1 = (3,4]  13.1%
      Case "(4,5]"
        POST_PROB = 0.083  ' ICU_GR1 = (4,5]  8.3%
      Case ">5"
        POST_PROB = 0.083  ' ICU_GR1 = >5     8.3%
      Case "[0,1]"
        POST_PROB = 0.069  ' ICU_GR1 = [0,1]  6.9%

    End Select
  End If


' RULE_6
    If EMERG = "ELEC" And RENAL_DIS = "" Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.25   ' ICU_GR1 = (1,2]  25.0%
```

```
      Case "(2,3]"
        POST_PROB = 0.219  ' ICU_GR1 = (2,3]  21.9%
      Case "(3,4]"
        POST_PROB = 0.125  ' ICU_GR1 = (3,4]  12.5%
      Case "(4,5]"
        POST_PROB = 0.062  ' ICU_GR1 = (4,5]  6.2%
      Case ">5"
        POST_PROB = 0.344  ' ICU_GR1 = >5     34.4%
      Case "[0,1]"
        POST_PROB = 0      ' ICU_GR1 = [0,1]  0%

    End Select
  End If


' RULE_7
    If EMERG = "EMER" And RENAL_DIS = "Y" Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0      ' ICU_GR1 = (1,2]  0%
      Case "(2,3]"
        POST_PROB = 0.333      ' ICU_GR1 = (2,3]  33.3%
      Case "(3,4]"
        POST_PROB = 0      ' ICU_GR1 = (3,4]  0%
      Case "(4,5]"
        POST_PROB = 0.667      ' ICU_GR1 = (4,5]  66.7%
      Case ">5"
        POST_PROB = 0      ' ICU_GR1 = >5     0%
      Case "[0,1]"
        POST_PROB = 0          ' ICU_GR1 = [0,1]  0%

    End Select
  End If

 lg94_41 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)


End Function

'File lg94_42
   Function lg94_42(ICU_GR1, B1225, IABP, B1142, PROCED_GRO)

   Select Case ICU_GR1
      Case "(1,2]"
        PRIOR_PROB = 0.441
      Case "(2,3]"
        PRIOR_PROB = 0.143
      Case "(3,4]"
        PRIOR_PROB = 0.076
      Case "(4,5]"
        PRIOR_PROB = 0.024
      Case ">5"
        PRIOR_PROB = 0.055
      Case "[0,1]"
```

```
        PRIOR_PROB = 0.261
    End Select


'RULE_1
    If B1225 = "" And IABP = "" And B1142 = "" And _
    (PROCED_GRO = "TRANSPLANT" Or PROCED_GRO = "AORTIC SURG" Or PROCED_GRO = ""
Or PROCED_GRO = "CABG+VALVE+OTHER") Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0      ' ICU_GR1 = (1,2]  0%
      Case "(2,3]"
        POST_PROB = 0.143      ' ICU_GR1 = (2,3]  14.3%
      Case "(3,4]"
        POST_PROB = 0.143      ' ICU_GR1 = (3,4]  14.3%
      Case "(4,5]"
        POST_PROB = 0.286      ' ICU_GR1 = (4,5]  28.6%
      Case ">5"
        POST_PROB = 0.429      ' ICU_GR1 = >5     42.9%
      Case "[0,1]"
        POST_PROB = 0          ' ICU_GR1 = [0,1]  0%

    End Select
  End If



'RULE_2
    If B1225 = "" And IABP = "" And B1142 = "" And _
    (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO = "CABG+VALVE"
Or PROCED_GRO = "CABG+OTHER" Or PROCED_GRO = "VALVE+OTHER") Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.528      ' ICU_GR1 = (1,2]   52.8%
      Case "(2,3]"
        POST_PROB = 0.169      ' ICU_GR1 = (2,3]   16.9%
      Case "(3,4]"
        POST_PROB = 0.089      ' ICU_GR1 = (3,4]   8.9%
      Case "(4,5]"
        POST_PROB = 0.029      ' ICU_GR1 = (4,5]   2.9%
      Case ">5"
        POST_PROB = 0.026      ' ICU_GR1 = >5      2.6%
      Case "[0,1]"
        POST_PROB = 0.159      ' ICU_GR1 = [0,1]   15.9%

    End Select
  End If


'RULE_3
    If B1225 = "" And IABP = "" And B1142 = "" And _
    PROCED_GRO = "OTHER" Then
      Select Case ICU_GR1
```

```
    Case "(1,2]"
      POST_PROB = 0.294     ' ICU_GR1 = (1,2]   29.4%
    Case "(2,3]"
      POST_PROB = 0.471     ' ICU_GR1 = (2,3]   47.1%
    Case "(3,4]"
      POST_PROB = 0     ' ICU_GR1 = (3,4]   0%
    Case "(4,5]"
      POST_PROB = 0       ' ICU_GR1 = (4,5]   0%
    Case ">5"
      POST_PROB = 0     ' ICU_GR1 = >5      0%
    Case "[0,1]"
      POST_PROB = 0.235     ' ICU_GR1 = [0,1]   23.5%

  End Select
End If



'RULE_4
  If B1225 = "" And IABP = "" And B1142 = "Y" Then
    Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.4       ' ICU_GR1 = (1,2]   40.0%
    Case "(2,3]"
      POST_PROB = 0.233     ' ICU_GR1 = (2,3]   23.3%
    Case "(3,4]"
      POST_PROB = 0.067     ' ICU_GR1 = (3,4]   6.7%
    Case "(4,5]"
      POST_PROB = 0.1         ' ICU_GR1 = (4,5]   10.0%
    Case ">5"
      POST_PROB = 0.167     ' ICU_GR1 = >5      16.7%
    Case "[0,1]"
      POST_PROB = 0.033     ' ICU_GR1 = [0,1]   3.3%

  End Select
End If



'RULE_5
  If B1225 = "" And IABP = "IABP" Then
    Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.105     ' ICU_GR1 = (1,2]   10.5%
    Case "(2,3]"
      POST_PROB = 0.211     ' ICU_GR1 = (2,3]   21.1%
    Case "(3,4]"
      POST_PROB = 0.211     ' ICU_GR1 = (3,4]   21.1%
    Case "(4,5]"
      POST_PROB = 0.105       ' ICU_GR1 = (4,5]   10.5%
    Case ">5"
      POST_PROB = 0.368     ' ICU_GR1 = >5      36.8%
    Case "[0,1]"
```

```
        POST_PROB = 0          ' ICU_GR1 = [0,1]   0%

    End Select
  End If



' RULE_6
   If B1225 = "Y" Then
      Select Case ICU_GR1
      Case "(1,2]"
         POST_PROB = 0.182     ' ICU_GR1 = (1,2]    18.2%
      Case "(2,3]"
         POST_PROB = 0.152     ' ICU_GR1 = (2,3]    15.2%
      Case "(3,4]"
         POST_PROB = 0.152     ' ICU_GR1 = (3,4]    15.2%
      Case "(4,5]"
         POST_PROB = 0.197       ' ICU_GR1 = (4,5]    19.7%
      Case ">5"
         POST_PROB = 0.288     ' ICU_GR1 = >5       28.8%
      Case "[0,1]"
         POST_PROB = 0.03        ' ICU_GR1 = [0,1]    3.0%

      End Select
   End If

lg94_42 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function
```

```
'File sm94_22

Function SM94_22(ICU_GR4, AGE, IABP, PROCED_GRO, EMERG)


  Select Case ICU_GR4
    Case ">2"
      PRIOR_PROB = 0.274
    Case "[0,2]"
      PRIOR_PROB = 0.726
  End Select

'RULE_1
   If IABP = "" And (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
    AGE < 59 And (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO =
"CABG+OTHE" Or PROCED_GRO = "VALVE+OTHE") Then
     If ICU_GR4 = ">2" Then
        POST_PROB = 0.235  'ICU_GR4 = >2     23.5%
     Else
        POST_PROB = 0.765   'ICU_GR4 = [0,2]    76.5%
     End If
  End If



'RULE_2
   If IABP = "" And (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
   (AGE >= 59 And AGE < 76) And (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or
PROCED_GRO = "CABG+OTHE" Or PROCED_GRO = "VALVE+OTHE") Then
     If ICU_GR4 = ">2" Then
        POST_PROB = 0.322  'ICU_GR4 = >2     32.2%
     Else
        POST_PROB = 0.678   'ICU_GR4 = [0,2]   67.8%
     End If
  End If


'RULE_3
  If IABP = "" And (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
    AGE >= 76 And (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO =
"CABG+OTHE" Or PROCED_GRO = "VALVE+OTHE") Then
   If ICU_GR4 = ">2" Then
        POST_PROB = 0.456  'ICU_GR4 = >2     45.6%
     Else
        POST_PROB = 0.544   'ICU_GR4 = [0,2]   54.4%

     End If
  End If


'RULE_4
   If IABP = "" And (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
```

```
    (PROCED_GRO = "CABG+VALV" Or PROCED_GRO = "OTHER" Or PROCED_GRO =
"AORTIC SU") Then
   If ICU_GR4 = ">2" Then
      POST_PROB = 0.486  'ICU_GR4 = >2    48.6%
     Else
      POST_PROB = 0.514   'ICU_GR4 = [0,2]   51.4%

     End If
  End If



'RULE_5
   If IABP = "" And (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
     PROCED_GRO = "TRANSPLAN" Then
     If ICU_GR4 = ">2" Then
       POST_PROB = 1   'ICU_GR4 = >2     100.0%
     Else
       POST_PROB = 0  'ICU_GR4 = [0,2]   0%

     End If
   End If



'RULE_6
   If IABP = "" And EMERG = "EMER" Then
     If ICU_GR4 = ">2" Then
       POST_PROB = 0.658   'ICU_GR4 = >2     65.8%
     Else
       POST_PROB = 0.342  'ICU_GR4 = [0,2]  34.2%

     End If
   End If



'RULE_7
  If IABP = "IABP" Then
    If ICU_GR4 = ">2" Then
       POST_PROB = 0.912   'ICU_GR4 = >2    91.2%
     Else
       POST_PROB = 0.088  'ICU_GR4 = [0,2]  8.8%

     End If
   End If

SM94_22 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function


'File sm94_31
```

Visual Basic program for calculation information based score for the SM94_** models.

```
Function Entropy(ParamArray PC())
  For Each X In PC
    LOG2 = Log(X) / Log(2)
    A = X * LOG2
    S = S + A
  Next X
  Entropy = -S
End Function

Function CALC_INF_SCORE(PRIOR_PROB, POST_PROB)
  If POST_PROB >= PRIOR_PROB Then
    CALC_INF_SCORE = -(Log(PRIOR_PROB) / Log(2)) + (Log(POST_PROB) / Log(2))
  Else
    CALC_INF_SCORE = -(-(Log(1 - PRIOR_PROB) / Log(2)) + (Log(1 - POST_PROB) / Log(2)))
  End If
End Function


'INFOM SCORE FOR SM94_21

Function SM94_21(ICU_GR4, AGE, EMERG, REDO, GENDER)


  Select Case ICU_GR4
    Case ">2"
      PRIOR_PROB = 0.274
    Case "[0,2]"
      PRIOR_PROB = 0.726
  End Select


'RULE_1
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And AGE < 59 Then
    If ICU_GR4 = ">2" Then
      POST_PROB = 0.266  'ICU_GR4 = >2    26.6%
    Else
      POST_PROB = 0.734  'ICU_GR4 = [0,2]  73.4%
    End If
  End If


'RULE_2
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (AGE >= 59 And AGE < 76) And
REDO = "" Then
    If ICU_GR4 = ">2" Then
      POST_PROB = 0.343  'ICU_GR4 = >2    34.3%
    Else
      POST_PROB = 0.657  'ICU_GR4 = [0,2]  65.7%
```

```
   (AGE >= 59 And AGE < 73) And CHF = "" Then
        Select Case ICU_GR3
      Case ">2"
        POST_PROB = 0.351   ' ICU_GR3 = >2     35.1%
      Case "[0,2]"
        POST_PROB = 0.649   ' ICU_GR3 = [0,2]  64.9%

   End Select
End If


' RULE_3
   If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (NYHA_CLASS = "" Or
NYHA_CLASS = "2" Or NYHA_CLASS = "1") And _
   (AGE >= 59 And AGE < 73) And CHF = "Y" Then
        Select Case ICU_GR3
      Case ">2"
        POST_PROB = 0.909   ' ICU_GR3 = >2     90.9%
      Case "[0,2]"
        POST_PROB = 0.091   ' ICU_GR3 = [0,2]  9.1%

   End Select
End If



'RULE_4 IF
   If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (NYHA_CLASS = "" Or
NYHA_CLASS = "2" Or NYHA_CLASS = "1") And _
   AGE >= 73 Then
        Select Case ICU_GR3
      Case ">2"
        POST_PROB = 0.479   ' ICU_GR3 = >2     47.9%
      Case "[0,2]"
        POST_PROB = 0.521   ' ICU_GR3 = [0,2]  52.1%

   End Select
End If



' RULE_5
    If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (NYHA_CLASS = "3" Or
NYHA_CLASS = "4") Then
        Select Case ICU_GR3
      Case ">2"
        POST_PROB = 0.711   ' ICU_GR3 = >2     71.1%
      Case "[0,2]"
        POST_PROB = 0.289   ' ICU_GR3 = [0,2]  28.9%

   End Select
End If
```

```
' RULE_6
  If EMERG = "EMER" Then
     Select Case ICU_GR3
     Case ">2"
        POST_PROB = 0.771   ' ICU_GR3 = >2      77.1%
     Case "[0,2]"
        POST_PROB = 0.229   ' ICU_GR3 = [0,2]   22.9%

  End Select
End If

LG94_21 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function

'File lg94_22
  Function lg94_22(ICU_GR3, CHF, B1225, IABP, B1142, PROCED_GRO)

     Select Case ICU_GR3
     Case ">2"
        PRIOR_PROB = 0.299
     Case "[0,2]"
        PRIOR_PROB = 0.701
  End Select

' RULE_1
  If B1225 = "" And IABP = "" And B1142 = "" And _
  (PROCED_GRO = "CABG" Or PROCED_GRO = "CABG+OTHER" Or PROCED_GRO =
"VALVE+OTHER" Or PROCED_GRO = "CABG+VALVE+OTHER") Then
    Select Case ICU_GR3
     Case ">2"
        POST_PROB = 0.295   ' ICU_GR3 = >2      29.5%
     Case "[0,2]"
        POST_PROB = 0.705   ' ICU_GR3 = [0,2]   70.5%

  End Select
End If



' RULE_2
  If B1225 = "" And IABP = "" And B1142 = "" And _
  (PROCED_GRO = "VALVE" Or PROCED_GRO = "CABG+VALVE" Or PROCED_GRO =
"OTHER") Then
      Select Case ICU_GR3
     Case ">2"
        POST_PROB = 0.409   ' ICU_GR3 = >2      40.9%
     Case "[0,2]"
        POST_PROB = 0.591   ' ICU_GR3 = [0,2]   59.1%
  End Select
End If


' RULE_3
```

```
    End If
  End If


'RULE_3
   If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And (AGE >= 59 And AGE < 76)
And REDO = "REDO" Then
      If ICU_GR4 = ">2" Then
        POST_PROB = 0.478  'ICU_GR4 = >2     47.8%
      Else
        POST_PROB = 0.522  'ICU_GR4 = [0,2]  52.2%
      End If
  End If


'RULE_4
   If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And AGE >= 76 And GENDER = "M"
Then
      If ICU_GR4 = ">2" Then
        POST_PROB = 0.459  'ICU_GR4 = >2     45.9%
      Else
        POST_PROB = 0.541  'ICU_GR4 = [0,2]  54.1%
      End If
  End If


'RULE_5
   If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And AGE >= 76 And _
    GENDER = "F" Then
      If ICU_GR4 = ">2" Then
        POST_PROB = 0.677  'ICU_GR4 = >2     67.7%
      Else
        POST_PROB = 0.323  'ICU_GR4 = [0,2]   32.3%
      End If
  End If


'RULE_6
   If EMERG = "EMER" Then
      If ICU_GR4 = ">2" Then
        POST_PROB = 0.755  'ICU_GR4 = >2     75.5%
      Else
        POST_PROB = 0.245  'ICU_GR4 = [0,2]   24.5%
      End If
  End If

SM94_21 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function
```

```
Function SM94_31(ICU_GR2, AGE, EMERG, REDO)

Select Case ICU_GR2
   Case "(2,4]"
      PRIOR_PROB = 0.189
   Case ">4"
      PRIOR_PROB = 0.085
   Case "[0,2]"
      PRIOR_PROB = 0.726
End Select

'RULE_1
 If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And AGE < 63 Then
    Select Case ICU_GR2
       Case "(2,4]"
          POST_PROB = 0.232   'ICU_GR2 = (2,4]   23.2%
       Case ">4"
          POST_PROB = 0.052   'ICU_GR2 = >4      5.2%
       Case "[0,2]"
          POST_PROB = 0.717   'ICU_GR2 = [0,2]   71.7%
    End Select
 End If


'RULE_2
 If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
 (AGE >= 63 And AGE < 76) And REDO = "" Then
    Select Case ICU_GR2
     Case "(2,4]"
      POST_PROB = 0.241         'ICU_GR2 = (2,4]   24.1%
     Case ">4"
      POST_PROB = 0.107         'ICU_GR2 = >4      10.7%
     Case "[0,2]"
      POST_PROB = 0.652         'ICU_GR2 = [0,2]   65.2%
    End Select
 End If




'RULE_3
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
  (AGE >= 63 And AGE < 76) And REDO = "REDO" Then
    Select Case ICU_GR2
     Case "(2,4]"
      POST_PROB = 0.315         'ICU_GR2 = (2,4]  31.5%
     Case ">4"
      POST_PROB = 0.192         'ICU_GR2 = >4      19.2%
     Case "[0,2]"
      POST_PROB = 0.493         'ICU_GR2 = [0,2]   49.3%
    End Select
End If
```

```
'RULE_4
  If (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And AGE >= 76 Then
    Select Case ICU_GR2
     Case "(2,4]"
      POST_PROB = 0.326          'ICU_GR2 = (2,4]  32.6%
     Case ">4"
      POST_PROB = 0.207          'ICU_GR2 = >4     20.7%
     Case "[0,2]"
      POST_PROB = 0.467          'ICU_GR2 = [0,2]  46.7%
    End Select
  End If



'RULE_5
  If EMERG = "EMER" Then
    Select Case ICU_GR2
     Case "(2,4]"
     POST_PROB = 0.358           'ICU_GR2 = (2,4]  35.8%
    Case ">4"
     POST_PROB = 0.396           'ICU_GR2 = >4     39.6%
    Case "[0,2]"
     POST_PROB = 0.245           'ICU_GR2 = [0,2]  24.5%
    End Select
  End If

SM94_31 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function

'File sm94_32


Function SM94_32(ICU_GR2, IABP, PROCED_GRO, EMERG)

  Select Case ICU_GR2
    Case "(2,4]"
     PRIOR_PROB = 0.189
    Case ">4"
     PRIOR_PROB = 0.085
    Case "[0,2]"
     PRIOR_PROB = 0.726
  End Select


'RULE_1
  If IABP = "" And PROCED_GRO = "AORTIC SU" Then
   Select Case ICU_GR2
    Case "(2,4]"
     POST_PROB = 0           'ICU_GR2 = (2,4]  0%
    Case ">4"
     POST_PROB = 0.75        'ICU_GR2 = >4   75.0%
```

```
    Case "[0,2]"
     POST_PROB = 0.25          'ICU_GR2 = [0,2]   25.0%
    End Select

End If


'RULE_2
  If IABP = "" And (EMERG = "ELEC" Or EMERG = "URG" Or EMERG = "") And _
     (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO =
"VALVE+OTH") Then

  Select Case ICU_GR2
     Case "(2,4]"
      POST_PROB = 0.234          'ICU_GR2 = (2,4]   23.4%
     Case ">4"
      POST_PROB = 0.065          'ICU_GR2 = >4       6.5%
     Case "[0,2]"
      POST_PROB = 0.701          'ICU_GR2 = [0,2]   70.1%

  End Select
End If


'RULE_3
 If IABP = "" And (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO =
"VALVE+OTH") And _
     EMERG = "EMER" Then
   Select Case ICU_GR2
      Case "(2,4]"
       POST_PROB = 0.296          'ICU_GR2 = (2,4]   29.6%
      Case ">4"
       POST_PROB = 0.296          'ICU_GR2 = >4       29.6%
     Case "[0,2]"
       POST_PROB = 0.407          'ICU_GR2 = [0,2]   40.7%

  End Select
 End If


'RULE_4
  If IABP = "" And PROCED_GRO = "CABG+OTHE" Then
   Select Case ICU_GR2
     Case "(2,4]"
      POST_PROB = 0.1            'ICU_GR2 = (2,4]   10.0%
     Case ">4"
      POST_PROB = 0.25           'ICU_GR2 = >4       25.0%
     Case "[0,2]"
      POST_PROB = 0.65           'ICU_GR2 = [0,2]   65.0%
  End Select
 End If
```

```
'RULE_5
  If IABP = "" And (PROCED_GRO = "CABG+VALV" Or PROCED_GRO = "OTHER") Then
    Select Case ICU_GR2
    Case "(2,4]"
     POST_PROB = 0.358          'ICU_GR2 = (2,4]   35.8%
    Case ">4"
     POST_PROB = 0.132          'ICU_GR2 = >4      13.2%
    Case "[0,2]"
     POST_PROB = 0.509          'ICU_GR2 = [0,2]   50.9%
  End Select
End If


'RULE_6
   If IABP = "" And PROCED_GRO = "TRANSPLAN" Then
     Select Case ICU_GR2
       Case "(2,4]"
         POST_PROB = 0.538          'ICU_GR2 = (2,4]   53.8%
       Case ">4"
         POST_PROB = 0.385          'ICU_GR2 = >4      38.5%
       Case "[0,2]"
         POST_PROB = 0.077          'ICU_GR2 = [0,2]   7.7%
      End Select
End If

'RULE_7
   If IABP = "IABP" Then
    Select Case ICU_GR2
    Case "(2,4]"
      POST_PROB = 0.368          'ICU_GR2 = (2,4]   36.8%
    Case ">4"
     POST_PROB = 0.544          'ICU_GR2 = >4      54.4%
    Case "[0,2]"
     POST_PROB = 0.088          'ICU_GR2 = [0,2]   8.8%
  End Select
End If

SM94_32 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function

'File sm94_41

  Function SM94_41(ICU_GR1, AGE, PROCED_GRO, EMERG, REDO, NYHA_CLASS)
   Select Case ICU_GR1
     Case "(1,2]"
      PRIOR_PROB = 0.44
     Case "(2,3]"
       PRIOR_PROB = 0.123
     Case "(3,4]"
       PRIOR_PROB = 0.066
     Case "(4,5]"
       PRIOR_PROB = 0.024
     Case ">5"
```

```
        PRIOR_PROB = 0.061
      Case "[0,1]"
        PRIOR_PROB = 0.286
    End Select

'RULE_1
  If (EMERG = "URG" Or EMERG = "") And _
  (PROCED_GRO = "CABG+VALV" Or PROCED_GRO = "CABG+OTHE" Or PROCED_GRO =
"TRANSPLAN" Or PROCED_GRO = "AORTIC SU" Or PROCED_GRO = "VALVE+OTH" Or
PROCED_GRO = "OTHER") Then


    Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.238  'ICU_GR1 = (1,2]  23.8%
      Case "(2,3]"
        POST_PROB = 0.143  'ICU_GR1 = (2,3]  14.3%
      Case "(3,4]"
        POST_PROB = 0.167  ' ICU_GR1 = (3,4]  16.7%
      Case "(4,5]"
        POST_PROB = 0.071  ' ICU_GR1 = (4,5]  7.1%
      Case ">5"
        POST_PROB = 0.238  ' ICU_GR1 = >5     23.8%
      Case "[0,1]"
        POST_PROB = 0.143  ' ICU_GR1 = [0,1]  14.3%
    End Select

End If


'RULE_2
    If (EMERG = "URG" Or EMERG = "") And _
    (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE") Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.45   'ICU_GR1 = (1,2]  45.0%
      Case "(2,3]"
        POST_PROB = 0.148  'ICU_GR1 = (2,3]   14.8%
      Case "(3,4]"
        POST_PROB = 0.094  ' ICU_GR1 = (3,4]  9.4%
      Case "(4,5]"
        POST_PROB = 0.051  ' ICU_GR1 = (4,5]   5.1%
      Case ">5"
        POST_PROB = 0.048  ' ICU_GR1 = >5     4.8%
      Case "[0,1]"
        POST_PROB = 0.208  ' ICU_GR1 = [0,1]  20.8%
    End Select

End If


'RULE_3
    If EMERG = "ELEC" And REDO = "" And AGE < 63 And (NYHA_CLASS = "" Or NYHA_CLASS =
"1") Then
```

```
 Select Case ICU_GR1
     Case "(1,2]"
       POST_PROB = 0.545   'ICU_GR1 = (1,2]   54.5%
     Case "(2,3]"
       POST_PROB = 0.168   'ICU_GR1 = (2,3]   16.8%
     Case "(3,4]"
       POST_PROB = 0.058   ' ICU_GR1 = (3,4]   5.8%
     Case "(4,5]"
       POST_PROB = 0.007   ' ICU_GR1 = (4,5]    0.7%
     Case ">5"
       POST_PROB = 0.01    ' ICU_GR1 = >5      1.0%
     Case "[0,1]"
       POST_PROB = 0.212   ' ICU_GR1 = [0,1]   21.2%
   End Select

End If



'RULE_4
   If EMERG = "ELEC" And REDO = " " And AGE < 63 And (NYHA_CLASS = "2" Or
NYHA_CLASS = "3" Or NYHA_CLASS = "4") Then
      Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.368   'ICU_GR1 = (1,2]   36.8%
      Case "(2,3]"
        POST_PROB = 0.211   'ICU_GR1 = (2,3]   21.1%
      Case "(3,4]"
        POST_PROB = 0.263   ' ICU_GR1 = (3,4]   26.3%
      Case "(4,5]"
        POST_PROB = 0.053   ' ICU_GR1 = (4,5]    5.3%
      Case ">5"
        POST_PROB = 0.053   ' ICU_GR1 = >5      5.3%
      Case "[0,1]"
        POST_PROB = 0.053   ' ICU_GR1 = [0,1]    5.3%
   End Select

End If


'RULE_5
   If EMERG = "ELEC" And REDO = "" And (AGE >= 63 And AGE < 76) And _
   (PROCED_GRO = "CABG" Or PROCED_GRO = "VALVE" Or PROCED_GRO = "VALVE+OTH"
Or PROCED_GRO = "OTHER" Or PROCED_GRO = "TRANSPLAN" Or PROCED_GRO = "AORTIC
SU") Then

   Select Case ICU_GR1
      Case "(1,2]"
        POST_PROB = 0.585   'ICU_GR1 = (1,2]   58.5%
      Case "(2,3]"
        POST_PROB = 0.157   'ICU_GR1 = (2,3]   15.7%
      Case "(3,4]"
        POST_PROB = 0.093   ' ICU_GR1 = (3,4]   9.3%
```

```
    Case "(4,5]"
      POST_PROB = 0.03   ' ICU_GR1 = (4,5]   3.0%
    Case ">5"
      POST_PROB = 0.025  ' ICU_GR1 = >5      2.5%
    Case "[0,1]"
      POST_PROB = 0.11   ' ICU_GR1 = [0,1]   11.0%
  End Select

End If



'RULE_6
   If EMERG = "ELEC" And REDO = " " And (AGE >= 63 And AGE < 76) And _
  (PROCED_GRO = "CABG+VALV" Or PROCED_GRO = "COBG+OTHE") Then

  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.361   'ICU_GR1 = (1,2]  36.1%
    Case "(2,3]"
      POST_PROB = 0.083   'ICU_GR1 = (2,3]  8.3%
    Case "(3,4]"
      POST_PROB = 0.167   ' ICU_GR1 = (3,4]  16.7%
    Case "(4,5]"
      POST_PROB = 0.139   ' ICU_GR1 = (4,5]  13.9%
    Case ">5"
      POST_PROB = 0.194   ' ICU_GR1 = >5     19.4%
    Case "[0,1]"
      POST_PROB = 0.056   ' ICU_GR1 = [0,1]  5.6%
  End Select

End If


'RULE_7
   If EMERG = "ELEC" And REDO = "" And AGE >= 76 Then
  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.347   'ICU_GR1 = (1,2]  34.7%
    Case "(2,3]"
      POST_PROB = 0.265   'ICU_GR1 = (2,3]  26.5%
    Case "(3,4]"
      POST_PROB = 0.102   ' ICU_GR1 = (3,4]  10.2%
    Case "(4,5]"
      POST_PROB = 0.082   ' ICU_GR1 = (4,5]  8.2%
    Case ">5"
      POST_PROB = 0.122   ' ICU_GR1 = >5     12.2%
    Case "[0,1]"
      POST_PROB = 0.082   ' ICU_GR1 = [0,1]  8.2%
  End Select

End If
```

```
'RULE_8
  If EMERG = "ELEC" And REDO = " REDO" Then

  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.554   'ICU_GR1 = (1,2]   55.4%
    Case "(2,3]"
      POST_PROB = 0.12    'ICU_GR1 = (2,3]   12.0%
    Case "(3,4]"
      POST_PROB = 0.087   ' ICU_GR1 = (3,4]   8.7%
    Case "(4,5]"
      POST_PROB = 0.076  ' ICU_GR1 = (4,5]   7.6%
    Case ">5"
      POST_PROB = 0.109   ' ICU_GR1 = >5      10.9%
    Case "[0,1]"
      POST_PROB = 0.054    ' ICU_GR1 = [0,1]   5.4%
  End Select

End If



'RULE_9
  If EMERG = "EMER" Then
  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.226   'ICU_GR1 = (1,2]   22.6%
    Case "(2,3]"
      POST_PROB = 0.226   'ICU_GR1 = (2,3]   22.6%
    Case "(3,4]"
      POST_PROB = 0.132   ' ICU_GR1 = (3,4]   13.2%
    Case "(4,5]"
      POST_PROB = 0.094  ' ICU_GR1 = (4,5]   9.4%
    Case ">5"
      POST_PROB = 0.302   ' ICU_GR1 = >5      30.2%
    Case "[0,1]"
      POST_PROB = 0.019    ' ICU_GR1 = [0,1]   1.9%
  End Select

End If

SM94_41 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function


'File sm94_42

Function SM94_42(ICU_GR1, AGE, IABP, PROCED_GRO, EMERG, REDO, GENDER, PUMPTIME)

  Select Case ICU_GR1
    Case "(1,2]"
```

```
      PRIOR_PROB = 0.44
    Case "(2,3]"
      PRIOR_PROB = 0.123
    Case "(3,4]"
      PRIOR_PROB = 0.066
    Case "(4,5]"
      PRIOR_PROB = 0.024
    Case ">5"
      PRIOR_PROB = 0.061
    Case "[0,1]"
      PRIOR_PROB = 0.286
  End Select


'RULE_1
  If IABP = "" And PUMPTIME < 116 And (EMERG = "URG" Or EMERG = " ") Then
  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.462   'ICU_GR1 = (1,2]   46.2%
    Case "(2,3]"
      POST_PROB = 0.155   'ICU_GR1 = (2,3]  15.5%
    Case "(3,4]"
      POST_PROB = 0.076  ' ICU_GR1 = (3,4]  7.6%
    Case "(4,5]"
      POST_PROB = 0.034  ' ICU_GR1 = (4,5]    3.4%
    Case ">5"
      POST_PROB = 0.038  ' ICU_GR1 = >5      3.8%
    Case "[0,1]"
      POST_PROB = 0.235   ' ICU_GR1 = [0,1]   23.5%
  End Select

End If


'RULE_2
  If IABP = "" And PUMPTIME < 116 And EMERG = "URG" And AGE < 49 Then
  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.469   'ICU_GR1 = (1,2]   46.9%
    Case "(2,3]"
      POST_PROB = 0.172   'ICU_GR1 = (2,3]  17.2%
    Case "(3,4]"
      POST_PROB = 0.062  ' ICU_GR1 = (3,4]  6.2%
    Case "(4,5]"
      POST_PROB = 0  ' ICU_GR1 = (4,5]    0%
    Case ">5"
      POST_PROB = 0   ' ICU_GR1 = >5      0%
    Case "[0,1]"
      POST_PROB = 0.297   ' ICU_GR1 = [0,1]  29.7%
  End Select

End If
```

```
'RULE_3
   If IABP = "" And PUMPTIME < 116 And EMERG = "ELEC" And (AGE >= 49 And AGE < 76)
Then

  Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.592    'ICU_GR1 = (1,2]    59.2%
    Case "(2,3]"
      POST_PROB = 0.163    'ICU_GR1 = (2,3]  16.3%
    Case "(3,4]"
      POST_PROB = 0.068    ' ICU_GR1 = (3,4]  6.8%
    Case "(4,5]"
      POST_PROB = 0.016    ' ICU_GR1 = (4,5]   1.6%
    Case ">5"
      POST_PROB = 0.021    ' ICU_GR1 = >5      2.1%
    Case "[0,1]"
      POST_PROB = 0.14     ' ICU_GR1 = [0,1]  14.0%
  End Select

End If


'RULE_4
    If IABP = "" And PUMPTIME < 116 And EMERG = "ELEC" And (AGE >= 76 And AGE < 88)
Then

    Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.344    'ICU_GR1 = (1,2]    34.4%
    Case "(2,3]"
      POST_PROB = 0.219    'ICU_GR1 = (2,3]  21.9%
    Case "(3,4]"
      POST_PROB = 0.156    ' ICU_GR1 = (3,4]  15.6%
    Case "(4,5]"
      POST_PROB = 0.094    ' ICU_GR1 = (4,5]   9.4%
    Case ">5"
      POST_PROB = 0.094    ' ICU_GR1 = >5      9.4%
    Case "[0,1]"
      POST_PROB = 0.094    ' ICU_GR1 = [0,1]  9.4%
  End Select

End If



'RULE_5
   If IABP = " " And PUMPTIME < 116 And EMERG = "EMER" Then
    Select Case ICU_GR1
     Case "(1,2]"
       POST_PROB = 0.321    'ICU_GR1 = (1,2]    32.1%
     Case "(2,3]"
       POST_PROB = 0.286    'ICU_GR1 = (2,3]    28.6%
```

```
    Case "(3,4]"
      POST_PROB = 0.107    ' ICU_GR1 = (3,4]   10.7%
    Case "(4,5]"
      POST_PROB = 0.107    ' ICU_GR1 = (4,5]   10.7%
    Case ">5"
      POST_PROB = 0.143    ' ICU_GR1 = >5      14.3%
    Case "[0,1]"
      POST_PROB = 0.036    ' ICU_GR1 = [0,1]   3.6%
  End Select

End If


'RULE_6
  If IABP = "" And PUMPTIME >= 116 Then
 Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.462    'ICU_GR1 = (1,2]    46.2%
    Case "(2,3]"
      POST_PROB = 0.119    'ICU_GR1 = (2,3]    11.9%
    Case "(3,4]"
      POST_PROB = 0.142    ' ICU_GR1 = (3,4]   14.2%
    Case "(4,5]"
      POST_PROB = 0.085    ' ICU_GR1 = (4,5]    8.5%
    Case ">5"
      POST_PROB = 0.085    ' ICU_GR1 = >5       8.5%
    Case "[0,1]"
      POST_PROB = 0.104    ' ICU_GR1 = [0,1]   10.4%
  End Select

End If


'RULE_7
  If IABP = "IABP" Then
Select Case ICU_GR1
    Case "(1,2]"
      POST_PROB = 0.035    'ICU_GR1 = (1,2]    3.5%
    Case "(2,3]"
      POST_PROB = 0.211    'ICU_GR1 = (2,3]  21.1%
    Case "(3,4]"
      POST_PROB = 0.158    ' ICU_GR1 = (3,4]   15.8%
    Case "(4,5]"
      POST_PROB = 0.123    ' ICU_GR1 = (4,5]   12.3%
    Case ">5"
      POST_PROB = 0.421    ' ICU_GR1 = >5      42.1%
    Case "[0,1]"
      POST_PROB = 0.053    ' ICU_GR1 = [0,1]   5.3%

  End Select

End If
```

SM94_42 = CALC_INF_SCORE(PRIOR_PROB, POST_PROB)

End Function