University of Zurich
Department of Informatics

Evaluation of novel algorithms to optimize Risk Stratification Scores in Myocardial Infarction

Diploma Thesis
January 11, 2006

Katerina Hunt
of Rüti, ZH, Switzerland

Student-ID: 00-707-000
katrin@hunt.ch

Prof. Abraham Bernstein, PhD
Department of Informatics
University of Zurich
http://www.ifi.unizh.ch/ddis

Acknowledgements

This diploma thesis could not have been accomplished without the input and help of Dr. David Kurz. Thanks for your time and dedication.
The advice given by Peter Vorburger and Jiwen Li enhanced the quality of the thesis considerably. Thanks a lot for your time.
Last but not least I would like to thank Prof. Bernstein for the insightful discussions and the open door to his office at all times.

Abstract

Risk Predictors currently used in the field of Acute Myocardial Infarction (AMI) were developed on data cohorts collected in the early 90's using traditional statistical methods. Considering the progress in the therapy of AMI as well as in the field of Data Mining, it was hypothesized that a better Risk Predictor could be developed. Working on the AMIS PLUS registry (n=7520) existing scores were evaluated and a new Risk Prediction Model developed, using the AODE algorithm from the Bayes family. The most accepted Risk Score (TIMI Risk Score for ST-Elevation) yielded an Area under the ROC Curve (AUC) of 0.803. The newly developed Risk Model called AMIS Model achieved an AUC of 0.875 using less input variables. Tests showed that the prediction capacity of the AMIS Model was especially good with patients undergoing PCI treatment (AUC=0.885 compared to AUC=0.783 of TIMI Risk Score).

Zusammenfassung

Gegenwärtig benutzte Risiko Prädiktoren im Umfeld des Myokardinfarkts wurden auf Daten entwickelt, welche in den frühen 90er Jahren erhoben worden sind. Zur Entwicklung der Prädiktionsmodelle wurden traditionelle statistische Methoden eingesetzt. In Anbetracht der Entwicklungen im Bereich des Data Mining, sowie auch den weiter entwickelten Therapiemöglichkeiten in den letzten Jahren wurde angenommen, dass es möglich ist ein besseres Risikoprädiktionsmodell zu entwickeln. Existierende Prädiktionsmodelle wurden auf den AMIS Plus Daten (n=7520) evaluiert und ein neues Prädiktionsmodell, basierend auf dem AODE Algorithmus der Bayes Familie, entwickelt. Das von der medizinischen Fachwelt am besten akzeptierte Prädiktionsmodell (TIMI Risk Score for ST-Elevation) erreichte eine Area under the ROC Curve (AUC) von 0.803. Das neu entwickelte Prädiktionsmodell namens AMIS Model erreichte auf denselben Daten eine AUC von 0.875, obwohl es weniger Angaben zur Berechnung benötigt. Weitere Tests zeigten, dass die Prädiktionskapazität des AMIS Model am grössten bei den Patienten ist, welche mit PCI therapiert werden (AUC=0.885 gegenüber AUC=0.783 des TIMI Risk Scores).

Table of Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

Since 1997 the AMIS Plus National Registry has collected and analyzed data of patients suffering an Acute Myocardial Infarction [1] (AMI) in Switzerland. In the past 12 years 18'000 records of Acute Coronary Syndrome (ACS) patients have been assembled. The records are collected from over 50 hospitals all over Switzerland[2]. These records have been analyzed with traditional statistical methods regularly. An analysis with data mining methods has been attempted once, but not on a professional level.

The steering committee of AMIS Plus hoped that, given enough expertise was invested, more information could be extracted from the data cohort using data mining methods. The development of a mortality risk prediction model was the specified aim. Under these circumstances the Department of Informatics was contacted.

1.2 The Assignment

The first step of the assignment was to acquire all skills, techniques and knowledge required to meet with the above described scenario. The problem being of an interdisciplinary kind not only information concerning the technical part had to be read up. More importantly the field of application, in this case the field of AMI, had to be gleaned.

[1] Acute Myocardial Infarction (AMI or MI) is also known as heart attack. It is caused by the death of heart muscle due to a obstruction of one or more coronary arteries.

[2] An up to date list of all participating hospitals is available on the AMIS Plus website (http://www.amis-plus.ch/participants.htm).

The second step of the assignment was the preprocessing of the data cohort. Part of this consisted of the study and understanding of the variables of the AMIS Plus dataset and their medical significance. Another part consisted of the cleaning of the database, i.e. the identification and removal of faulty and therefore disturbing records or variables. In a third step the data had to be prepared in the best way possible for the application of the induction algorithms.

As a third step existing risk stratification scores were to be applied and validated on the AMIS Plus dataset. This step bore three goals. Firstly the data could be analyzed by the application of existing risk scores. Secondly the risk scores themselves could be analyzed and tested for weaknesses and strengths. Thirdly a benchmark could be established, serving to grade any newly developed risk predictors.

The application of the data mining algorithms made up the fourth step. Using selected algorithms a prediction model should be generated whose output is the probability of death. The input variables must be carefully selected and discussed with an expert in the field of application to prevent incorrect interpretations and the usage of hidden dependencies.

The fifth step contained the presentation of the results to all stakeholders involved in the project, namely the steering committee of AMIS Plus, the AMIS Data centre [3], and the involved medical staff. Part of this presentation is a prototype that allows the medical professionals to test the created prediction model.

If the outcome permits, the newly developed prediction model could be used as a support when diagnosing and as a benchmark for the hospitals.

1.3 Previous Work

In the field of AMI there is a number of publications on mortality risk prediction models. The most relevant to our work are the works of the Gusto Investigators[4] (Lee et al., 1995) with their multivariate model for mortality prediction, and the TIMI investigators with the TIMI Risk Score for ST-Elevation (Morrow et al., 2000), the TIMI Risk Score for Unstable Angina / Nont-ST Elevation (Antman et al., 2000) and the Simple Risk Score (Morrow et al., 2001).

[3] See section 2.1

[4] In the medical field the risk models are usually named after the database they were developed on. Accordingly the research group is usually named after the cohort they worked on.

1.4 Outcome

After performing the necessary steps of data cleaning and data preparation, the data available in the AMIS Plus registry is of sufficient quality and quantity to yield useful data mining results.
In the course of this diploma work the AMIS Model was developed, prediction in hospital mortality after an AMI. The AMIS Model is based on an AODE algorithm, uses 7 input variables and outputs the probability of death. All 7 variables are available at presentation[5], independent of the course of treatment, and can be measured unambiguously.
In chapter 5 it is shown that the AMIS Model is superior in many regards to the existing risk prediction scores and inferior in none tested.
The AMIS Model has already been presented to the AMIS Plus steering committee and was received well.

1.5 The Structure of this Diploma Thesis

In chapter 2 the characteristics of the AMIS Plus data registry are outlined and all undertaken preprocessing steps are described.
Chapter 3 contains the evaluation of the existing risk scores, describing the methods of evaluation as well as the results and possible interpretations.
The application of the data mining algorithms is described in chapter 4. Training sets, choice of variables and algorithms are compared and explained for the endpoints death and *mace*[6]. At the end of chapter 4 all decisions leading to the AMIS Model are summed up.
Chapter 5 contains the evaluation of the AMIS Model. Comparisons of the AMIS Model to the TIMI Risk Score are shown as well as treatment subgroup analyses and probability distributions.
Chapter 6 contains a summary of the results achieved with this diploma thesis and chapter 7 lists possible interests of future work.

All medical terms and acronyms can be looked up in the glossary in appendix F.

[5]For the medical professionals it is very valuable if all information needed to calculate the risk score is available immediately at first presentation of the patient.

[6]Mace is the combined endpoint of death and major adverse cardiac events. For further information consult the description in chapter 4.

2

The data and necessary steps of Preprocessing

2.1 The AMIS Plus Registry

The AMIS Plus (Acute Myocardial Infarction in Switzerland) project was founded in 1997 by a steering committee representing the Swiss Societies of Cardiology, Intensive care and Internal Medicine. The AMIS Plus has two main purposes. The maintenance of a nation wide registry to enable the description of the patient population as well as the characteristics of treatment and hospitalization and the carrying out of epidemiological studies. The second purpose is the quality control of the medical care in Switzerland.
As mentioned in the introduction, the AMIS Plus Registry is filled with data collected from over 50 Swiss hospitals. Any person suffering an Acute Myocardial Infarction[1] in one of the collaborating hospitals is entered into the database if permitted by the patient[2].

Data Entry The entering of the data is done by the hospitals. The data is entered into the AMIS Plus questionnaire (shown in B.1). The Questionnaire can either be filled in on paper, or on the Internet. The questionnaire is updated frequently. The AMIS Plus team issues guidelines instructing the personnel on the exact input information wanted to the questionnaire. See Appendix B.2 for the current AMIS Plus guidelines.
It is known to the AMIS Plus team that not all hospitals carry out the job of filling in the questionnaire with the same care and zest. This results in a notable variability of the data quality. For data protection reasons that dataset received did not contain any information enabling the identification of the hospitals. Hence it was not possible to include the origin of the data into the thinking process when carrying out the data cleaning.

[1] since 01.01.2005 year also patients suffering from Unstable Angina (UA)

[2] Each patient is handed an information sheet which also asks for the patients consenting signature. See Appendix B.3 for the AMIS Plus Patient Information and Consent Form

Data Analysis The AMIS Plus group employs five people in a Data Center in Zurich. This Data Center is responsible for the data care, the updating of the questionnaire, as well as the analysis of the data. The Data Center is headed by Dragana Radovanovic, MD. The Data Center is very productive, generating numerous abstracts and papers yearly. Most of these papers are based on analyses of statistical observations.

Characteristics of Data The Gusto and the TIMI investigators carried out their prediction modeling on medical studies which were prospective randomized controlled trials (RCT). RCTs have a set of exclusion criteria like old age and high blood pressure. The patients entered into the database are selected carefully. Because of this the data quality of RCTs is usually good and the control possibilities high.
The AMIS Plus data is a registry and has no exclusion criteria but an inclusion criteria (ACS). The data is less selective than with RCTs and represents a better replica of real life as everybody is entered into the cohort. The development of a prediction model is possibly more difficult on a registry than on a RCT which is more homogeneous. If it should be possible to develop such a model, the model would be of more relevance to the general cases, and not only to a specified set of patients.

The copy of the database received from AMIS Plus consisted of approximately 18'000 records and 210 variables stored in one table. The data characteristics are shown in table 2.1.

Characteristics of patients in AMIS Databse

(10.2001 - 05.2005, n=7520)			
Demographics			
Sex			
	Male	n=5415	72.01%
Age			
	mean=65.89	min=22.7	max=99
History of Patient			
Hypertension		n=4075	54.19%
Dyslipidemia		n=4169	55.44%
Smoking			
	Never	n=2475	32.91%
	Former	n=1750	23.27%
	Current	n=2836	37.71%
Diabetes		n=1506	20.02%

Moderate to severe renal disease	n=408	5.43%
Cardiac Insuffieciency	n=341	4.53%
Cerebrovascular Disease	n=422	5.61%
Previous MI or stable angina	n=2560	34.04%
Presenting characteristics		
ECG at presenation		
ST-elevation	n= 4300	57.18%
Q wave	n= 1228	16.33%
ST-depression	n= 2264	30.11%
T-wave changes	n= 2120	28.19%
Left bundle branch block	n= 372	4.95%
Right bundle branch block	n= 428	5.69%
Heart Rhythm		
Sinus rhythm	n= 6801	90.44%
Atrial fibrillation	n= 376	5.0%
Blood Pressure		
Systolic BP	mean=134.06	
Diastolic BP	mean=77.98	
Killip Classification		
1	n=5617	74.69%
2	n=1302	17.31%
3	n=348	4.63%
4	n=208	2.77%
Heartrate mean=78.88		
Resuscitation prior to admision	n=341	4.53%
Cardio pulmonary resuscitation (mechanical)	n=232	3.09%
Defibrillation only	n=270	3.59%
Outcome		
In hospital mortality		
Dead	n=562	7.47%
MACE	n=896	11.91%

Table 2.1: Characteristics of patients in AMIS Databse

2.2 Steps of Data Cleaning

The process of data cleaning was conducted in close collaboration with the AMIS Data Center and Dr. Kurz. Any quality problems identified were first discussed with the AMIS Data Center, which was usually aware of the problem. All my solutions were discussed with either Dr. Radovanovic or Nicole Duvoisin to prevent incorrect actions based on erroneous assumptions.
The process of data cleaning was conducted entirely with the program Clementine from SPSS. The most time consuming step of the data cleaning process was to specify a range and type for each variable and filter out all values lying beyond the defined space. This challenge was tackled in close collaboration with Dr. Kurz. See Appendix A.1 for a full table of all ranges specified.
Another step was the identification and elimination of duplicate records. Patients that were transferred from one hospital to another were at times entered into the database twice.
A pattern of missing values in boolean variables was discovered. Boolean variables were only filled in, if the unexpected / abnormal value occurred. For example for the variable diabetes a value would only be specified, if the patient had diabetes. If the patient didn't have diabetes, the variable would simply not be filled in. This missing pattern was rectified by automatically filling all missing values with a negative boolean value, if one boolean value of a group of alternative values had been specified positively. In our diabetes example this would mean, that the variable diabetes would be filled with a 0, if another comorbidity[3] had been specified. Another group of alternative values were the variables specifying ECG[4] - findings. Care was taken that all variables belonging to a specified subgroup had been issued on the same date.
The variable `basicins` specifying the kind of insurance a patient possesses had to be filtered out altogether. In the old questionnaire the question asking for the insurance had been ambiguous and the answers therefore unreliable. As this problem has been mended since, the variable could possibly be useful for future data mining projects.
All variables of the type string had to be filtered out as well, as there were not enough values specified to attempt any kind of text mining.
Records missing the values for variables date of birth (`birthdat`) , date of admission (`admisdat`) and in hospital survival (`alive`) were discarded. Records missing this vital information were usually hardly filled and dismissed as not to disrupt the process.
As it was not yet clear which variables were to be part of the new risk prediction model it was refrained from discarding any other records because of insufficient filling. The possibility remained that a pattern could be found in the missing values. Another possibility was that a variable was

[3] Comorbidities are coexisting / additional diseases.
[4] Electrocardiogram

only filled in insufficiently, but specified often for records with a negative outcome[5] and therefore especially interesting.
Please see Appendix A.1 for a complete list of all steps undertaken, or refer to the Clementine data stream cleaning.str added on the CD accompanying this diploma thesis.

2.3 Steps of Data Preparation

In order to enable a successful application of the algorithms the data had to be prepared further. In the following sections I would like outline the most important steps and ideas shortly. Most of the data preparation is not very interesting, but vital for a good outcome.

Field creation The AMIS Plus questionnaire is updated frequently and new questions are added. Accordingly new fields are added into the database. This means that some variables needed for further calculations are present in the database, but only filled in sparsely because they are new. Therefore some variables that seemingly were already present had to be recreated and filled. Naturally this was only possible, if the value of the field could be inferred by other variables.[6]

Discretisation Most variables (eg. Age, Weight, etc...) could be discretised automatically by specifying a fixed bin width. Some of the variables though had to be specified manually. Among these was the Body Mass Index with the four categories underweight, normal weight, overweight and obesity that follow an international standard. For the laboratory values Dr. Kurz had to define categories. A target variable-dependent discretisation was attempted, but did not yield any superior results.

The Time and Date Variables As traditional / propositional data mining algorithms can't get information out of encoded dates, the dates were recoded into more useful variables. For example out of the variable `onsetti` which stands for the time of symptom onset, the hour was extracted, enabling to detect a possible difference between AMI's suffered at night and AMI's suffered during daytime. Accordingly information about the month, the year, the weekday etc. was extracted from both the time and date of symptom onset, as well as the time and date of hospitalization and stored in separate variables.

[5] The negative outcome, meaning death of a patient, occurs with about 7.5 % of all records.

[6] The variable `stemi` for example can be calculated from the values of st-elevation and left bundle branch block. In the latest questionnaire it is now a specified variable.

Please see Appendix A.2 for a list of all steps carried out, or refer to the Clementine data stream preparation.str added on the CD accompanying this diploma thesis.

3

Evaluation of existing Risk Scores

To develop a new risk prediction model the first step must be to analyze existing risk scores in the field of myocardial infarction. A valid benchmark is required to facilitate evaluation of newly found methods. The evaluation of an existing score also helps to identify the potentials for new development.

In the medical literature the TIMI Risk Score for ST-Elevation[1] is currently considered to be the most significant risk score in mortality prediction after an AMI. The TIMI Risk Score is a so-called bedside[2] score and easy to calculate. This aspect increases the relevance of the TIMI score for our purposes, as we are also looking to develop a bedside score.
Another valuable predictor is the killip classification. The killip classification categorizes the presence and severity of heart failure at the time of initial presentation into the categories 1-4, with increasing severity for higher numbers. Although the killip classification is a very simple method, its predictive value is not to be underestimated (Khot et al., 2003).

There were three more scores that were checked on their benchmark abilities. The Score of the Gusto Team was discarded, because of its elaborateness. The Gusto Score definitely doesn't qualify as a bedside score. Furthermore it is hardly used in practical hospital life.
The TIMI Score for nonStemi was discarded because too many characteristics required for its calculation are not available in the AMIS dataset.[3] The Simple Risk Score (Morrow et al., 2001) is a risk score predicting the mortality using just the three values age, heart rate and systolic blood

[1] The medical term ST-Elevation (STEMI) indicates that the ECG findings suggest ST-Elevation, left bundle branch block or both. *Stemi* MI, in contrast to *nonStemi* MI, have been found to be more dangerous for the patient.

[2] In this context bedside denotes that all characteristics required for the calculation of the score can be retrieved at first presentation with the patient, so to speak by the patients bedside. This guarantees on the one hand that the score can be calculated swiftly. On the other hand it ensures that no variables are included in the score, which are dependent on the course of the disease and therefore dependent of the outcome to be predicted.

[3] Among other the information whether aspirin was used in prior 7 days, or the occurence of at least 2 anginal events in prior 24 hours (Antman et al., 2000).

pressure. It is less established in the clinical world.

All evaluation will be assessed using the Area under the ROC curve as an index of model performance (Hanley & McNeil, 1982). Values of the AUC from 0.7 on can be interpreted as modestly successful, results above 0.8 as usable (Ohman, Granger, Harrington, & LEE, 2000). In addition to the ROC curve evaluation the mortality rates will be compared.

3.1 The TIMI Risk Score for ST-Elevation Myocardial Infarction

(Morrow et al., 2000) The TIMI Risk score for ST-Elevation Myocardial Infarction was developed on a population of 14114 patients and published in the year 2000. The aim was to design a convenient bedside predictor for everyday hospital life.
The study population consisted of *stemi* patients that were fibrinolytic-eligible[4] and was collected in 800 hospitals around the world. The data was collected in the early nineties (Morrow et al., 2000). Exclusion criteria included history of cerebrovascular disease,[5] systolic blood pressure greater than 180 mm Hg, diastolic blood pressure greater than 110 mm Hg, cardiogenic shock,[6] or increased risk of severe bleeding.

To calculate the risk score the TIMI investigators proceeded as follows: First the univariate predictors of all variables of the original TIMI cohort were calculated. Keeping only variables with $p > 0.05$, a stepwise logistic regression was carried out. All variables were ranked by the z-score and the ones with the lowest score sequentially removed, until the characteristics boiled down to 10. These ten variables, capturing 97% of all prognostic information, were used to compile the TIMI Risk Score as seen in table 3.1.
The TIMI Risk score was evaluated on a cohort of 3700 patients. On the original cohort of 14114 patients an AUC of 0.779 was achieved. On the validation cohort an AUC of 0.746 could be measured.

[4]See Appendix D.1 for the full list of indications for fibrinolytic therapy

[5]Cerebrovascular disease is an interruption of the blood supply to any part of the brain, resulting in damaged brain tissue. Also called stroke.

[6]Cardiogenic shock is a disease state where the heart is too damaged to supply sufficient blood to the body.

TIMI Risk Score for STEMI	
Historical	
Age 65-74	2 points
Age $\geq$ 75	3 points
DM/HTN or angina	1 point
Exam	
SBP $<$ 100	3 points
HR $>$ 100	2 points
Killip II-IV	2 points
Weight $<$ 67kg	1 point
Presentation	
Anterior STE or LBBB	1 point
Time to rx $>$ 4 hrs	1 point
Risk Score = Total	(0-14)

Table 3.1: Calculation of TIMI Risk Score (Morrow et al., 2000)

3.1.1 How to apply the TIMI risk score

The TIMI Risk score can be calculated easily according to the spread sheet displayed in table 3.1. The patient is given points for the 11 chosen characteristics, these points are then accumulated. A patient with a low TIMI Risk Score, for example 3, has got a better chance of survival than a patient with a high TIMI Risk Score, for example 11. Fig. 3.1 shows the mortality rate prediction for all TIMI Risk Score categories as appointed by the TIMI investigators. It also shows what percentage of patients were at risk in the original cohort in each category.

The TIMI Risk Score is an often relied upon measure. There is however a drawback which suggests the TIMI Risk Score to be unsuitable. The TIMI Risk Score was developed on a very uniform cohort. The exclusion criteria for the original cohort are numerous. In the AMIS dataset these exclusion criteria alone would exclude almost 18 % of the cohort. Furthermore all patients of the TIMI cohort received thromobolysis therapy. The TIMI Risk Score is biased by its reliance on a thrombolysis only data cohort. The question arises whether a risk score that was derived from a uniform cohort can also perform well in every day life, where many records don't fit the exclusion criteria. As the AMIS dataset represents a very mixed cohort, it is possible to answer this question.

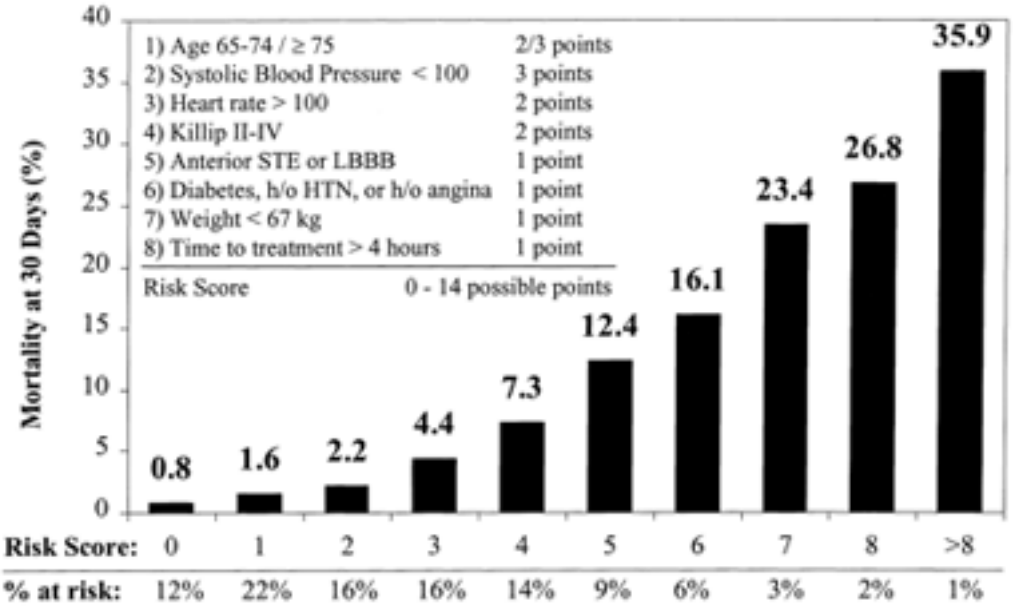


Figure 3.1: Mortality Rate with TIMI Risk Score (Morrow et al., 2000)

3.2 TIMI Risk Score applied to the AMIS dataset

There are two aspects that play a role when applying the TIMI Risk Score to the AMIS dataset. First we want to find out how well the TIMI Risk Score predicts the mortality on our data. Second we use the TIMI Risk Score as a benchmark for the AMIS dataset. Different points of interest arise from these two angles. Please note the the whole cohort has been used for the following evaluations.

First we focus on the performance of the TIMI Risk Score. How does the TIMI Risk Score perform on data that has been filtered according to the exclusion criteria above mentioned. Has time and progress in treatment had an impact? Then we want to see how the TIMI Risk Score performs when the exclusion criteria are disregarded. Thirdly we want to analyze the performance of the TIMI Risk Score on a non-*stemi* cohort.

Next the focus shifts to analyzing the AMIS dataset. More knowledge about the invisible structure of the AMIS dataset can be gained by dividing it into subsets according to certain characteristics, and the comparison of the newly created subsets. The most interesting and obvious characteris-

tics are *stemi / nonStemi*, the different treatment groups (PCI[7], thrombolysis[8] and no treatment[9]) and with or without exclusion criteria.
The TIMI Risk Score is designed to predict the mortality rate, its aim is therefore the probability of death. It can be imagined though, that a successful score could also target complications (including death) in the course of the disease. For this reason the combined endpoint of major adverse cardiac events (MACE) was also investigated.

3.2.1 Differences in datasets

When applying the TIMI Risk Score to the AMIS dataset there are differences in the datastructure and variables that had to be overcome. The most important shall be accounted for in the next paragraph. The complete list of all calculations necessary to calculate the TIMI Risk Score on the AMIS dataset is recorded in appendix A.2.

The TIMI Risk Score aims to calculate the *30-day mortality* of a patient. In the AMIS dataset we only have a variable which designates the in-hospital mortality[10]. According to medical staff there is almost no difference between in-hospital and 30-day mortality. The difference was therefore ignored.
In the calculation of the TIMI Risk Score the term *Anterior STE*[11] or LBBB[12] shows up. In the AMIS dataset the variable for Anterior MI is missing. In the dataset of the Triemli hospital, which is a subset of the AMIS dataset, the missing variable exists. In the Triemli dataset 44% of all patients suffer an Anterior MI. A new variable was generated in the AMIS dataset that distributed 0/1 randomly according to the proportion in the Triemli dataset[13].
In the AMIS dataset the variable *time to treatment*[14] has to be calculated. At what time the treatment was applied can only be elicited, if the patient underwent PCI or thrombolysis treatment, as otherwise no time is recorded. This problem was solved by adding a point to every patients risk

[7] Percutaneous coronary intervention (PCI) is a treatment procedure that unblocks narrowed coronary arteries without performing surgery.

[8] Thrombolytic therapy involves the use of drugs that break up or dissolve blood clots, which are the main cause of both heart attacks and stroke.

[9] In the AMIS dataset no treatment is specified as true, if neither PCI nor thrombolytic therapy have been carried out.

[10] In the last update of the AMIS questionnaire variables about the follow up of patient history were included. For a future project the necessary information would be available.

[11] Anterior ST Elevation indicates the location of MI and is retrieved from the initial ECG finding.

[12] LBBB stands for Left bundle branch block and designates the location of the MI. This information is collected from the initial ECG finding.

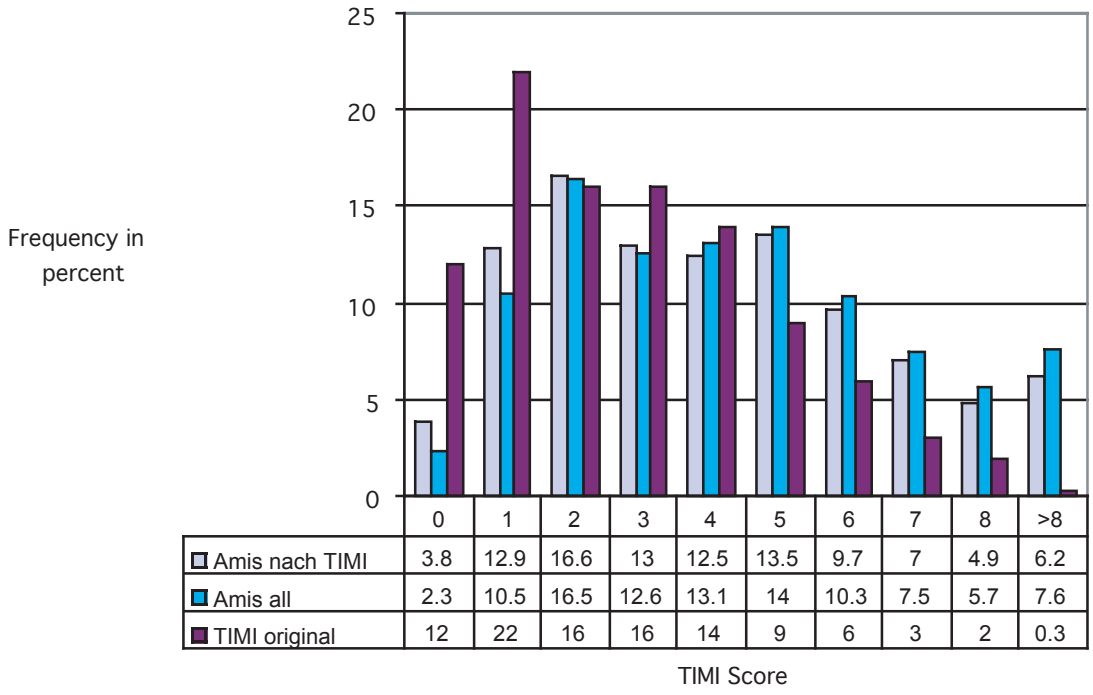
[13] This measure may lead to a overestimation. Note that the results show the TIMI Risk Score to be too low in risk levels. Overestimation is therefore the friendlier method.

[14] Time to treatment stands for the time of the symptom onset, until the patient receives treatment

score that lacked information concerning the time.
All decisions made concerning the above described adaptations were made by the medical staff, ensuring their acceptability.

3.3 Results of Evaluation

3.3.1 Comparison of people at risk



	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	3.8	12.9	16.6	13	12.5	13.5	9.7	7	4.9	6.2
Amis all	2.3	10.5	16.5	12.6	13.1	14	10.3	7.5	5.7	7.6
TIMI original	12	22	16	16	14	9	6	3	2	0.3

Figure 3.2: Original TIMI cohort vs. AMIS set and AMIS subsets - Distribution of percentage at risk

Our investigation shows that AMIS dataset differs noticeably from the original TIMI cohort. Figure 3.2 shows the distribution of people at risk. With the original TIMI cohort 50% of all patients are part of the risk groups < 3. In the AMIS dataset there are only few low risk patients. The patients are altogether more evenly distributed. Unlike in the original TIMI cohort the section > 6 is still amply populated.
As anticipated the dataset AMIS all[15] has more high risk patients as the AMIS according to TIMI

[15]The dataset *AMIS all* includes the whole AMIS database without appliance of any restrictions or exclusions.

dataset[16].

The origin of discrepancies in the two datasets (original TIMI and AMIS cohort) has two main factors. Firstly a span of about 3 years has passed since the TIMI investigation and the AMIS investigations. Although some of the AMIS data was collected around the same time, the big bulk of data was collected from 2002 onwards. Secondly, unlike the AMIS dataset the TIMI cohort is very homogenous, having been collected from study populations. Because the AMIS data is a registry and large and small hospitals take part in collecting the data, virtually any person suffering an AMI is included.

3.3.2 Performance of TIMI Risk Score

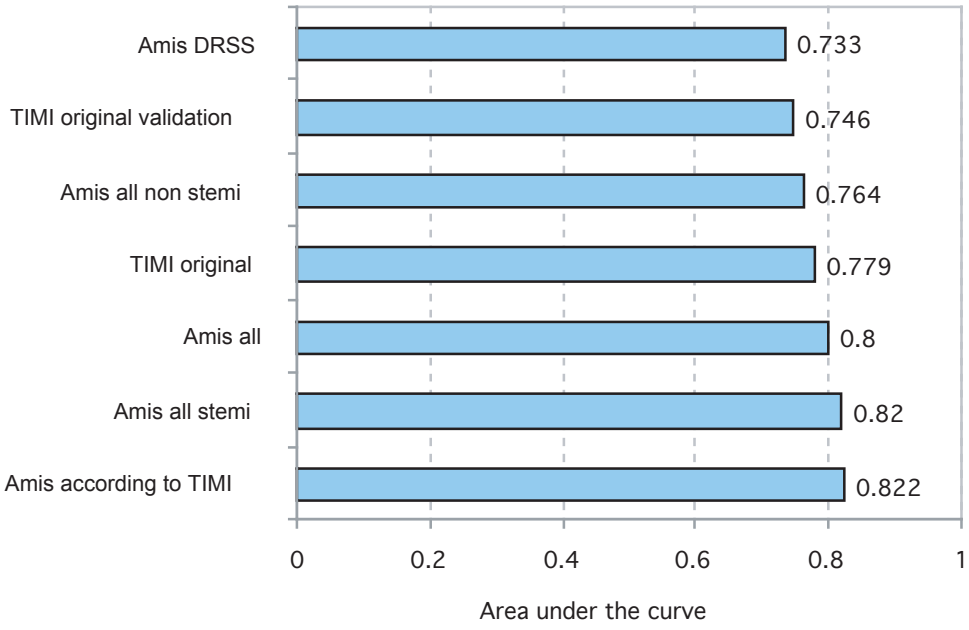


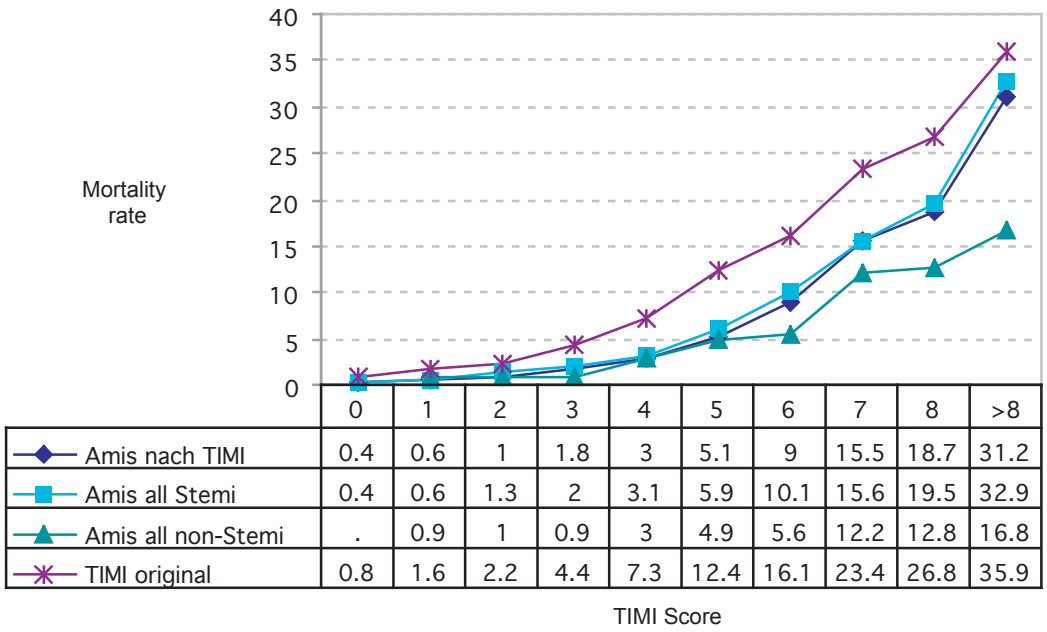
Figure 3.3: AUC comparison of TIMI Risk Score on various datasets

To determine the validity of the TIMI Risk Score the AUC of the different datasets are compared, as suggested by Ohman et. al. (Ohman et al., 2000). The following datasets are compared to the original TIMI cohort (*TIMI original*) and the original TIMI validation cohort(*TIMI original validation*): *AMIS according to TIMI* (see footnote 16), *AMIS all* (see footnote 15), *AMIS all STEMI*,[17]

[16]The *AMIS according to TIMI* dataset contains the remainder of the AMIS datset, after applying all restrictions that were specified in the original TIMI cohort.

[17]The AMIS all STEMI dataset consists of all *stemi* cases in the AMIS database, no restrictions applied.

AMIS all non STEMI[18] and *AMIS MACE*.[19]. Fig. 3.3 shows the outcome.
Surprisingly the TIMI Risk Score achieves better results on our AMIS cohort than on the original TIMI cohort. The TIMI Risk Scores performs generally better than expected, as even the TIMI Risk Score of the *nonStemi* population outperforms the original validation material. The TIMI Risk Score applied to the whole AMIS dataset achieves a good AUC of 0.8.



	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	0.4	0.6	1	1.8	3	5.1	9	15.5	18.7	31.2
Amis all Stemi	0.4	0.6	1.3	2	3.1	5.9	10.1	15.6	19.5	32.9
Amis all non-Stemi	.	0.9	1	0.9	3	4.9	5.6	12.2	12.8	16.8
TIMI original	0.8	1.6	2.2	4.4	7.3	12.4	16.1	23.4	26.8	35.9

Figure 3.4: Comparison of TIMI Risk Score Mortality Rates

Fig. 3.4 compares the mortality rate prediction of the TIMI Risk Score (purple line) with the actual outcome on the AMIS datasets. The actual mortality rates of all datasets are considerably lower than predicted.
The question arises, why is the predicted mortality rate so much higher. A reason for this is certainly the advancment of therapy methods and that the survival of patients is better in general. This suggestion is endorsed by Fig. 3.5, which shows that especially patients who received PCI treatment have a much better chance of surviving.

[18]The AMIS all non STEMI dataset consists of all *nonStemi* records in the database, no restrictions applied
[19]The AUC for AMIS MACE is not calculated on the mortality rate, but on the mace rate (death, reinfarction, stroke, cardiogenic shock quota). The idea is to find out, whether the TIMI Risk Score is also valuable for the prediction of complications in general

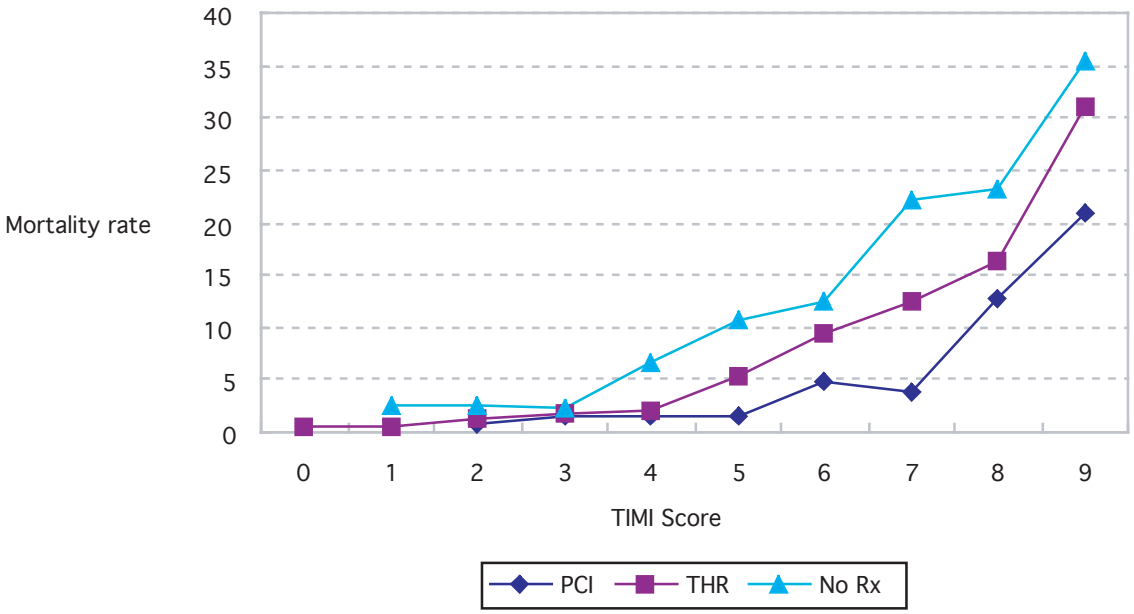


Figure 3.5: Comparison of TIMI Risk Score Mortality Rates split by treatments on AMIS according to TIMI dataset

Fig. 3.4 shows that especially the predicted mortality rates of *nonStemi* records are far above the actual outcome. This can not be taken as an argument for the failure of the TIMI Risk Score as it was only developed for *stemi* cases. It can be taken as a vital hint though, that the distinction between *stemi* and *nonStemi* is possibly vital for a new score with the aim to predict all cases.

3.3.3 MACE rate

The graph of the MACE rate (Fig. 3.6) shows us a different picture than the mortality rate graphs. Surprisingly the Amis all *nonStemi* graph lies below the original TIMI predictions of mortality, while the Amis all *stemi* graph shows a higher rate than the graph of the TIMI Risk Score. This is proof, that also for the development of a MACE predictor the variable *stemi* vs. *nonStemi* is of importance.

Recapitulating the results of this section, one can say two important things.
Firstly although the AUC of the TIMI Risk Score is good, the mortality predictions that go with it are no longer gripping tightly, especially when new therapies like PCI are applied. All tested possibilities were predicted too high by the TIMI Risk Score.

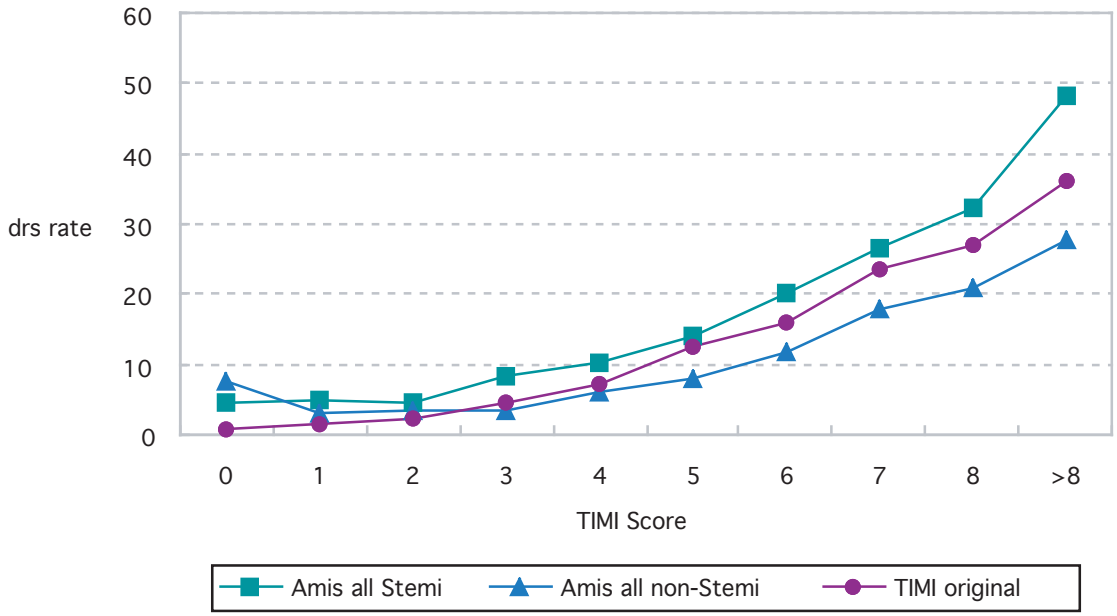


Figure 3.6: Comparison of TIMI Risk Score mortality prediction with MACE rate

Secondly *nonStemis* have a more even risk increase compared to *stemis*. The characteristic *stemi / nonStemi* as such is a very important indicator when it comes to mortality as well as complications.

4

Identification of significant Variables and Application of Data Mining Algorithms

This section describes the algorithms used and the test settings when determining the best prediction model. In my opinion this was the most interesting and intense part of the diploma thesis. This stage of the project had a iterative character, as all the components were interdependent. The difficulty was to reach the global optimum and not get stuck in a local one.
The components of the test settings were the algorithm, the parameter setting of the individual algorithm, the training set selected, the subset of interest and the variables picked. For the testing of the variables the WEKA workbench (Witten & Frank, 2005) was used.

Methods of validation and evaluation To test the different algorithms and settings a quick but reliable method of validation and evaluation was necessary. 10-fold cross validation[1] was used as validation.
To evaluate the algorithms and settings the area under the ROC curve was chosen. Because of the uneven target distribution the normal misclassification error would not have been a satisfactory estimate. An algorithm predicting every patient to survive would have classified 93% correctly, as 93% of the patients do survive. Because of the skewed class distribution, an evaluation based on accuracy is not helpful. Foster and Provost show that in this environment the best results are achieved with an ROC evaluation (Provost & Fawcett, 2001).

[1]Cross-validation is a typical method of validation in the field of data mining, especially since Kohavi suggested its superiority to the bootstrap method in 1995 (Kohavi, 1995). With 10-fold cross-validation the dataset is split into 10 parts of approximately the same size and target-distribution. The algorithm is then trained 10 times on the dataset, each time leaving out one part, with which the validation is carried out. The resulting 10 validation figures are averaged. This way more data remains for the training of the algorithm whereupon yielding more solid validation results.

4.1 The Training Sets

Quality vs. Quanitity From Mrs. Radovanovic I had been given the tip, that the older data was mostly disrupting and not of much help. She personally usually leaves out the older data when carrying out any statistical analyses for two reasons. On the one hand the therapy methods have altered remarkably during the past years, changing the mortality risk when suffering from AMI. On the other hand, the quality of the older data is often suboptimal. Partly because some of the variables have only been introduced at a later stage.
The cutoff point was set to October 2001. In October 2001 all comorbidity variables had been introduced. The comorbidity variables describe all other diseases a patient might have other than AMI. The comorbidity variables were likely to be good predictors. When tested with an *AODE* algorithm[2] using the same parameter settings the new records (n= 7520) performed clearly better with an AUC = 0.8755 than the whole dataset (n= 16205, AUC = 0.8595). The validation used was a 10-fold cross validation. This showed clearly that the quality had a higher impact than the quantity. Therefore the decision was made to develop and validate the model only on the newer records.

Clustering From a medical point of view it seemed very probable, that there would be a great difference between *stemi* and *nonStemi* patients. This fact was underlined in section 3.3, showing the big difference between the *stemi* and the *nonStemi* cohort. Most scores developed so far, are only applicable to either the one or the other category. For this reason 2 clusters were manually made. Algorithms were tested on the *stemi* patients, *nonStemi* patients and both of them together. When tested with the *AODE* algorithm using the same parameter settings, the *stemi* cohort (n = 4598) achieved an AUC of 0.875, the *nonStemi* cohort (n = 2885) achieved an AUC of 0.860 and the combined cohort (n = 7516) achieved an AUC of 0.875. Since the prediction achieved on either cluster is not superior to the prediction achieved on the full cohort, the decision was made to continue with the more practical version and develop the model on the full cohort.

A balancing act The uneven target variable distribution was probably the biggest difficulty to overcome. The problems of uneven target variable distribution and different costs attributed to the target classes are related. It has been suggested, that by increasing the prevalence of the more costly instances, better training results can be achieved (Breiman, Friedman, & Olsen, 1984). In an attempt to ameliorate the learning environment for the algorithms a set of training sets was constructed with balanced data. From the filtered out older data the records with negative outcome were extracted and merged with the training sets. This raised the target distribution

[2]See section 4.4 for more information on the AODE algorithm.

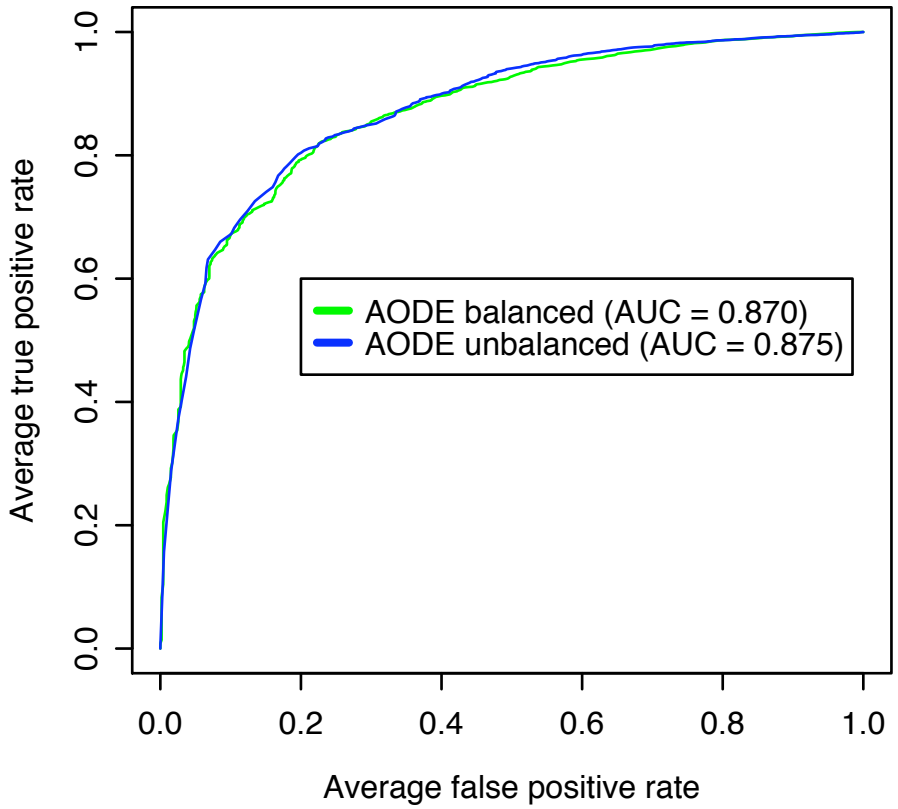


Figure 4.1: ROC Curve comparison of AODE performance on balanced vs. unbalanced data

form 93:7 to 86:14. Then new records were discarded until the distribution of the training set reached 50:50 [3]. Through the removal of records with a positive outcome the number of records was reduced from n ≈ 6790 to n ≈ 2450 per fold.

All algorithms were tested on the original unbalanced datasets, as well as on the balanced datasets [4]. The validation of the results was carried out with the original datasets.

None of the algorithms performed significantly better when trained on the balanced training sets. Using the same parameter settings the *AODE* algorithm achieved on the balanced data an AUC of 0.870 and on the unbalanced data an AUC of 0.875. See figure 4.1 for a comparison of the ROC curves.

Again the less practical version was not convincing enough. Hence the decision to continue with the unbalanced datasets.

[3] Boosting was also tested but turned out to deteriorate the prediction capability.

[4] Other distributions were also tested but did not achieve any interesting results.

4.2 The Target

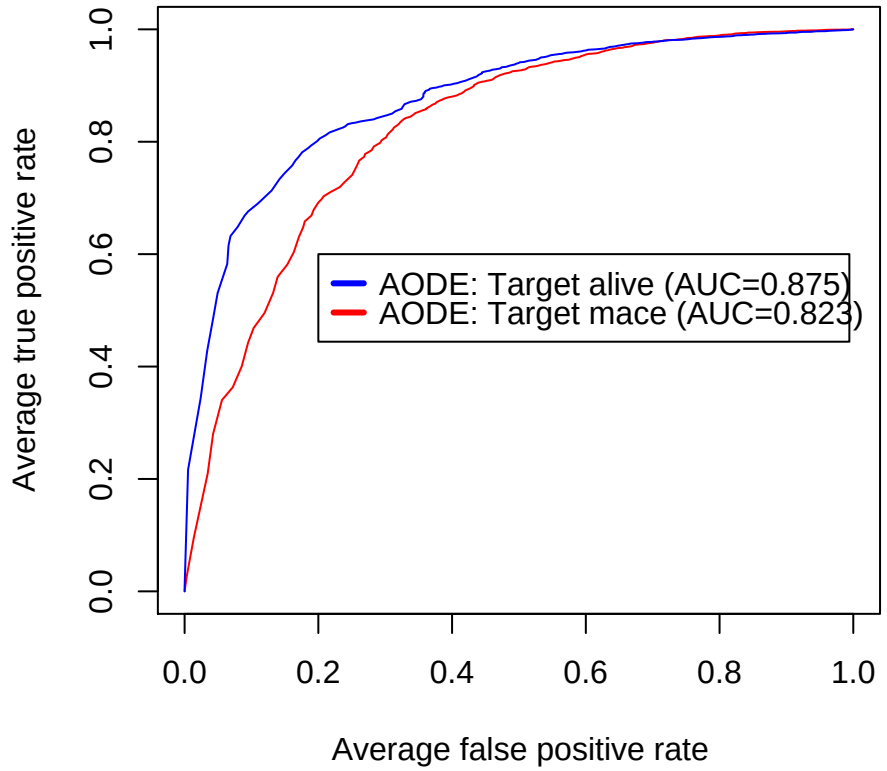


Figure 4.2: ROC Curve comparison of AODE performance using endpoint alive vs. the combined endpoint mace

The obvious and given target to calculate the mortality rate is the variable `alive` defining the outcome of in hospital patient survival. Apart from the mortality risk it would be interesting for the attending physician to know the risk of complications as specified in chapter 3. A combined endpoint including death, reinfarct, shock and stroke (mace) was constructed and tested. The prediction of the mace-rate is also interesting from a different point of view. The target distribution of 88:12 is more friendly for the algorithms. On the other hand, a combined endpoint is usually more difficult to predict, especially because in this case it could not be split up into the four sub-targets, as there were not enough records with negative outcome for each sub-target. See table 4.1 for the sub-target distributions.

Name	
Re-infarction	2.14 %
Cardiogenic shock	6.7 %
Cerebrovascular event	1 %
Death	7.47 %

Table 4.1: Distribution of sub-targets of the target mace

All tests conducted in this phase were applied to both target structures. See figure 4.2 for a comparison of the ROC curves. The combined endpoint *mace* proved more difficult to predict then the endpoint *alive*, in spite of the above mentioned advantage of having a better target-variable distribution. The best dataset and algorithm choices for the endpoint *mace* were identical to those for the endpoint *alive*. Please note that more thought and time was invested in developing a model predicting the mortality rate.
The *AODE* algorithm using the same parameter settings achieved an AUC of 0.875 with the endpoint *alive*, and an AUC of 0.823 with the combined endpoint *mace*.

4.3 The Identification of critical Variables

The choice of the variables was made in close collaboration with Dr. Kurz. Next to the prediction capability of the individual variable, all variables were also scrutinized regarding their interdependencies on other variables as well as their validity. The variable `transfer` (defining whether a patient has been transferred hospitals) for example has a good prediction capability, yet was discarded because the decision of transferring a patient is taken once therapy has begun. Patients are usually transferred from small hospitals to a bigger ones because of better therapy possibilities. The transfer of a patient is usually a therapy decision based on the state of the patient.
The results of a *C 4.5* decision tree with pruning, a subset evaluator combined with a best first search algorithm and a single attribute evaluator combined with a ranking algorithm were compared to reduce the number of variables to the thirty most interesting.
The 30 variables were discussed with Dr. Kurz and many of them crossed out because of their possible ambiguity. Variables with a strong dependence on each other were tested pairwise, keeping only the stronger one. For example the variables `resusci` (Resuscitation prior to arrival at hospital), `cprarr` (Cardiopulmonary Resuscitation) and `defibarr` (Cardioversion / defibrillation) are interconnected. `resusci` specifies whether any kind of reanimation was conducted, the variables `cprarr` and `defibarr` specify what kind. All three characteristics are strong predic-

Testset	AUC	Variables chosen
1 variable	0.8718	killip
4 variables	0.8452	age, killip, systbp, hrtrate
7 variables	0.8755	age, killip, systbp, hrtrate, cprarr, cmcardin, cmcevdis
17 variables	0.8675	dyspnea, stelev, twavec, lbbblck, rbbblck, hrtrythm, diastpb, resusci, defibarr, rgaspir, age, killip, systbp, hrtrate, cprarr, cmcardin, cmcevdis
30 variables	0.8722	sex, height, dyspnea, stelev, qwave, stdepres, twavec, lbbblck, rbbblck, hrtrythm, diastpb, resusci, defibarr, rgaspir, rgaceinh, rgnitrat, rgdigoxi, rgdiuret, diabetes, cmrenald, histhta, histhlip, histsmok, age, killip, systbp, hrtrate, cprarr, cmcardin, cmcevdis

Table 4.2: Test results of different sets of variables

tors. Tests showed that the variable cprarr is the best predictor of the three. Therefore cprarr was retained and the other two variables were discarded.

Next variables were left out stepwise, testing which could be discarded without reducing the prediction capability of the model. See table 4.2 for the comparison of the performance of the different variable subsets. They were all tested with the *AODE* algorithm using the same parameter settings. See the remaining 7 variables in table 4.3. The 7 selected variables were readily accepted by the steering committee of AMIS Plus, which is an important sign as to their clinical significance. They are easily and unambiguously measured, can be recorded immediately at first presentation of the patient, and independent of the course of disease.

Variable	Name	Description
age	Age	Age at time of infarction
systbp	Systolic Blood Pressure	Systolic Blood Pressure at admission
hrtrate	Heart rate	Heart rate at admission
killip	Killip classification	Killip classification at admission
cprarr	Cardiopulmonary resuscitation	Cardiopulmonary resuscitation prior to arrival at hospital
cmcardin	Cardiac insufficiency	Comorbidity: Cardiac insufficiency
cmcevdis	Cerebrovascular disease	Comorbidity: Cerebrovascular disease

Table 4.3: Chosen variables for the AMIS Model ranked according to Ranker algorithm

4.4 The Algorithms

The algorithm needed two important characteristics: it had to be able to predict probabilities and it had to overcome the uneven target distribution.
Naive Bayes algorithms are known for their good probability prediction capability. With the *AODE* algorithm the WEKA workbench offers a naive bayes based algorithm with less strong independence assumptions (Webb, Boughton, & Wang, 2005, 2002). This seemed useful as with medical data there exist more interdependencies than in other fields of application.
Another interesting algorithm was the *C 4.5* tree. Provost and Domingos showed that the *C 4.5* algorithm yields good probability estimations when no pruning is applied and the laplace correction is used (Provost & Domingos, 2003). The tree generated by the *C 4.5* algorithm has the advantage of representing the gained information in an understandable form. See Appendix E for the tree generated by the *C 4.5* algorithm, when trained on the described training set.

Other interesting candidates were the logistic regression, support vector machines, the multilayer perceptron and the RBF network. The logistic regression was primarily used by the other investigators and therefore an interesting testing object. The support vector machines are very powerful and can optionally be combined with logistic regression yielding better estimations. The multilayer perceptron and the RBF network being neural networks and therefore supposedly mighty were definitely worth a try.

Metaclassifiers After having built prediction models with the above mentioned algorithms, it was attempted to build a metaclassifier. The building of a ROC convex hull (ROCCH) was inves-

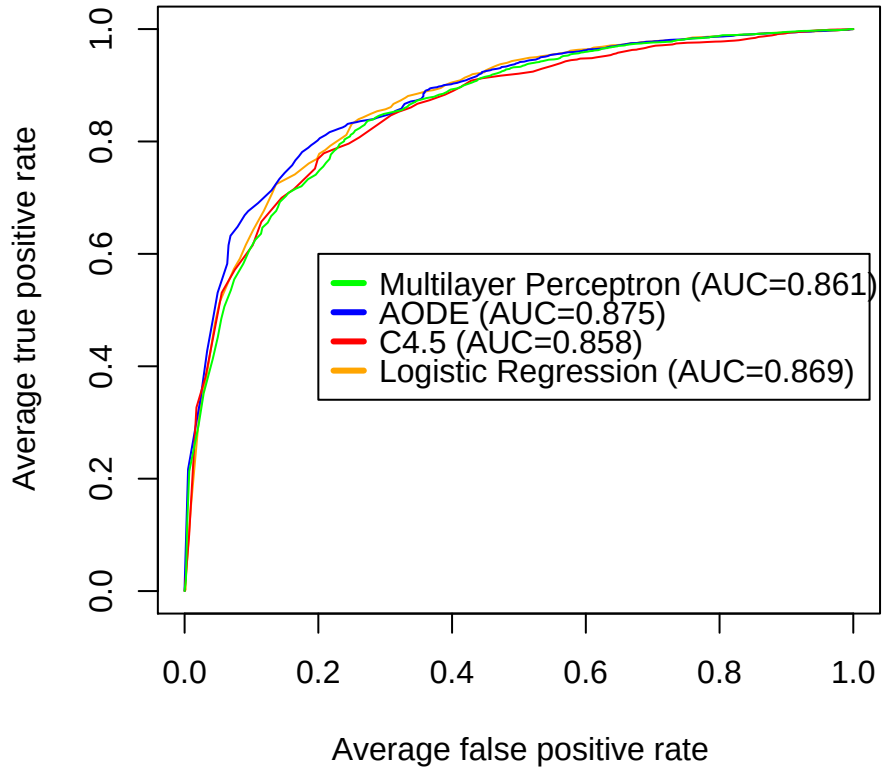


Figure 4.3: ROC Curve comparison best performing algorithms

tigated according to the methods described by Foster and Provost (Provost & Fawcett, 2001).

4.4.1 Weapon of choice

The tested algorithms performed all astonishingly similar on the prepared data. See figure 4.3 for the comparison of the ROC curves of the best performing algorithms. Figure 4.3 shows that the building of a ROCCH is not going to yield any interesting results. Scott et al. show that metaclassifiers cannot achieve better results than the ROCCH (Scott, Niranjan, Melvin, & Prager, 1998). The building of metaclassifiers was therefore also discarded. In a way this fact supported the practicability of the solution. As a rule of thumb one can state that the simpler the model, the easier it is accepted and the more robust it survives over time.

The support vector machine, even though combined with logistic models, did not yield acceptable probabilities and was dropped.

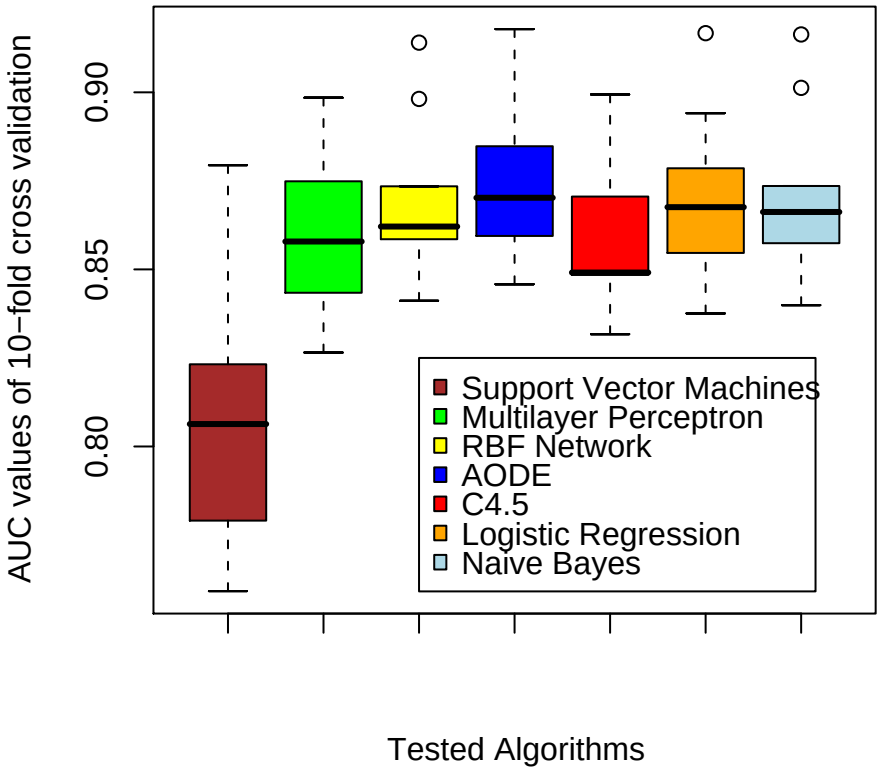


Figure 4.4: Area under the curve comparison of different algorithms

When comparing the ROC curves in figure 4.3 it is not at once clear which algorithm is superior to the others. See figure 4.4 for another visual comparison of the tested algorithms. Note that the graph in figure 4.4 does not show the whole scale from 0 to 1 on the y-axis. The boxplots show that the *AODE* algorithm is the best performing algorithm.

4.5 Summary: The AMIS Model

The AMIS Model was built on the newer records (from October 2001 onwards) with unbalanced training sets. All input variables were subjected to a fixed bin width discretization. No other filters were applied. The AMIS Model applies for *stemi* as well as *nonStemi* records. The AMIS Model can be used as a predictor of the mortality risk when trained with the target `alive`, or as a predictor of the mace-risk, when trained with the constructed target `mace`.
The 7 input variables are age (`age`), killip classification (`killip`), systolic blood pressure (`systbp`),

heart rate (`hrtrate`), cardio-pulmonary resusciation (`cprarr`), history of heart failure (`cmcardin`) and history of cerebrovascular disease (`cmcevdis`).
The best algorithm turned out to be the *AODE* algorithm with no special parameter settings.

5

Evaluation of AMIS Model

To compare the AMIS Model to our established benchmark (TIMI Risk Score) the values of the AUC and the ROC curves were compared. Points of interest were the endpoints *alive* and the combined endpoint death, reinfarct, shock and stroke (*mace*). The results of the prediction models were split up and compared according to the therapy subgroups.
In this section it is shown that in every tested situation the ROC curve of the AMIS Model lies at all points above the ROC curve of the TIMI Risk Score. It achieves the best prediction capacity on the subgroup treated with PCI with an AUC of 0.885. The distribution of the predicted probabilities is even, making the model superior to the TIMI Risk Score regarding low risk patients.

5.1 Method of Evaluation

As our class distribution is very skewed, evaluation measures based on accuracy don't work. A cost matrix specifying the cost of misclassifications is neither available nor conceivable, as it would be ethically problematic to calculate the cost of death. It is however clear, that a false positive error is more costly than a false negative error.
ROC graphs describe the trade off between the hit rate and the false alarm rate, independent of the class distribution and independent of the costs. The ROC graph allows us to compare our results without having to specify any kind of loss function.
As the comparison by ROC curve is widely applied in the medical field, this validation has an additional practical value. It simplifies the acceptance of the prediction model.
Note that the data cohort is different from the one used in chapter 3. The prediction models are no longer applied on all data, but, for the reasons stated in chapter 4, only on the data recorded since octobre 2001.

To ensure a thorough evaluation a 10-fold cross-validation was carried out. The ten resulting ROC curves were averaged by threshold averaging. Threshold averaging averages points sam-

pled based on the thresholds that produced these points (Fawcett, 2004). Unlike vertical or horizontal averaging the true positive rate as well as the true negative rate are both directly included. For the AUCs the mean was calculated.

5.2 Endpoint ALIVE

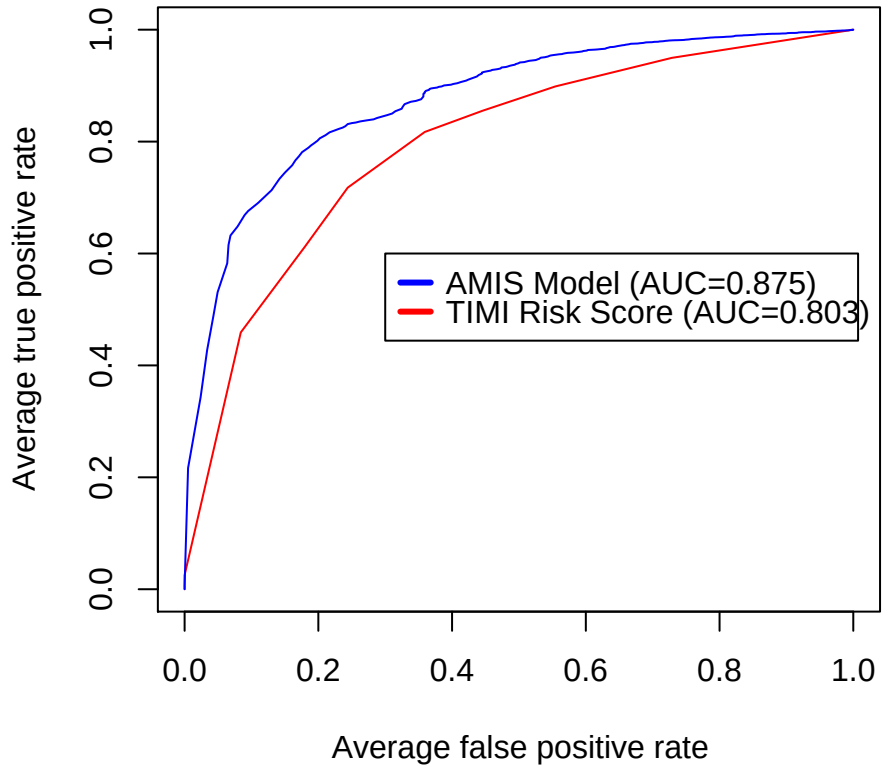


Figure 5.1: ROC Curve comparison of AMIS Model vs. TIMI Risk Score target alive

The primary target of this project was to create a predictor of the mortality risk for all ACS patients. The AMIS Model achieves an AUC of 0.875 on the collective records. The ROC curve of the AMIS Model lies above the ROC curve of the TIMI Risk Score on any given point of the curve. This means that under any cost assumptions the AMIS Model is superior to the established risk score. Figure 5.2 and 5.3 show the error rate and accuracy of the AMIS Model and the TIMI Risk Score. The graphs have not been averaged and all 10 graphs resulting from the cross validation are visible.

To establish the significance of the result a pairwise t-test was calculated yielding a t value of 5.45 and a p-value of 6.212.

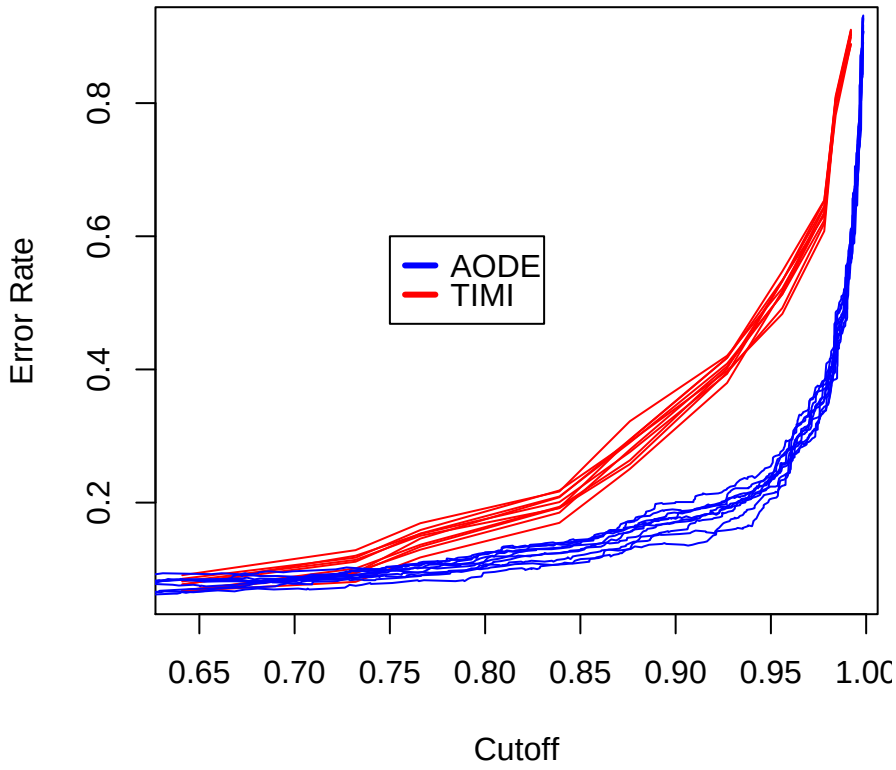


Figure 5.2: Error graph comparison of AMIS Model vs. TIMI Risk Score target alive

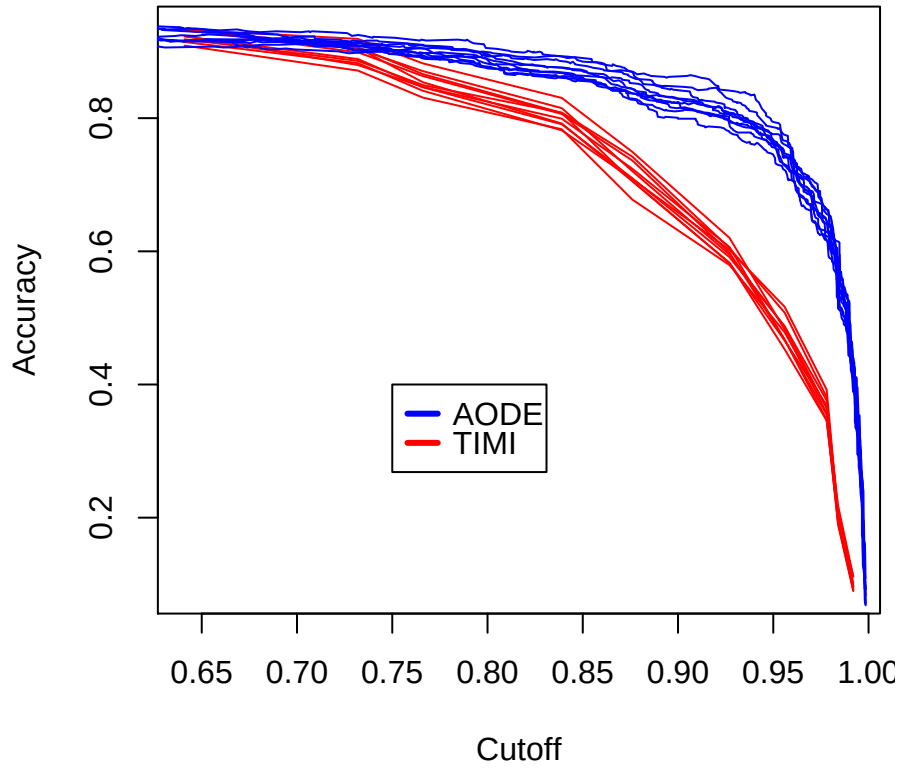


Figure 5.3: Accuracy graph comparison of AMIS Model vs. TIMI Risk Score target alive

5.2.1 Subgroup Analysis according to Therapy

The TIMI Risk Score was developed on a data cohort of thrombolysis patients. Accordingly it is expected to yield very good results on that therapy group. The AMIS Model in comparison was developed on a more heterogeneous cohort with no indication of which therapy was chosen. The model is therefore applicable to all patients, regardless which therapy they shall undergo. It is nevertheless interesting to analyze the performance on the different subgroups. Table 5.1 lists the distribution of the subgroups.

At the time of the development of the TIMI Risk Score the most important therapy was thrombolysis. This has now shifted with a clear advantage towards the PCI therapy. In the bigger hospitals of Switzerland almost all patients undergo PCI therapy.

See figure 5.4 for a comparison of the ROC curves split up according to the therapy groups. The AMIS Model on a thrombolysis cohort achieves an AUC of 0.852, on the no treatment cohort it achieves an AUC of 0.788 and finally on the PCI cohort it achieves a AUC of 0.885.

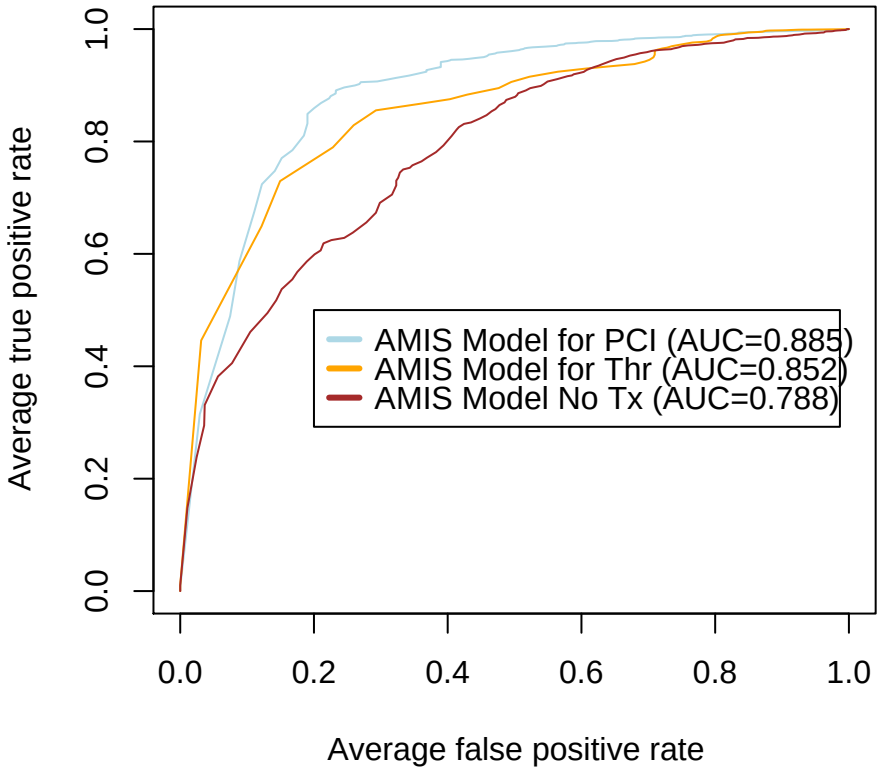


Figure 5.4: ROC Curve comparison of AMIS Model therapy subgroups

No treatment

On a data cohort consisting of patients who received neither thrombolysis nor PCI treatment the AMIS Model achieves an AUC of 0.788. The TIMI Risk Score achieves an AUC of 0.673. For this data cohort the TIMI Risk Score is not a very good predictor. The AMIS Model does not do as well as over all, still achieves an acceptable result. See figure 5.5 for the comparison of the curves.

Subgroup	
Thrombolysis	n≈920
PCI	n≈4930
No treatment	n≈2000

Table 5.1: Distribution of the treatment subgroups

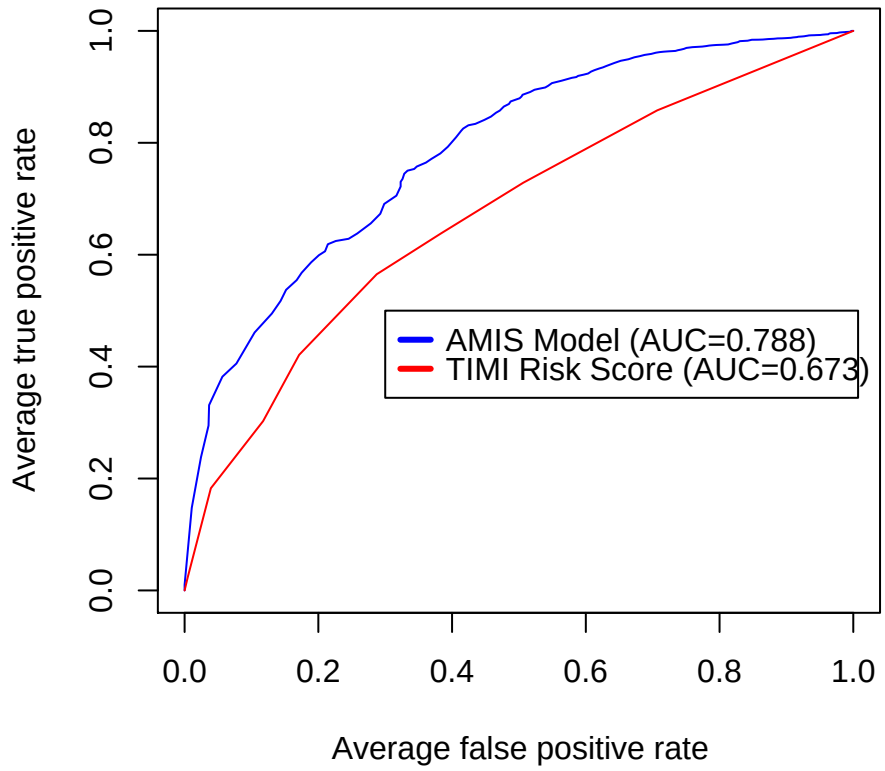


Figure 5.5: ROC Curve comparison of AMIS Model on a cohort receiving no treatment

Thrombolysis

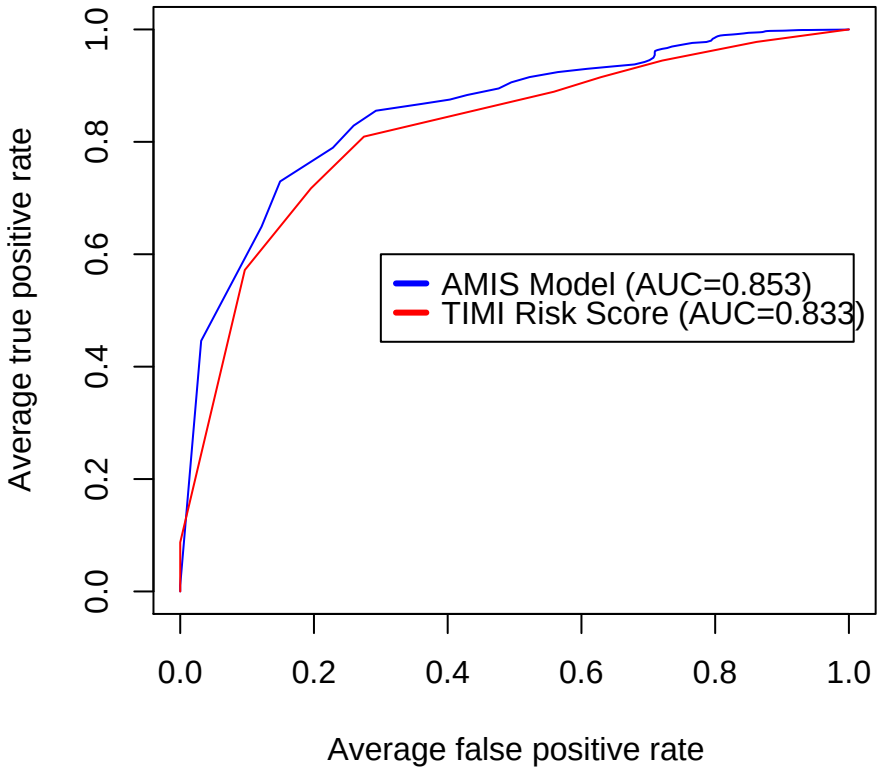


Figure 5.6: ROC Curve comparison of AMIS Model on a cohort receiving thrombolysis treatment

On a data cohort consisting of patients receiving thrombolysis therapy the AMIS Model achieves an AUC of 0.852. The TIMI Risk Score achieves an AUC of 0.833. Although the TIMI Risk Score was developed on this exact data cohort, it still does not beat the prediction capacity of the AMIS Model who was trained regardless of the therapy. See figure 5.6 for a comparison of the curves.

PCI

On a data cohort consisting of patients receiving PCI therapy the AMIS Model achieves an AUC of 0.885. This is the best result the AMIS Model achieved in all tests carried out. The TIMI Risk Score achieves an AUC of 0.783. See figure 5.7 for a comparison of the curve.

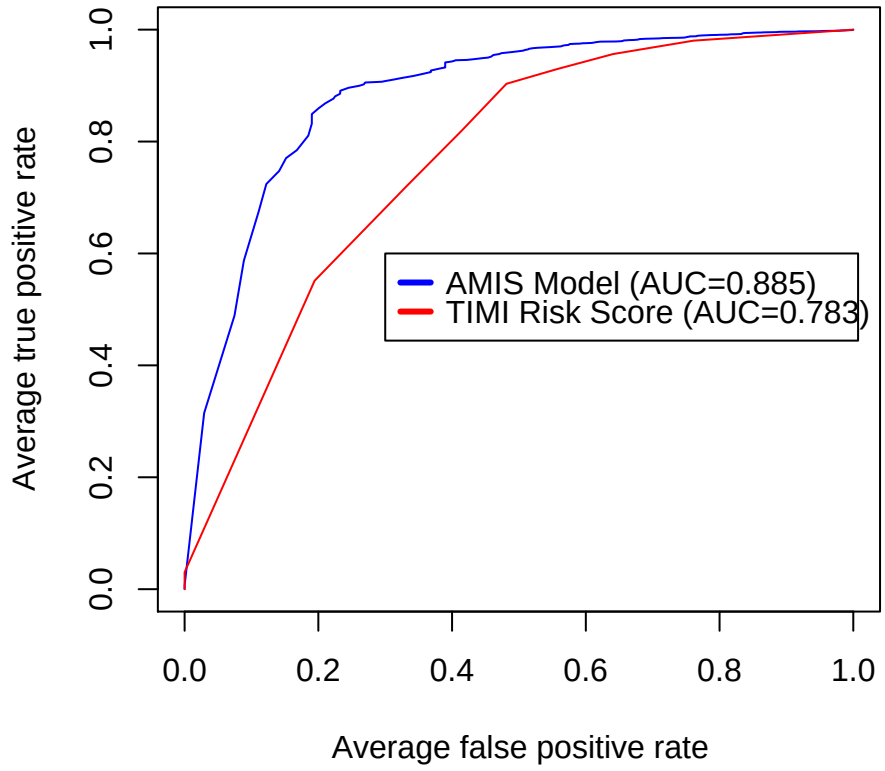


Figure 5.7: ROC Curve comparison of AMIS Model on a cohort receiving PCI treatment

5.2.2 Subgroup Analysis according to ST-Elevation

ST-Elevation

See figure 5.8 for a comparison of the ROC curves of the TIMI Risk Score and the AMIS Model. The AMIS Model achieves an AUC of 0.880. The TIMI Risk Score achieves an AUC of 0.816 on the ST-Elevation data cohort.

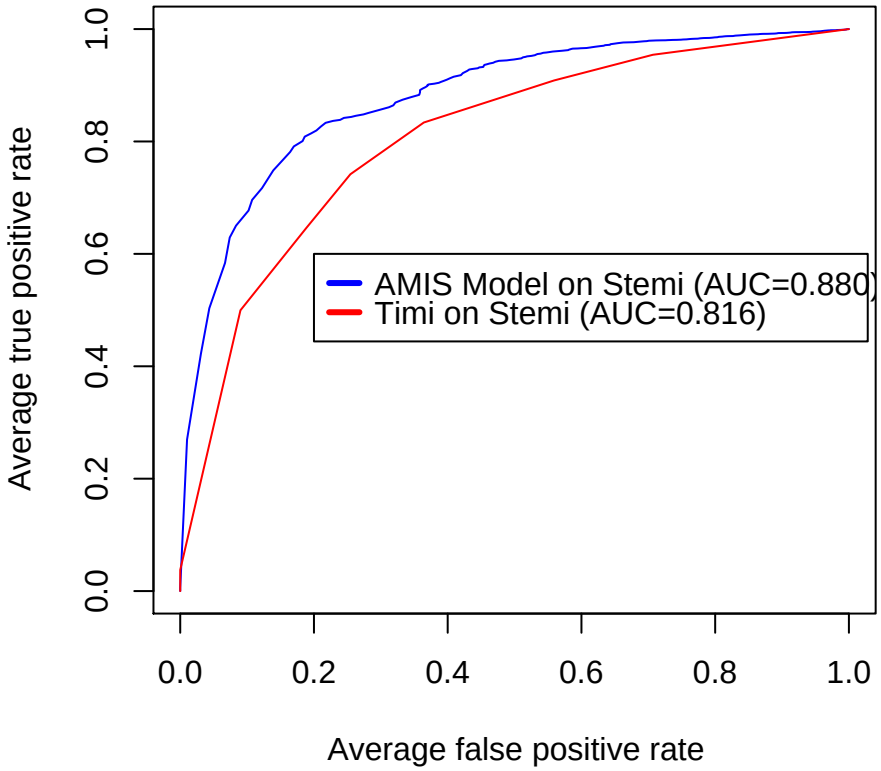


Figure 5.8: ROC Curve comparison of AMIS Model vs. TIMI Risk Score on a ST-Elevation cohort

non ST-Elevation

See figure 5.9 for a comparison of the ROC curves of the TIMI Risk Score and the AMIS Model. The AMIS Model achieves an AUC of 0.869. The TIMI Risk Score achieves an AUC of 0.794 on the non-ST-Elevation data cohort.

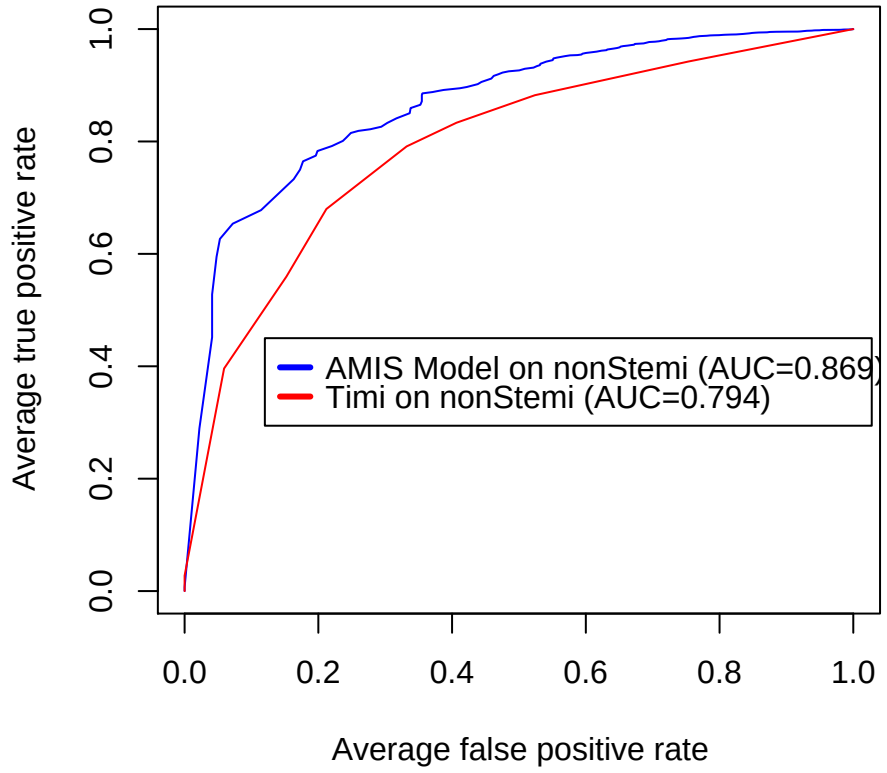


Figure 5.9: ROC Curve comparison of AMIS Model vs. TIMI Risk Score on a non-ST-Elevation cohort

5.3 Combined endpoint Death, Reinfarct, Shock, Stroke (MACE)

The combined endpoint death, reinfarct, shock and stroke was the secondary target. The only parameter changed when training the AMIS Model was the target. The rest was kept alike for reasons explained in chapter 4.

See figure 5.10 for a comparison of the ROC curves of the TIMI Risk Score and the AMIS Model. The AMIS Model achieves an AUC of 0.823. The TIMI Risk Score achieves an AUC of 0.757 with the combined endpoint death, reinfarct, shock and stroke.

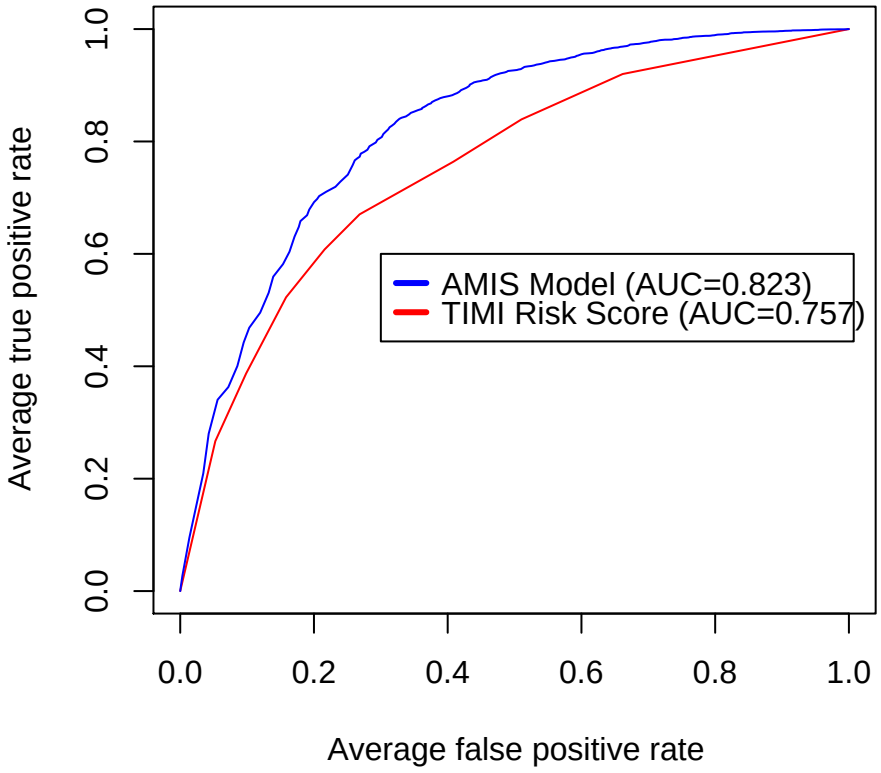


Figure 5.10: ROC Curve comparison of AMIS Model vs. TIMI Risk Score target mace

5.4 Analysis of negative Outcome

As mentioned before, the false positive error is critical for the model. In figure 5.11 the distribution of probabilities predicted by the AMIS Model for all records with negative outcome is visualized. The probabilities were categorized into 5 categories. Category 1 contains all records with a risk prediction <1% (very low risk), category 2 contains all records with a predicted risk >= 1 < 5% (low risk), category 3 contains all records with a risk prediction >= 5 < 15% (high risk), and finally category 4 contains all records with a predicted risk >= 15 (very high risk). Then all records with positive outcome were filtered out. The figure shows that 63.88 % of all records with negative outcome were predicted to be in the top category. Only 3.2 % of the records with negative outcome were attributed to the bottom category.

Value	Proportion	%	Count
0		3.2	18
1		13.17	74
2		19.75	111
3		63.88	359

Figure 5.11: Predicted Mortality Distribution by AMIS Model of records with negative outcome

5.5 Analysis of Probability Distribution

In figure 3.5 it was shown, that the mortality rates hardly change for the PCI and thrombolysis treatment groups until they reach the category 5 or 6 of the TIMI Risk Score. This indicates that the prediction capability of the TIMI Risk Score for these treatment groups is low when low risk patients are concerned. To test the AMIS Model for the equivalent characteristic, the predicted probabilities were categorized into the 6 categories shown in figure 5.12. All predictions above 50% were put in one bucket, because there were too few records to allow a statement in this area. Then the predicted mortality rate was compared to the actual mortality rate in each bucket. The AMIS Model shows a very even and accurate distribution.

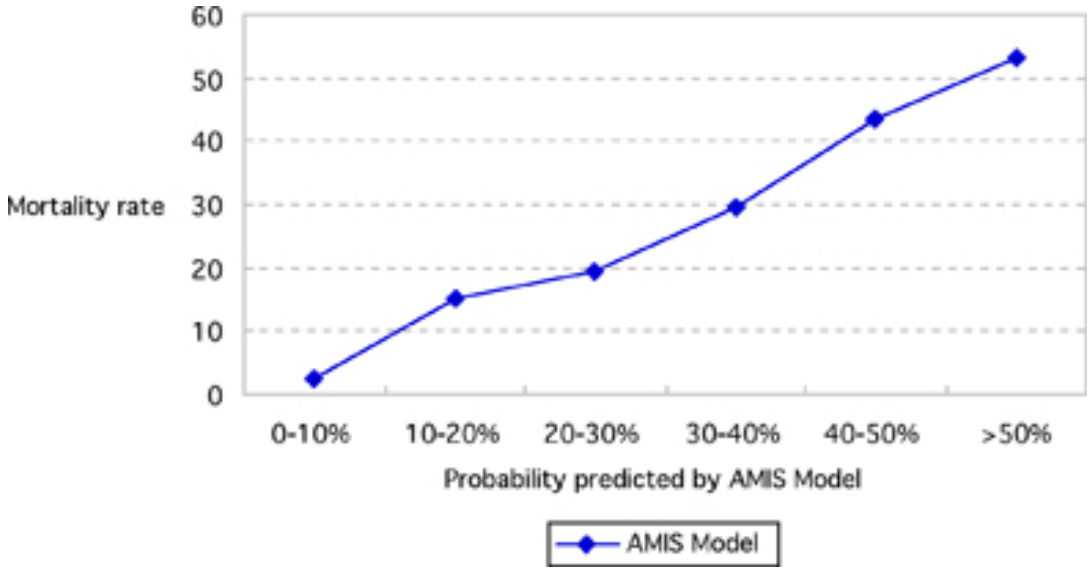


Figure 5.12: Predicted Probability Distribution by AMIS Model

6

Conclusions

After the application of a thorough preprocessing we found that the AMIS Plus registry is a very interesting ground for data mining methods. The quality as well as the quantity of the data is sufficient to carry out data mining successfully. During preprocessing it is necessary that all data alterations are discussed with a medical expert, as the field of application is highly specialized. The danger misinterpretation, as well as the danger of missing hidden influences is great. The heterogeneity turns out to be more interesting than complicating for the development.

It was shown that the AMIS Model based on an AODE algorithm has many advantages over the existing risk prediction method TIMI Risk Score, which was developed using a logistic regression. This was shown by comparison of the AUC values (0.875 vs. 0.802 on the full cohort) and the ROC curves (the ROC curve of the AMIS Model lies above the ROC curve of the TIMI Risk Score at any point on the curve).
It was shown that especially for patients with negative outcome the AMIS Model is a better predictor, placing over 63% in the highest risk category, and as few as 3.2% in the lowest. For low risk patients the AMIS Model has been shown to be superior to the TIMI Risk Score, displaying an even probability distribution over all risks.

On the treatment subgroup of PCI patients the AMIS Model showed the best prediction capability (AUC of 0.885). For this treatment group the mortality rates predicted by the TIMI Risk Score were shown to be the least applicable, as the TIMI Risk Score was developed on a data cohort of thrombolysis patients. The good performance of the AMIS Model in this section is especially gratifying because it represents the most relevant treatment group at present.

The combined endpoint death, reinfarct, shock and stroke proved to be a more difficult target to predict. It might be an interesting project in future, when more records are available. Momentarily the AMIS Model achieves an acceptable AUC of 0.824 when trained for the target *mace*.

7

Future Work

It is definitely desirable to develop more prediction models on the AMIS Plus registry in future. Because the whole preprocessing process has already been performed and saved in an easy accessible way, further work is facilitated.
The AMIS Plus steering adapted a welcoming attitude towards new projects.

Validation of AMIS Model Newly collected records could be used to validate the AMIS Model. Even more interesting would be the use of an outside database. The AMIS Steering committee has pointed out the possibility of attaining a database from a French hospital for this purpose.

Development of Mortality Risk Prediction As the therapies and mortality rates change over time and more records are collected, it would be desirable to develop a new model every few years. Soon values for the newly specified variables about the patient follow up will be available. This would enable a mortality risk prediction for in-hospital mortality, as well as 6 months and a year. It would be very interesting to identify which variables play a role.

Development of a MACE Risk Prediction For the same reasons as mentioned above, it would be interesting to carry on developing a new *mace* predictor.

Research data mining More challenging but also very interesting would be to apply the data mining methods not to predict, but to identify factors and combination of factors that lead to death. This work would have to be tackled in close collaboration with medical personal.

Predict the best therapy The AMIS Plus group would be very interested in a predictor, predicting which therapy would achieve the best outcome. See the paper by Brohpy and Lawrence (James M. Brophy & Lawrence Joseph, 1995) for a possible approach.

A

Appendix: Data Preprocessing

A.1 Data Cleaning Steps

In the following all steps of the data cleaning process are listed. Please note that original Clementine data stream (cleaning.str) is available on the CD accompanying the diploma thesis.

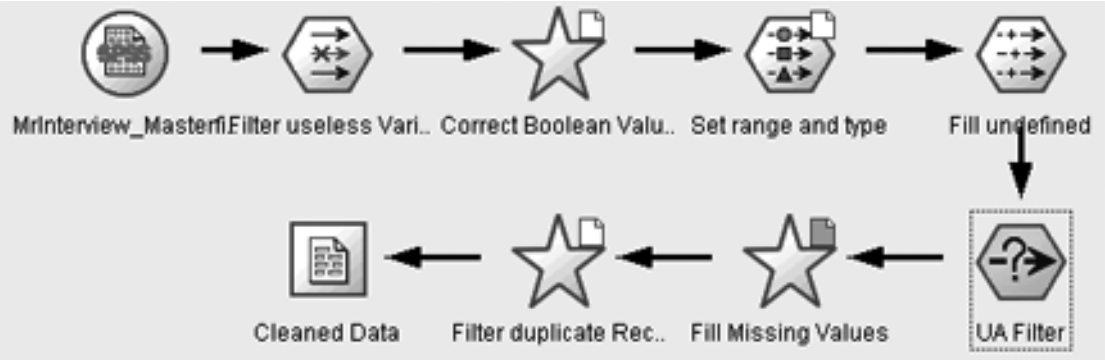


Figure A.1: Graphic from data stream of the data cleaning steps from Clementine workbench

1. The variables `basicins`[1], `othsymp`, `othrythm`, `cm0` and `deathco`[2] were filtered out.
2. Boolean variables of the following variables were corrected. With the variables `transfer`, `sympadm`, `pain`, `dyspnea`, `resusci`, `rgaspir`, `rgclop`, `rganticg`, `rgbbloc`, `rgaceinh`, `rgangioa`, `rgcabloc`, `rgnitrat`, `rgdigoxi`, `rgdiuret`, `rgliplow`, `comorbid`, `histfam`, `histhta`, `histhtatr`, `histhlip`, `histhliptr`, `histdiab`, `histdiabtre`, `thrmblys`, `pci`, `pcistent`, `pcides`, `pciams`, `gvaspir`, `gvclop`, `gvgp`, `gvsthepa`, `gvlohepa`, `gvbbloc`, `gvaceinh`, `gvangioa`, `gvcabloc`, `gvnitrat`, `gvliplow`, `gwasopr`, `gvnesir`, `gvdiuret`,

[1] The old AMIS Plus questionnaire used an ambiguous wording, when asking for the value of basic insurance. The variable is therefore useless. In the last questionnaire update the wording was changed.

[2] The mentioned variables are all strings and not usable for the intended purpose of Data Mining.

gvinsulin, compl, theint, echox, stresisch, intcare, intcare, imcare, alive, thaspir, thticlo, thhep, thanticg, thbbloc, thaceinh, thaantag, thcabloc, thnitrat, thdigoxi, thdiuret, thliplowst, thliplowez, thliplowoth, thliplow, thamiod, thother, thordiab, thinsulin, thantidep and thsedat the boolean values had to be recoded. Instead of the boolen 1/0 values, they all had the values 1/2/3 (1=yes, 2=no, 3=missing). In a first step all values equal 2 were replaced with a 0. In a second step all values equal 3 were replaced with missing.

3. The range and type for all variables were defined. Please check the table below for the exact specification of all bedside variables. To see the definition of all 210 variables please check the above mentioned Clementine stream.

Table A.1: Ranges and types of variables in AMIS Plus database

Ranges and types of variables			
Variable	Type	Range	Description
birthdat	date	1900-01-01 - ...	Date of Birth
sex	set	1,2	Sex, 1= Male, 2 = Female
weight	numeric	30 - ...	Weight of Patient at admission
height	numeric	120 - ...	Height of Patient
admisdat	date	1997-01-01 - ...	Date of admission
admisti	time	00:00:00 - 23:59:59	Time of admission
firstdat	date	2003-03-14	Date of Hospitalization, if different from Date of admission
firstti	time	00:00:00 - 23:59:59	Time of Hospitalization, if different from Time of admission
transfer	boolean	1/0	Has the patient been transfered from one hospital to another
sympadm	boolean	1/0	Symptoms at admission
pain	boolean	1/0	Pain at admission
dyspnea	boolean	1/0	Dyspnea at admission
onsetdat	date	1996-12-29 - ...	Date of symptom onset
onsetti	time	00:00:00 - 23:59:59	Time of symptom onset
killip	ordered set	1,2,3,4	Killip Classification
stelev	boolean	1/0	ECG-indication of ST-Elevation
Continued ...			

Variable	Type	Range	Description
qwave	boolean	1/0	ECG-indication of Q-Wave
stdepres	boolean	1/0	ECG-indication of ST-Depression
twavec	boolean	1/0	ECG-indication of T-Wave Changes
lbbblck	boolean	1/0	ECG-indication of Left Bundle Branch Block
rbbblck	boolean	1/0	ECG-indication of Right Bundle Branch Block
nochange	boolean	1/0	No ECG-indication or changes
ecgoth	boolean	1/0	Other ECG-indications
locisant	boolean	1/0	Anterior Location of ST-Elevation
locisinf	boolean	1/0	Inferior Location of ST-Elevation
locispost	boolean	1/0	Posterior Location of ST-Elevation
locisundet	boolean	1/0	Undetermined Location of ST-Elevation
locisoth	boolean	1/0	Other Location of ST-Elevation
locis0	boolean	1/0	No answer to Location of ST-Elevation
hrtrythm	set	1,2,3,4,5,6	Heart rythm
systbp	numeric	0-...	Systolic Blood Pressure
diastbb	numeric	0-...	Diastolic Blood Pressure
hrtrate	numeric	0-....	Heartrate
resusci	boolean	1/0	Resusciation prior to arrival at hospital
cprarr	boolean	1/0	Cardio-pulmonary Resuscitation
defibarr	boolean	1/0	Cardioversion / defibrillation
resusc0	boolean	1/0	No answer towards resuscitation
rgaspir	boolean	1/0	Aspirin, ASA (regular)
rgclop	boolean	1/0	Clopidogrel (regular)
rganticg	boolean	1/0	Oral anticoagulant (regular)
rgbbloc	boolean	1/0	Beta-blocker (regular)
rgaceinh	boolean	1/0	ACE inhibitor (regular)
rgangioa	boolean	1/0	Angiotensin II receptor antagonist (regular)
rgcabloc	boolean	1/0	Ca-channel blocker (regular)
rgnitrat	boolean	1/0	Nicorandil, molsidomine and/or long-acting nitrates (regular)
rgdigoxi	boolean	1/0	Digoxin (regular)
rgdiuret	boolean	1/0	Diuretic (regular)

Continued . . .

Variable	Type	Range	Description
rgliplow	boolean	1/0	Lipid-lowering Drug (regular)
comorbid	boolean	1/0	Comorbidity
cmphmi	boolean	1/0	Comorbidity: Past History of myocardial infarction
cmcardin	boolean	1/0	Comorbidity: Cardiac insufficiency
cmpvdis	boolean	1/0	Comorbidity: Peripheral vascular disease
cmcevdis	boolean	1/0	Comorbidity: Cerebrovascular disease
cmhemipl	boolean	1/0	Comorbidity: Hemiplegia
cmdement	boolean	1/0	Comorbidity: Dementia
cmchlung	boolean	1/0	Comorbidity: Chronic lung disease
cmconntd	boolean	1/0	Comorbidity: Connective tissue disease
cmpepdis	boolean	1/0	Comorbidity: Peptic ulcer disease
cmdiabet	boolean	1/0	Comorbidity: Diabetes
cmdmtod	boolean	1/0	Comorbidity: Diabetes with target organ damage
cmlivmil	boolean	1/0	Comorbidity: Mild liver disease
cmlivsev	boolean	1/0	Comorbidity: Moderate to severe liver disease
cmrenald	boolean	1/0	Comorbidity: Moderate to severe renal disease
cmmalign	boolean	1/0	Comorbidity: Malignant neoplasm
cmleukem	boolean	1/0	Comorbidity: Leukemia
cmlympho	boolean	1/0	Comorbidity: Lymphoma
cmmetstu	boolean	1/0	Comorbidity: Metastatic solid tumor
cmaidsc	boolean	1/0	Comorbidity: AIDS (stage C)
histap	boolean	1/0	History of previous stable angina
histpci	boolean	1/0	History of previous angioplasty (PCI)
histami	boolean	1/0	History of previous AMI
histbypass	boolean	1/0	History of previous coronary artery bypass grafting
hist0	boolean	1/0	No answer concerning history
histpmi	boolean	1/0	History of MI or stable angina
histfam	boolean	1/0	Family history (in first degree relative ¡60y)
histhta	boolean	1/0	History of arterial hypertension
histhtatr	boolean	1/0	History of arterial hypertension: Treatment
Continued . . .			

Variable	Type	Range	Description
histhlip	boolean	1/ 0	History of dyslipidemia
histhliptr	boolean	1/0	History of dyslipidemia: Treatment
histdiab	boolean	1/0	History of diabetes mellitus
histdiabtre	boolean	1/0	History of diabetes mellitus: treated or untreated
histdiabtr	boolean	1/0	History of diabetes mellitus: Treatment
histsmok	boolean	1/0	History of Smoking
The End			

4. All undefined values were replaced by missing values.

5. All records with `diagall = 3` were discarded, as this means that they were diagnosed with Unstable Angina and not Acute Myocardial Infarction and are therefore not to be included in our trial.

6. The ECG variables `qwave`, `stdepres`, `twavec`, `lbbblck`, `rbbblck` and `nochange` were filled with a 0, if their value was missing, but the variable `stelev` was specified. This was carried out on the assumption, that when filling in the questionnaire the personnel only ticked what concurred and neglected to tick the negative case.

7. The comorbidity variables `cmphmi`, `cmcardin`, `cmpvdis`, `cmcevdis`, `cmhemipl`, `cmchlung`, `cmconntd`, `cmpepdis`, `cmdiabet`,`cmdmtod`,`cmlivmil`,`cmlivsev`,`cmrenald`, `cmmalign`, `cmleukem`, `cmlympho`, `cmmetstu`, `cmaids` and `cmdement` were filled with a 0, if the value was missing, and the variable `comorbid` was specified as 0. The variable `comorbid` indicates whether any comorbidities were present. If this value is set to 0, indication that this is not the case, the missing values of the comorbidities variables may be filled in with 0.

8. The variables `pci` was filled with a 0, if the value was missing.

9. All records missing the values for the variables `birthdat`, `admisdat` and `alive` were discarded. If the outcome (alive) is not available, the record is useless. The same goes for the birth date and the admission date.

10. All records were sorted in ascending order according to the variable `birthdat`.

11. If two consecutive records had the same values for the variables `birthdat`, `admisdat` one of the records was discarded.

A.2 Data Preparation Steps

In the following all steps of the data preparation process are listed. The scripting language used to specify and carry out the commands is called CLEM. Please note that original Clementine data stream (preparation.str) is available on the CD accompanying the diploma thesis.

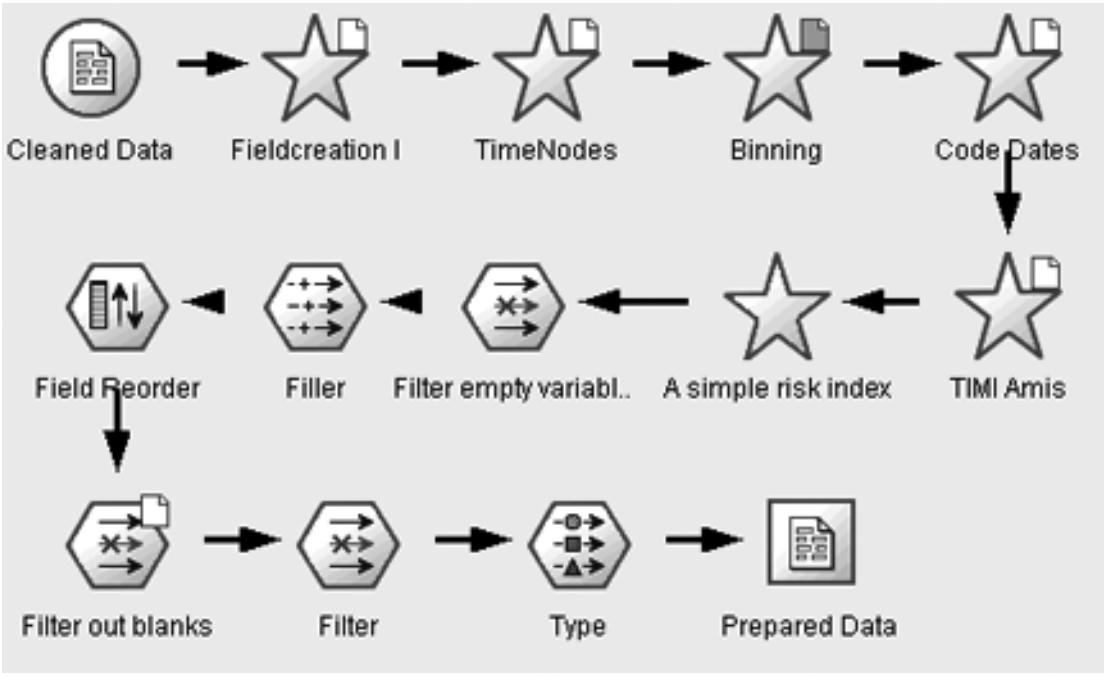


Figure A.2: Graphic from data stream of the data preparation steps from Clementine workbench

1. The field `age` was created - `admisdat - birthdat`
2. The field `stemi`[3] was created - `if (stelev = 1 or lbbblck = 0 then 1 elseif (stelev = 0 and lbbblck = 0) then 0 else undef endif`
3. The field `diabetes`[4] was created - `if (cmdiabet = 1 or cmdmtod = 1 or histdiab = 1) then 1 elseif (cmdiabet = 0 and histdiab = 0) then 0 else undef endif`

[3] `stemi` stands for ST-Elevation. The field has been included in the last update of the AMIS Plus questionnaire. There are virtually no records with the value specified in this very important field

[4] Whether a patient suffers from diabetes is unfortunately specified in three different fields. `cmdiabet` specifies whether the patient suffers from diabetes as a comorbidity. `histdiab` specifies whether the patient has a history of diabetes. `cmdmtod` stands for a different form of diabetes.

4. The field `bmi` was created - `(weight / (height*height)) * 10000`
5. The field `drss` was created - `if alive = 0 then 1 elseif complreinf = 1 then 1 elseif complcvins = 1 then 1 elseif complshock = 1 then 1 elseif alive = 1 and complreinf = 0 and complcvins = 0 and compleshock = 0 then 0 else undef endif`
6. The auxiliary variable `needle_indicator` was created, specifying whether any kind of treatment has been carried out - `if @NULL(firstdat) or @NULL(firstti) then 0 else 1 endif`
7. The field `id` was created, for internal identification purposes - `if @OFFSET(id,1)=undef then 1 else @OFFSET(id,1) + 1 endif`
8. The auxiliary variable `needle` was created, indicating the first time any treatment was carried out in hospital. This variable can only be filled, if either PCI or Thrombolysis treatment was conducted.
   `if needle_indicator = 1 then (datetime_in_seconds(firstdat) + time_in_secs(firstti)) elseif thrmblys = 0 and pci = 0 then undef elseif thrmblys = 1 and pci = 1 then min((datetime_in_seconds(pcidat)+ time_in_secs(pciti)),(datetime_in_seconds(thrmbdat) + time_in_secs(thrmbti))) elseif pci = 1 then (datetime_in_seconds(pcidat)+ time_in_secs(pciti)) elseif thrmblys = 1 then (datetime_in_seconds(thrmbdat) + time_in_secs(thrmbti)) else undef endif`
9. The field `doorNeedle` was calculated, indication the time elapsed from admission to hospital until treatment was carried out
   `(needle - (datetime_in_seconds(admisdat)+time_in_secs(admisti)))/60/60`
10. The field `painDoor` was calculated, indicating how much time elapsed from symptom onset until admission at hospital
    `((datetime_in_seconds(admisdat) + time_in_secs(admisti))-(time_in_secs(onsetti) + datetime_in_seconds(onsetdat)))/60/60`
11. The field `painNeedle` was calculated, indicating the time elapsed between symptom onset and treatment - `(needle - datetime_in_seconds(onsetdat) + time_in_secs(onsetti)))/ 60 / 60`
12. Set ranges of `doorNeedle`, `painDoor` and `painNeedle` to positive and filter all negative values as well as those ranged above 100[5]

[5] In the date and time variables there is unfortunately a lot of noise

13. The variable `bmi` was categorized into 4 medically predefined categories creating the new field `bmi_BIN - if bmi<= 18.5 then 0 elseif bmi<= 24.9 then 1 elseif bmi <= 29.9 then 2 elseif bmi > 29.9 then 3 else undef endif`

14. The variable `maxcpk` was binned into the by Dr. Kurz predefined categories creating the new field `maxcpk_bin - if maxcpk < 2000 then 0 elseif maxcpk < 4000 then 1 elseif maxcpk < 6000 then 2 elseif maxcpk < 8000 then 3 elseif maxcpk >= 8000 then 4 else undef endif`

15. The variable `maxcpkmb` was binned into the by Dr. Kurz predefined categories creating the new field `maxcpkmb_bin - if maxcpkmb < 200 then 0 elseif maxcpkmb < 400 then 1 elseif maxcpkmb < 600 then 2 elseif maxcpkmb < 800 then 3 elseif maxcpkmb >= 800 then 4 else undef endif`

16. The variable `tropi` was binned into the by Dr. Kurz predefined categories creating the new field `tropi_bin - if maxcpkmb < 200 then 0 elseif maxcpkmb < 400 then 1 elseif maxcpkmb < 600 then 2 elseif maxcpkmb < 800 then 3 elseif maxcpkmb >= 800 then 4 else undef endif`

17. The variable `tropitot` was binned into the by Dr. Kurz predefined categories creating the new field `tropitot_bin - if tropitot < 100 then 0 elseif tropitot < 200 then 1 elseif tropitot < 300 then 2 elseif tropitot < 400 then 3 elseif tropitot >= 400 then 4 else undef endif`

18. The variable `tropt` was binned into the by Dr. Kurz predefined categories creating the new field `tropt_bin - if tropt < 40 then 0 elseif tropt < 80 then 1 elseif tropt < 120 then 2 elseif tropt < 160 then 3 elseif tropt >= 160 then 4 else undef endif`

19. The variable `cholestl` was binned into the by Dr. Kurz predefined categories creating the new field `cholestl_bin - if cholestl < 2 then 0 elseif cholestl < 4 then 1 elseif cholestl < 6 then 2 elseif cholestl < 8 then 3 elseif cholestl >= 8 then 4 else undef endif`

20. The variable `hdl` was binned into the by Dr. Kurz predefined categories creating the new field `hdl_bin - if hdl < 0.4 then 0 elseif hdl < 0.8 then 1 elseif hdl < 1.2 then 2 elseif hdl < 1.6 then 3 elseif hdl >= 1.6 then 4 else undef endif`

21. The hour was extracted from the variable `admisti` creating the new field `admis_hour - datetime_hour(admisti)`

22. The month was extracted from the variable `admisdat` creating the new field `admis_month` - `datetime_month(admisdat)`

23. The weekday was extracted from the variable `admisday` creating the new field `admis_day` - `datetime_weekday(admisdat)`

24. The hour was extracted from the variable `onsetti` creating the new field `onset_hour` - `datetime_hour(onsetti)`

25. The month was extracted from the variable `onsetdat` creating the new field `onset_month` - `datetime_month(onsetdat)`

26. The year was extracted from the variable `admisyear` creating the new field `admis_year` - `datetime_year(admisdat)`

27. The value for the TIMI Risk Score for STEMI was calculated for each record and stored in the variable `TIMI` with the following steps:

    (a) variable `AnteSte` was created, and filled with a 0 for 58% of the records and with a 1 for the remaining 42% with a random distribution[6]

    (b) The field `TIMI` was created and filled with 0

    (c) Points for the age of the patient were added with a Clementine Filler Node
    Condition: `65<= age and age <= 74`
    Replace with: `TIMI + 2`

    (d) Points for the history of the patient were added with a Clementine Filler Node
    Condition: `diabetes = 1 or histpmi = 1or histhta = 1`
    Replace with: `TIMI + 1`

    (e) Points for the age of the patient were added with a Clementine Filler Node
    Condition: `age >= 75`
    Replace with: `TIMI + 3`

    (f) Points for the systolic blood pressure of the patient were added with a Clementine Filler Node
    Condition: `systbp < 100`
    Replace with: `TIMI + 3`

[6]The information about Anterior ST-Elevation is needed to calculate the TIMI Risk Score. In order to not influence the risk distribution, the variable was calculated and distributed randomly, according to the percentage observed in other data cohorts.

(g) Points for the heartrate of the patient were added with a Clementine Filler Node
Condition: `hrtrate > 100`
Replace with: `TIMI + 2`

(h) Points for the killip classification of the patient were added with a Clementine Filler Node
Condition: `killip > 1`
Replace with: `TIMI + 2`

(i) Points for the weight of the patient were added with a Clementine Filler Node
Condition: `weight < 67`
Replace with: `TIMI + 1`

(j) Points for the left bundle branch block and the randomly generated variable of anterior st-elevation of the patient were added with a Clementine Filler Node
Condition: `lbbblck = 1 or AnteSte = 1`
Replace with: `TIMI + 1`

(k) Points for the the time elapsed between symptom onset and treatment added with a Clementine Filler Node
Condition: `painNeedle > 4 or @NULL(painNeedle)`
Replace with: `TIMI + 1`

28. The Simple Risk Score was calculated for all records and stored in the variable `simpRisk - (hrtrate * (age/10)*(age/10))/systbp`

B

Appendix: AMIS Plus Documents

B.1 The AMIS Plus Questionnaire

University of Zurich
Institute of Social and
Preventive Medicine

AMIS Plus Questionnaire

02.2005

AMIS Acute Myocardial Infarction in Switzerland

AMIS DATA ENTRY IDENTIFICATION

Hospital ☐☐☐☐

Patient ID number /VO Set Nr./Code ☐☐☐☐☐☐☐/☐☐☐☐☐☐☐/☐

Physician ID ☐☐☐☐☐☐☐☐

Date of data entry Day ☐☐ Month ☐☐ Year ☐☐☐☐

PATIENT AT ADMISSION

Date of birth Day ☐☐ Month ☐☐ Year ☐☐☐☐

Gender ❍ Male ❍ Female

Weight *(eg.: 68.5 kg is rounded up to* 0 6 9 *kg)* ☐☐☐ kg

Height ☐☐☐ cm

Admission date to this hospital Day ☐☐ Month ☐☐ Year ☐☐☐☐

Admission time hh ☐☐ mn ☐☐

Time of first medical contact leading to hospitalization
(If available and different from admission time) Day ☐☐ Month ☐☐ Year ☐☐☐☐

hh ☐☐ mn ☐☐

Insurance coverage ❍ Basic ❍ Semiprivate/ Private

Transfer *(Was the patient transferred from another hospital?)* ❍ Yes ❍ No

IF YES ➔ Hospital name? _____

»Condition

Symptoms at admission ❍ Typical ❍ Atypical

- Pain ❍ Yes ❍ No ❍ Unknown
- Dyspnea ❍ Yes ❍ No ❍ Unknown
- Other ❍ Yes , _____

Symptom onset date Day ☐☐ Month ☐☐ Year ☐☐☐☐

Symptom onset time hh ☐☐ mn ☐☐

Killip classification
*(choose **only one** answer)*

❍ Class I = no clinical signs of heart failure (no rales, no S3)
❍ Class II = crackles, S3 gallop and elevated jugular venous pressure
❍ Class III = frank pulmonary edema
❍ Class IV = cardiogenic shock

ECG on admission
(several answers possible)

❍ ST-segment elevations
❍ Q-waves
❍ ST-segment depressions
❍ T-wave changes
❍ Left bundle branch block
❍ Right bundle branch block
❍ No changes/normal
❍ Other

Location of ischemic region

❍ Anterior
❍ Inferior
❍ Posterior
❍ Undetermined
❍ Other

Heart rhythm
*(choose **only one** answer)*

❍ Sinus rhythm
❍ Atrial fibrillation
❍ Advanced AV-block (II/III)
❍ Wide QRS complex tachycardia
❍ Paced rhythm
❍ Other?_____

AMIS Plus Data Center
Hirschengraben 84
CH-8001 Zürich

Tel: +41 (0)44-634 48 30
Fax: +41 (0)44-634 49 86

E-Mail: amis@ifspm.unizh.ch
www.amis-plus.ch

»Vital signs

Systolic blood pressure ☐☐☐ mmHg

Diastolic blood pressure ☐☐☐ mmHg

Heart rate ☐☐☐ bpm

REPORTED HISTORY

Resuscitation *prior to arrival at hospital ?* ❍ **Yes** ❍ **No**

IF YES➔ Kinds of resuscitation
- ❍ Cardiopulmonary (mechanical)
- ❍ Cardioversion/defibrillation

Regular medication

Has this patient been taking any of the following medication daily or regularly ?
(check each medication)

Aspirin, ASA	❍ Yes	❍ No	❍ Unknown
Clopidogrel	❍ Yes	❍ No	❍ Unknown
Oral anticoagulant	❍ Yes	❍ No	❍ Unknown
Beta-blocker	❍ Yes	❍ No	❍ Unknown
ACE inhibitor	❍ Yes	❍ No	❍ Unknown
Angiotensin II receptor antagonist	❍ Yes	❍ No	❍ Unknown
Ca-channel blocker	❍ Yes	❍ No	❍ Unknown
Nicorandil, molsidomine and/or long-acting nitrates	❍ Yes	❍ No	❍ Unknown
Digoxin	❍ Yes	❍ No	❍ Unknown
Diuretic	❍ Yes	❍ No	❍ Unknown
Lipid-lowering drug	❍ Yes	❍ No	❍ Unknown

Comorbidities (Charlson Index) ❍ **Yes** ❍ **No** ❍ **Unknown**

IF YES➔ Kinds of comorbidities:
- ❍ Past history of myocardial infarction
- ❍ Cardiac insufficiency (NYHA III+IV)
- ❍ Peripheral vascular disease (ST III+IV)
- ❍ Cerebrovascular disease
- ❍ Hemiplegia
- ❍ Dementia
- ❍ Chronic lung disease
- ❍ Connective tissue disease
- ❍ Peptic ulcer disease
- ❍ Diabetes
- ❍ Diabetes with target organ damage
- ❍ Mild liver disease
- ❍ Moderate to severe liver disease
- ❍ Moderate to severe renal disease
- ❍ Malignant neoplasm
- ❍ Leukemia
- ❍ Lymphoma
- ❍ Metastatic solid tumor
- ❍ AIDS (stage C)

»Past history

Ischemic heart diseases
- ❍ Previous stable angina
- ❍ Previous angioplasty (PCI)
- ❍ Previous AMI
- ❍ Previous coronary artery bypass grafting

Risk factors *(check each risk factor)*

• Family history *(in first degree relative <60y)*	❍ Yes	❍ No	❍ Unknown
• History of arterial hypertension	❍ Yes	❍ No	❍ Unknown ❍ Treated ❍ Untreated
• History of dyslipidemia	❍ Yes	❍ No	❍ Unknown ❍ Treated ❍ Untreated
• History of diabetes mellitus	❍ Yes	❍ No	❍ Unknown ❍ Oral treated ❍ Insulin treated ❍ Untreated

- Smoking
  - ❍ Never smoker
  - ❍ Former smoker
  - ❍ Current smoker
  - ❍ Unknown

IMMEDIATE THERAPY

Initial therapeutic strategy

- ❍ **Primary PCI**
- ❍ **PCI facilitated (before catheter laboratory)**
- ❍ **Thrombolysis**
- ❍ **Primary conservative treatment, angiography planned**
- ❍ **Conservative therapy, elective angiography if problems**
- ❍ **Primary palliative/symptomatic therapy**

Thrombolysis ❍ **Yes** ❍ **No**

IF YES ➔ Location of thrombolysis ❍ In your hospital ❍ In another hospital:_____

IF IN YOUR HOSPITAL ➔

- Date of thrombolysis: Day ☐☐ Month ☐☐ Year ☐☐☐☐
- Starting time of thrombolysis: hh ☐☐ mn ☐☐

IF NO ➔ **Reason for denial of thrombolysis**

*(choose **only one** of the following answers, the **most important** reason)*

- ❍ PCI preferred
- ❍ Not indicated ➔
  - ❍ Too late
  - ❍ Diagnosis uncertain
  - ❍ ECG-criteria not fulfilled
- ❍ Contraindicated ➔
  - ❍ Active bleeding
  - ❍ Non-compressible puncture site
  - ❍ Recent surgery
  - ❍ Uncontrolled hypertension
  - ❍ Recent cerebral event
  - ❍ Oral anticoagulation
- ❍ Refused
- ❍ Unknown or others

Percutaneous Coronary Intervention (PCI) ❍ **Yes** ❍ **No**

IF YES ➔ Location of PCI procedure ❍ In your hospital ❍ In another hospital:_____

- Date of PCI: Day ☐☐ Month ☐☐ Year ☐☐☐☐
- Starting time of PCI *(Time of first automatic blood pressure measurement in the heart catheter lab)*: hh ☐☐ mn ☐☐
- Reason for performing PCI *(check **only one**)*
  - ❍ First strategy (instead of thrombolysis)
  - ❍ Rescue after thrombolysis
  - ❍ Rescue after primary conservative therapy
  - ❍ Elective
- **TIMI flow at the end of PCI** *(if available)* ❍ 0 ❍ I ❍ II ❍ III
- **Angiographic findings**
  - ❍ One vessel
  - ❍ Two vessels
  - ❍ Three vessels
  - ❍ Left main
  - ❍ No angiographic abnormalities
- **Left ventricular ejection fraction**
  - ❍ <35%
  - ❍ 35-50%
  - ❍ > 50%
  - ❍ unknown
- **Vessel treated**
  - ❍ Left main
  - ❍ Left anterior descending (or one of its branches)
  - ❍ Left circumflex artery (or one of its branches)
  - ❍ Right coronary artery (or one of its branches)
- **PCI with stent?** ❍ **Yes** ❍ **No**
  - **IF YES** ➔ Drug eluting stent/s? ❍ Yes ❍ No
  - ➔ absorbable stent/s? ❍ Yes ❍ No

Medication for immediate therapy *(within 24 hours)*

(check each medication)

Aspirin, ASA	❍ Yes ❍ No
Clopidogrel	❍ Yes ❍ No
GP IIb/IIIa antagonists	❍ Yes ❍ No
Unfractionated heparin	❍ Yes ❍ No
Low molecular weight heparin	❍ Yes ❍ No
Beta-blocker	❍ Yes ❍ No
ACE inhibitor	❍ Yes ❍ No
Angiotensin II receptor antagonist	❍ Yes ❍ No
Ca-channel blocker	❍ Yes ❍ No
Nitrate	❍ Yes ❍ No
Lipid lowering drug	❍ Yes ❍ No
Vasopressors (dopamine, dobutamine and others)	❍ Yes ❍ No
Nesiritide	❍ Yes ❍ No
Diuretic	❍ Yes ❍ No
Insulin	❍ Yes ❍ No

LABORATORY PARAMETERS

Glycemia *(on admission)*	☐☐ mmol/l
Creatinine *(on admission)*	☐☐☐ µmol/l
Maximum CK *(max. level during hospitalization)*	☐☐☐☐☐ IU/l
Maximum CK-MB *(max. level during hospitalization)*	☐☐☐☐ IU/l
Troponin I (free) *(max. level during hospitalization)*	☐☐☐.☐ µg/l
Troponin T *(max. level during hospitalization)*	☐☐☐.☐ µg/l
Troponin I (total) *(max. level during hospitalization)*	☐☐☐.☐ µg/l
Total Cholesterol *(within 24 hours of chest pain onset)*	☐☐.☐☐ mmol/l
HDL Cholesterol *(within 24 hours of chest pain onset)*	☐☐.☐☐ mmol/l
CRP *(on admission, if available)*	☐☐.☐☐ mg/l
BNP *(if available)*	☐☐☐☐ pg/ml
NT-ProBNP *(max. level during hospitalization)*	☐☐☐☐ pg/ml

HOSPITALIZATION

Complications ❍ **Yes** ❍ **No**

IF YES ➔ Kinds of complications

(check each complication)

- ❍ AV block (needing pacing)
- ❍ Cardiogenic shock
- ❍ Recurrent ischemic episodes (Post-infarction angina)
- ❍ Infarction in patient admitted for unstable angina
- ❍ Re-infarction
- ❍ Cerebrovascular event
- ❍ Major bleeding (requiring special therapy)
- ❍ Acute renal failure (needing treatment)
- ❍ Sepsis/ SIRS/ Multiorgan failure
- ❍ Atrial fibrillation at discharge
- ❍ New heart failure (Killip III-IV)
- ❍ Other

Therapeutic interventions ❍ **Yes** ❍ **No**

IF YES ➔ Kinds of therapeutic interventions

(check each intervention)

- ❍ Temporary pacing
- ❍ Heart massage (CPR)
- ❍ Mechanical circulatory support (IABP or others)
- ❍ Invasive mechanical ventilation (intubation)
- ❍ Non-invasive mechanical ventilation (mask)
- ❍ Defibrillation/cardioversion
- ❍ Permanent pacemaker implantation
- ❍ Permanent defibrillator implantation (performed/ planned)

Performed diagnostic / therapeutic procedures

(check each procedure)

- **Echocardiography** ❍ Yes ❍ No

  IF YES ➔

  - Date of first echocardiography — Day ☐☐ Month ☐☐ Year ☐☐☐☐
  - Time — hh ☐☐ mn ☐☐

  - **Left ventricular ejection fraction**
    - ❍ <30%
    - ❍ 30-40%
    - ❍ >40%

- **Stress test** *(including stress echo and/or isotops)* ❍ Yes ❍ No ❍ Planned

  IF DONE ➔ Ischemia ❍ Yes ❍ No

- **Coronary angiography** ❍ Yes ❍ No ❍ Planned
- **Coronary artery bypass grafting (CABG)** ❍ Yes ❍ No ❍ Planned
- **Electrophysiology study** (invasive) ❍ Yes ❍ No ❍ Planned

Intensive care unit? ❍ **Yes** ❍ **No**

IF YES ➔ Number of days ☐☐☐ Days

(Number of days spent in the intensive care unit)

➔ **SAPS II-score** *(if available)* ☐☐.☐☐

Intermediate care unit?

IF YES ➔ Number of days ❍ **Yes** ❍ **No**

(Number of days spent in the intermediate care unit) ☐☐☐ Days

DISCHARGE

Survival ❍ **Yes** ❍ **No**

IF SURVIVAL "YES"
»Discharge

- **Date of discharge**

  Day ☐☐ Month ☐☐ Year ☐☐☐☐

- **Destination** *(check **only one**)*

  ❍ Rehabilitation ➔ { ❍ In-patient (stationary) / ❍ Out-patient (ambulatory) }

  ❍ Home ➔ { ❍ with support of Spitex / ❍ without support of Spitex }

  ❍ Transfer to another hospital
  IF YES ➔ To which hospital? _____

  ❍ Retirement home/Nursing home

  ❍ Other/unknown

- **Clinical diagnosis** *(check **only one** diagnosis)*

  ❍ Myocardial infarction
  ❍ ACS (Acute Coronary Syndrome) with minimal necrosis/infarction
  ❍ Unstable angina
  ❍ Non-cardiac or unclear

Medication at discharge

- **Total number of different drugs at discharge**
  (Count all drug classes)

  ☐☐

- **Cardiovascular medication at discharge**
  (check each medication)

Aspirin, ASA	❍ Yes ❍ No
Clopidogrel	❍ Yes ❍ No
Heparins	❍ Yes ❍ No
Oral anticoagulant	❍ Yes ❍ No
Beta-blocker	❍ Yes ❍ No
ACE inhibitor	❍ Yes ❍ No
Angiotensin II receptor antagonist	❍ Yes ❍ No
Ca-channel blocker	❍ Yes ❍ No
Nicorandil, molsidomine and/or long-acting nitrates	❍ Yes ❍ No
Digoxin	❍ Yes ❍ No
Diuretic	❍ Yes ❍ No
Lipid lowering drug	
• Statin	❍ Yes ❍ No
• Ezetimibe	❍ Yes ❍ No
• Other lipid lowering drug	❍ Yes ❍ No
Amiodarone	❍ Yes ❍ No
Other	❍ Yes ❍ No

- **Other medication at discharge**

Oral antidiabetic	❍ Yes ❍ No
Insulin	❍ Yes ❍ No
Antidepressant	❍ Yes ❍ No
Sedative/Tranquilizer	❍ Yes ❍ No

IF SURVIVAL "NO"
»Death

- **Date of death**

  Day ☐☐ Month ☐☐ Year ☐☐☐☐

- **Cause of death**
  (Check the most important cause)

  ❍ Pump failure
  ❍ Mechanical complications (e.g. rupture)
  ❍ Arrhythmia
  ❍ Sepsis/ SIRS/ Multiorgan failure
  ❍ Bleeding
  ❍ Non-cardiac
  ❍ Other

- **Clinical Diagnosis**

  ❍ Myocardial infarction
  ❍ ACS (Acute Coronary Syndrome) with minimal necrosis/infarction
  ❍ Unstable angina
  ❍ Non-cardiac or unclear

Questionnaire completed?	❍ Yes

PATIENT FOLLOW-UP

The patient has given his/her formal consent for follow-up interview.

Name of Patient _____

Tel. Nr.: _____

B.2 The AMIS Plus Data Entry Instructions

University of Zurich
Institute of Social and
Preventive Medicine

A Acute
M Myocardial
I Infarction in
S Switzerland

AMIS Plus
Data Entry Instructions

March 2005

All patients with initial suspicion of acute coronary syndrome (AMI, ACS with minimal necrosis, unstable angina) at hospital admission should be **considered** for inclusion in the AMIS Plus Registry. However, a completed questionnaire should only be submitted for patients who have been clinically diagnosed with acute coronary syndrome.

Data can be sent to us directly online www.amis-plus.ch under “Data Entry“ or by completing and posting the paper questionnaire to the AMIS Plus Data Center.

Please pay attention to:

- Diagnosis (does discharge diagnosis correspond to inclusion criteria?)
- Chronological order of dates and times
- Legibility (if illegible the paper questionnaire will have to be returned)
- Completeness (enter/write “unknown” in the comments field/on the paper questionnaire if data are unobtainable)

AMIS Plus Data Center
Hirschengraben 84
CH-8001 Zurich

Tel: +41 (0)44-634 48 30
Fax: +41 (0)44-634 49 86

E-Mail: amis@ifspm.unizh.ch
www.amis-plus.ch

AMIS DATA ENTRY IDENTIFICATION	
Hospital	Each participating hospital receives its own official four-letter code. This code is entered automatically with online data entry.
Patient ID number	This ID number is unique to the particular patient and remains the same for each subsequent hospitalization. This number should enable the physician to easily identify the patient, and with the additional information on admission date, to find the specific hospital record of a particular admission in the hospital archive. This is not the case number, which changes even for the same patient for each hospitalization**.**
VO Set Nr. / Code	Set Nr. and Code concerns Verein Outcome hospitals only.
Physician ID	Even though we use the word physician, this person could be a non-physician in some cases. For Online Data Entry, Physician ID will be provided by the Data Center. It enables login to online Data Entry together with a password. Please contact the AMIS Plus Data Center (amis@ifspm.unizh.ch) for your personal user name and password. For those using the paper questionnaire, the initial of your first name can be written together with your surname (e.g. Hans Muster – hmuster).
Date of data entry	The date the questionnaire was completed. This date is entered automatically for online entry.

PATIENT AT ADMISSION	
Date of birth	Mandatory field
Gender	Mandatory field
Weight	Rounded up to the nearest kilogram
Height	In centimeters
Admission date and time to this hospital	
Time of first medical contact leading to hospitalization	If the time of first medical contact is different from admission time and is known to you, please note this time. It is important for the calculation of the time taken to achieve immediate therapy.
Insurance coverage	Does the patient have basic, semi-private or private insurance?
Transfer	Please do not forget to indicate from which hospital the patient was transferred as this is essential when analyzing the data.
»Condition	
Symptoms at admission	Please note if the symptoms at admission were typical or atypical for acute coronary syndrome.
Symptom onset date and time	Date and time of symptom onset is usually **before** date and time of admission (however, please refer to special cases: ACS during hospitalization). With this data, patient delay time can be calculated. Time can be approximate. If unknown, please note this in the comments field.
Killip classification	Killip classification measured the severity of heart failure with myocardial infarction (*Killip and Kimball; Am J Card 1967; 20:457-64*). Patients are ranked by Killip class: • I: no clinical signs of heart failure • II: crackles, S3 gallop and elevated jugular venous pressure • III: frank pulmonary edema • IV: cardiogenic shock – hypertension (systolic<90mmHg) and evidence of peripheral vasoconstriction (oliguria, cyanosis, sweating).
ECG on admission	Please note all changes in the initial ECG. If there are no changes or ECG is normal then only this choice is possible.
Location of ischemic region	Please note the ischemic region if it has been localized. If not, check undetermined.
Heart rhythm	Only one answer is possible.
»Vital signs	
Systolic blood pressure	Note systolic blood pressure at admission.
Diastolic blood pressure	Note diastolic blood pressure at admission.

REPORTED HISTORY	
Resuscitation *prior to arrival at hospital?*	If resuscitation was performed, please note which kind: cardiopulmonary reanimation and/or cardioversion or defibrillation
Regular medication	Check each drug the patient has been taking daily or regularly at the time of admission
Comorbidities (Charlson Index) **IF YES➔** Kinds of comorbidities: **Myocardial infarction:** includes all patients with at least one definite or probable MI, these patients were hospitalised and had ECG and/or enzyme changes. **Cardiac insufficiency:** includes patients with stress or paroxysmal nocturnal dyspnea who take digitalis, diuretics or afterload lowering drugs. Not included are patients on medication without change of symptoms or improvement of clinical findings. NYHA III and IV correspond to dyspnea during walking on the flat or at rest. **Peripheral vascular disease:** includes patients with intermittent claudication, post-bypass status due to arterial circulatory disorder, gangrene or acute arterial occlusive disease or with untreated thoracal or abdominal aortic aneurysm of at least 6 cm in diameter. **Cerebrovascular disease:** includes patients with an old cerebrovascular insult with few or no residuals, as well as patients with transient ischemic attacks. **Hemiplegia:** is defined as a monolateral paresis/plegia, independent of its origin. **Dementia:** includes patients with chronic cognitive deficits. **Chronic lung disease:** includes patients with at least one hospitalization due to decompensated COPD in the past. Included are also patients with dyspnea at rest in spite of treatment, those with permanent oxygen therapy, with CO2 retention, as well as those with PO2 at rest below 50mmHg. **Connective tissue disorder:** includes patients with systemic lupus erythemathosus, polymyositis, mixed connective tissue disease, polymyalgia rheumatica and moderate to severe rheumatic arthritis. **Peptic ulcer disease:** includes patients who have been treated for peptic ulcer disease, including those who had a bleeding ulcer.	**Rules:** - only active comorbidities on the day of admission are considered - take all listed comorbidities into account (but only these!) **Diabetes:** defined by the fact that patient is treated with insulin or oral antidiabetics. **Diabetes with target organ damage:** includes diabetics with neuropathy, angiopathy, kidney disease and other manifested target organ damage. **Mild liver disease:** includes patients with an increase of transaminase levels below twice the upper normal limit. **Moderate liver disease:** includes patients with an increase of transaminase levels above twice the upper normal limit. **Severe liver disease:** includes patients with coagulopathy and/or ascites. **Moderate to severe renal disease:** includes patients with a serum creatinine of at least 260 micromol/l (3mg%). **Malignant neoplasm:** includes patients with solid malignant tumors without metastases but with initial treatment during the last five years (incl. carcinoma of the breast, colon, bronchial and other carcinoma). **Leukemia:** includes patients with acute and/or chronic myeloid leukemia, acute and/or chronic lymphatic leukemia and polycythemia vera. **Lymphoma:** includes patients with Hodgkin's disease, lymphosarcoma, Waldenström's disease, myeloma (plasmacytoma) and other lymphoma. **Metastatic solid tumor:** includes patients with metastatic solid tumors (incl. carcinoma of the breast, colon, bronchial and other carcinoma). **AIDS (stage C):** includes patients with definite or probable AIDS or AIDS-related complex.
»**Past history**	
Ischemic heart diseases	Please check each occurrence of previous heart disease.
Risk factors	Check each documented or treated risk factor.
Family history	Is there family history in a first-degree relative younger than 60 years?
History of arterial hypertension	Does this patient have a history of hypertension (>140/90) and has this been treated or not?
History of hyperlipidemia	Does this patient have a history of dyslipidemia and has this been treated or not?
Smoking	History confirming cigarette smoking: current smoker: has smoked at least 100 cig. (5 packs) in his life and is currently smoking, ex-smoker: stopped smoking cigarettes more than 1 year before this admission, never smoker: never smoked cigarettes.

IMMEDIATE THERAPY	
Initial therapeutic strategy	Please note the very first therapeutic strategy intended. This does not have to correspond to the therapy actually performed.
	Primary PCI: Is percutaneous coronary intervention the first intended choice of therapy.? Usually within 24 hours from symptom onset.
	PCI facilitated (before catheter laboratory): Pretreatment with lysis and/or GPIIb/IIIa antagonist to facilitate not immediate but early PCI.
	Thrombolysis: Is thrombolysis the first intended choice of therapy?
	Primary conservative treatment, angiography planned: Is conservative therapy the first intended choice of therapy along with a planned angiography?
	Conservative treatment, elective angiography if problems: Is conservative therapy the first intended choice of therapy with angiography as an option in case of additional problems?
	Primary palliative/symptomatic therapy: Is palliative and/or symptomatic therapy intended?
Thrombolysis	If thrombolysis was performed in your hospital, please note the date and time of thrombolysis. The time is usually within 24 hours from symptom onset. If thrombolysis was not performed, note the most important reason for denial.
Percutaneous Coronary Intervention (PCI) • **Starting time of PCI**	PCI is a more general term than the formerly used PTCA (percutaneous transluminal coronary angioplasty), which also includes interventions such as stenting. Starting time of PCI is defined as the time of the patient's first automatic arterial blood pressure measurement in the heart catheter laboratory.
• **Reason for performing PCI**	First strategy, which is usually within 24 hours from chest pain onset.
• **TIMI flow at the end of PCI**	Note the TIMI flow at the end of PCI. 0 = no antegrade flow beyond a coronary occlusion; I = minimal flow beyond the occlusion, filling the distal coronary bed is incomplete; II = delayed or sluggish flow with complete filling of the distal territory; III = normal flow which fills the distal coronary bed completely
• **Angiographic findings**	Are there angiographic findings? If yes, check whether one, two or three vessels or left main are affected. If there are no abnormalities then only this choice is possible.
• **Left ventricular ejection fraction**	Check the value of left ventricular ejection fraction. Only one choice is possible.
• **Vessel treated**	Check which vessel was treated. Multiple choices are possible.
• **PCI with stent?**	Was a stent used? If yes, was it a drug eluting or absorbable stent/s?
Medication for immediate therapy	Please check each medication given within 24 hours from admission including emergency medication.

LABORATORY PARAMETERS	
Glycemia (mmol/l) *(on admission)* **Creatinine (µmol/l)** *(on admission)* **Maximum CK (IU/l)** *(max. level during hospitalization)* **Maximum CK-MB (IU/l)** *(max. level during hospitalization)* **Troponin I (free) (µg/l)** *(max. level during hospitalization)* **Troponin T (µg/l)** *(max. level during hospitalization)* **Troponin I (total) (µg/l)** *(max. level during hospitalization)* **Total Cholesterol (mmol/l)** *(within 24 hours of chest pain onset)* **HDL Cholesterol (mmol/l)** *(within 24 hours of chest pain onset)* **CRP (mg/l)** *(on admission, if available)* **BNP (pg/ml)** *(if available)* **NT-ProBNP (pg/ml)** *(max. level during hospitalization)*	Check all the laboratory levels available and measured in your hospital. Please check parameter units. Note the maximum levels of CK, CK-MB, free Troponin I, Troponin T and total Troponin I during hospital stay. Cholesterol und HDL levels within 24 hours of chest pain onset. Note the admission levels of CRP, BNP and maximal level of NT-ProBNP (if available) in your hospital.

HOSPITALIZATION	
Complications	Check each complication which occurred during this hospitalization. Cardiogenic shock should only be checked if it occurred during hospitalization and was not an admission diagnosis. SIRS = systemic inflammatory response syndrome.
Therapeutic interventions	Check each intervention performed during this hospitalization.
Performed diagnostic / therapeutic procedures	Check each procedure.
• **Echocardiography** **IF YES ➔** • Date of first echocardiography • Time • Left ventricular ejection fraction	Note if echocardiography was performed. If yes, please note the date and time of the first performed echocardiography at this hospitalisation and note the value of left ventricular ejection fraction.
• **Stress test** *(including stress echo and/or isotops)* **IF DONE ➔** Ischemia	Was a stress test performed? This includes stress test, stress echo or stress test with isotopes. If yes, was ischemia found?
• **Coronary angiography** • **Coronary artery bypass grafting (CABG)** • **Electrophysiology study (invasive)**	Check each intervention.
Intensive care unit? **IF YES ➔ Number of days** **➔ SAPS II-score** *(if available)*	If patient was in ICU please note the number of days the patient spent in ICU. SAPS II Score (Simplified Acute Physiology Score) is usually calculated in every ICU (*Le Gall et al. JAMA 1993, 270:2957-63*). Note the score if available.
Intermediate care unit? **IF YES ➔ Number of days**	If patient was in an intermediate care unit please note the number of days the patient spent there.

DISCHARGE	
Survival	If the patient was alive at the end of hospitalization please fill in the date of discharge, destination, clinical diagnosis and check each medication given at discharge**.**
IF SURVIVAL "YES" **»Discharge**	
• **Date of discharge**	Mandatory field
• **Destination**	Where was the patient discharged to?
Clinical diagnosis	Please check the clinical diagnosis. **Myocardial infarction:** Symptoms and/or ECG changes compatible with ACS and enzymes (CK or CK-MB) at least twice the upper limits of normal. **ACS with minimal necrosis:** Symptoms and/or ECG changes compatible with ACS but enzymes (CK or CK-MB) below twice the upper limits of normal and troponin positive. **Unstable angina:** Symptoms and/or ECG typical for ACS (e.g. ischemic pain at rest, ECG changes during/after symptoms), enzymes normal. **Non-cardiac or unclear:** If patient was admitted for ACS but the diagnosis could not be confirmed at discharge.
Medication at discharge	
• **Total number of different drugs at discharge**	Please count each different drug at discharge.
• **Cardiovascular medication at discharge**	Check each drug given at discharge.
• **Other medication at discharge**	

IF SURVIVAL "NO"	If the patient died during this hospitalization please fill out date and cause of death as well as clinical diagnosis.
»Death	
Date of death	
Cause of death	
Clinical Diagnosis	**Myocardial infarction:** Symptoms and/or ECG changes compatible with ACS and enzymes (CK or CK-MB) at least twice the upper limits of normal. **ACS with minimal necrosis:** Symptoms and/or ECG changes compatible with ACS but enzymes (CK or CK-MB) below twice the upper limits of normal and troponin positive. **Unstable angina:** Symptoms and/or ECG typical for ACS (e.g. ischemic pain at rest, ECG changes during/after symptoms), enzymes normal. **Non-cardiac or unclear:** If patient was admitted for ACS but the diagnosis could not be confirmed at death.

PATIENT FOLLOW-UP	
The patient has given his/her formal consent for a follow-up interview. Name of Patient: **Tel. Nr.:**	If the patient has given his/her formal consent for a follow-up interview please note his/her name and telephone number.
Three and twelve month follow-up interviews will be carried out by the AMIS Plus Data Center Team. The results of this follow-up will be available in the "AMIS Plus Data Analysis" (Webb App) using your own login.	

B.3 The AMIS Plus Patient Information and Consent form

A Acute
M Myocardial
I Infarction in
S Switzerland

HOSPITAL	
PATIENT ID NUMBER	

Patient information

Dear Patient

The AMIS Plus Registry collects data on diagnostics and treatment of patients treated for heart complaints (Diagnosis: acute coronary syndrome). These data form the basis for continuous improvement in diagnostics and treatment of these complaints.

In order to collect even more accurate data on the treatment results of hospital stays we require long-term observations. We are therefore reliant on you.

We request your consent to personally ring you 3 and 12 months after this hospital stay to ask you about your state of health.

These data will be anonymised at the data collection center and can only be seen by the doctor responsible at your hospital.

Consent

I hereby agree, in accordance with the above information, that a doctor or member of the study personnel may telephone me to ask questions regarding my state of health and these data may then be anonymously evaluated.

NAME OF PATIENT	
DATE OF BIRTH	
TELEPHONE NUMBER	
DATE	
PATIENT'S SIGNATURE	

AMIS Plus Data Center
Hirschengraben 84
CH-8001 Zürich

Tel: +41 (0)44-634 48 30
Fax: +41 (0)44-634 49 86

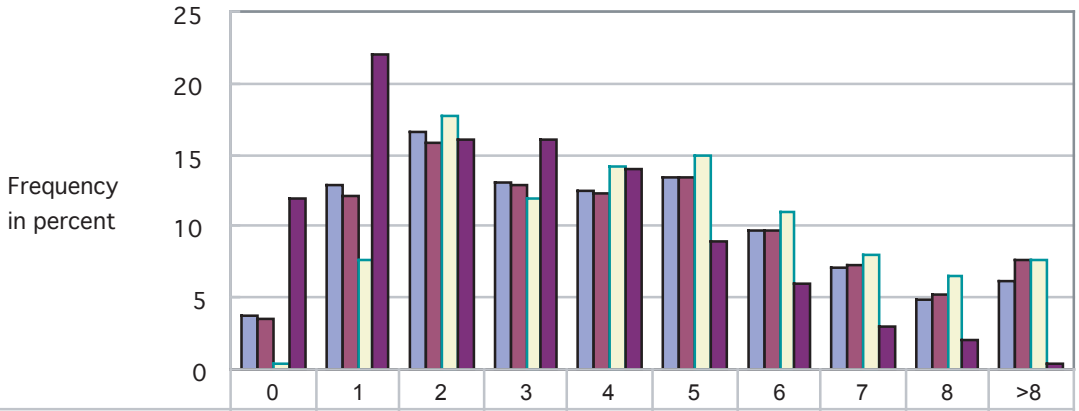
E-Mail: amis@ifspm.unizh.ch
www.amis-plus.ch

C

Appendix: Statistical Evaluation

C.1 Statistical Results of TIMI Evaluation on various datasets

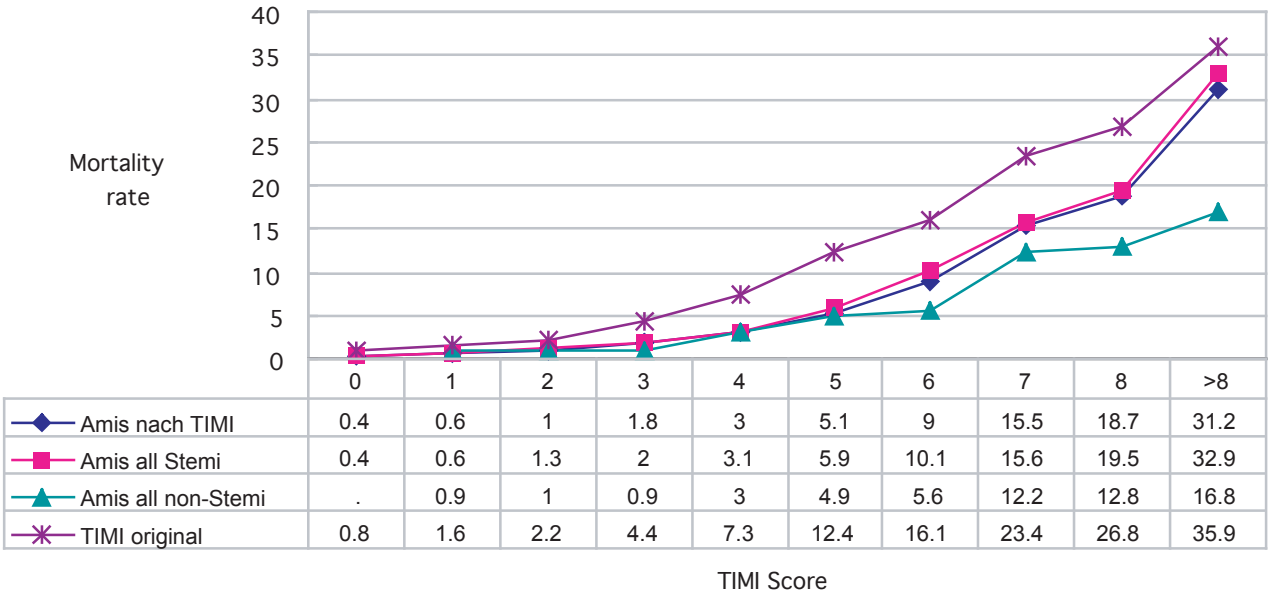
C.1.1 AMIS dataset - general results



	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	3.8	12.9	16.6	13	12.5	13.5	9.7	7	4.9	6.2
Amis all Stemi	3.5	12.2	15.8	12.9	12.4	13.4	9.7	7.3	5.2	7.6
Amis all non-Stemi	0.3	7.7	17.8	12	14.2	14.9	11	8	6.6	7.6
TIMI original	12	22	16	16	14	9	6	3	2	0.3

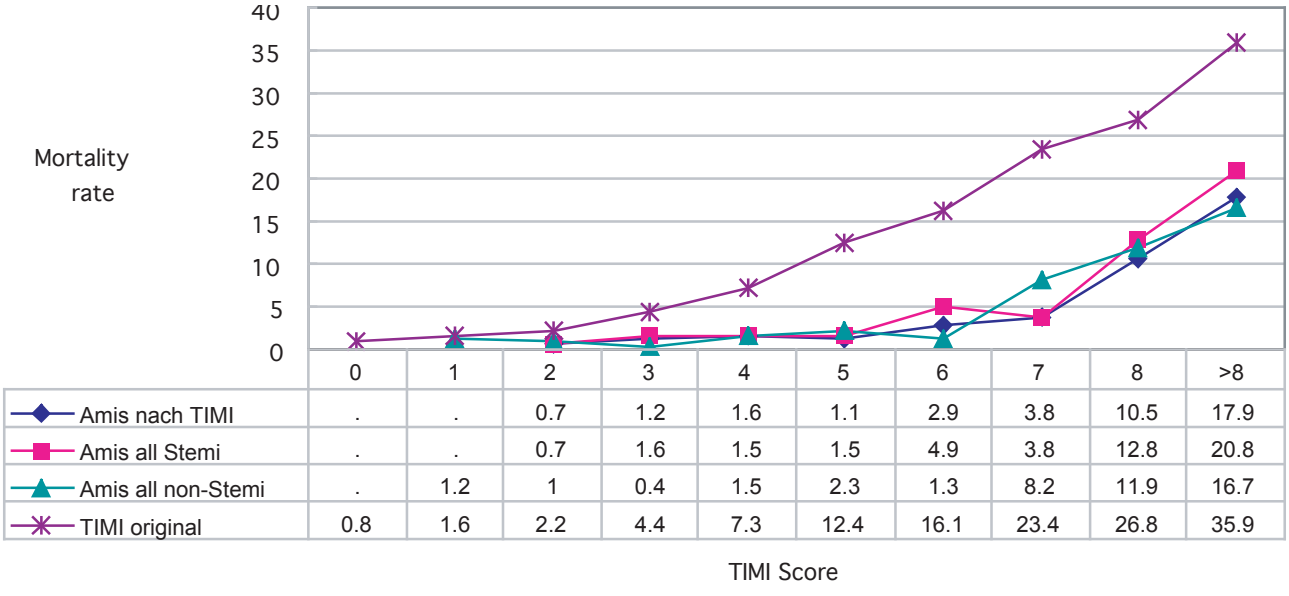
TIMI Score

Figure C.1: Original TIMI cohort vs. AMIS set and AMIS subsets - Distribution of patients at risk



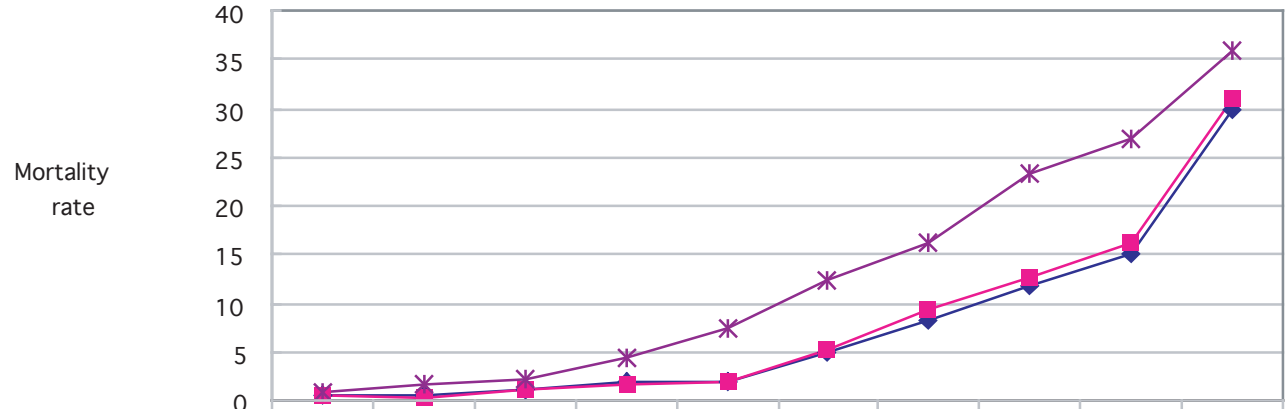
	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	0.4	0.6	1	1.8	3	5.1	9	15.5	18.7	31.2
Amis all Stemi	0.4	0.6	1.3	2	3.1	5.9	10.1	15.6	19.5	32.9
Amis all non-Stemi	.	0.9	1	0.9	3	4.9	5.6	12.2	12.8	16.8
TIMI original	0.8	1.6	2.2	4.4	7.3	12.4	16.1	23.4	26.8	35.9

Figure C.2: Original TIMI cohort vs. AMIS set and AMIS subsets- Comparison of mortality rates



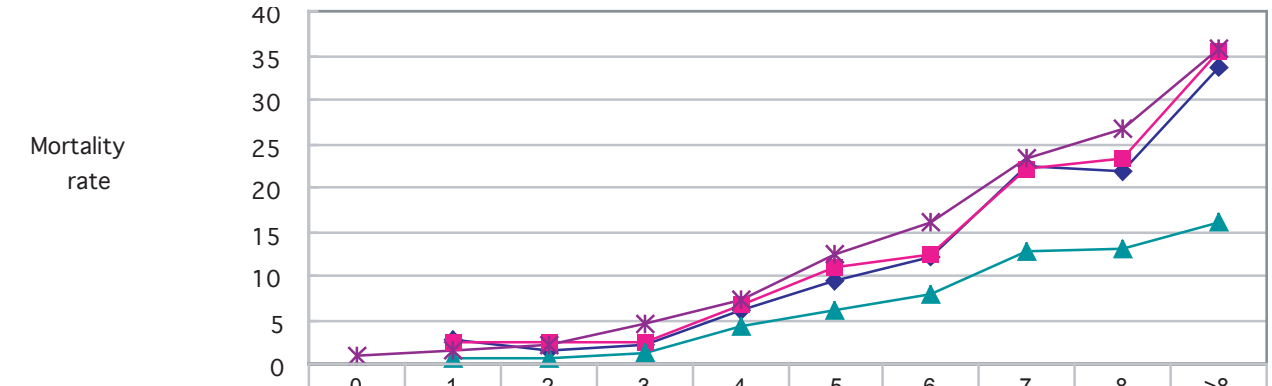
	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	.	.	0.7	1.2	1.6	1.1	2.9	3.8	10.5	17.9
Amis all Stemi	.	.	0.7	1.6	1.5	1.5	4.9	3.8	12.8	20.8
Amis all non-Stemi	.	1.2	1	0.4	1.5	2.3	1.3	8.2	11.9	16.7
TIMI original	0.8	1.6	2.2	4.4	7.3	12.4	16.1	23.4	26.8	35.9

Figure C.3: AMIS subsets - Mortality Rates of patient who received pci treatment



	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	0.6	0.5	1.1	1.9	2	4.8	8.2	11.7	15.2	30
Amis all Stemi	0.6	0.4	1.2	1.7	2	5.3	9.4	12.5	16.3	31
TIMI original	0.8	1.6	2.2	4.4	7.3	12.4	16.1	23.4	26.8	35.9

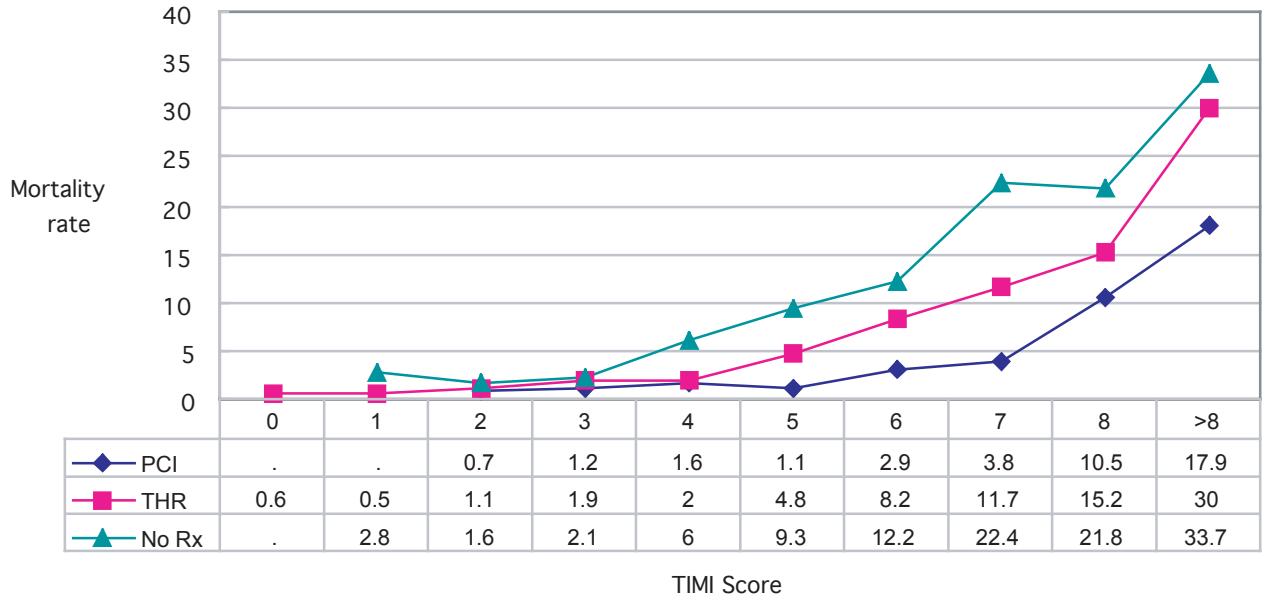
Figure C.4: AMIS subsets - Mortality Rates of patient who received thrombolysis treatment



	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	.	2.8	1.6	2.1	6	9.3	12.2	22.4	21.8	33.7
Amis all Stemi	.	2.5	2.5	2.3	6.7	10.8	12.5	22.2	23.3	35.4
Amis all non-Stemi	.	0.6	0.5	1.2	4.3	6.2	7.8	12.8	13.1	16.2
TIMI original	0.8	1.6	2.2	4.4	7.3	12.4	16.1	23.4	26.8	35.9

Figure C.5: AMIS subsets - Mortality Rates of patient who received no treatment

C.1.2 AMIS according to TIMI dataset



	0	1	2	3	4	5	6	7	8	>8
PCI	.	.	0.7	1.2	1.6	1.1	2.9	3.8	10.5	17.9
THR	0.6	0.5	1.1	1.9	2	4.8	8.2	11.7	15.2	30
No Rx	.	2.8	1.6	2.1	6	9.3	12.2	22.4	21.8	33.7

Figure C.6: Comparison of mortality rates among treatment groups

Treatment	Number of Records	AUC
no treatment	2116	0.75
thrombolysis	2203	0.809
PCI	2718	0.797

Table C.1: AUC-comparsion of different treatments on AMIS according to TIMI dataset

C.1.3 AMIS all stemi dataset

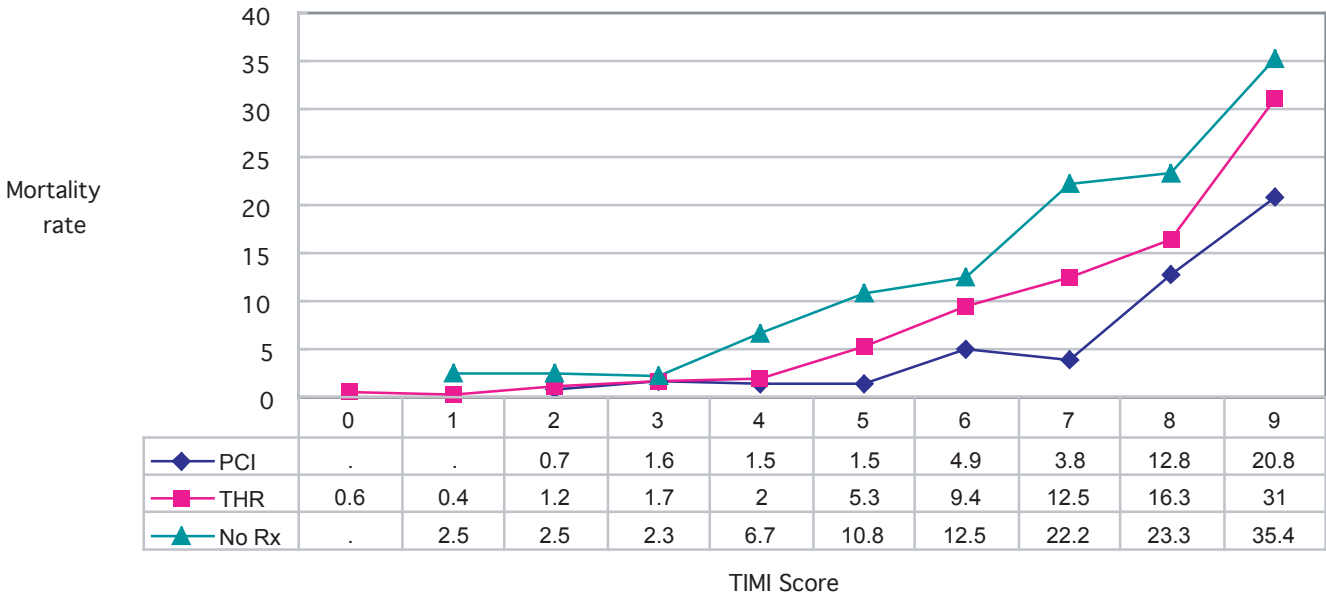
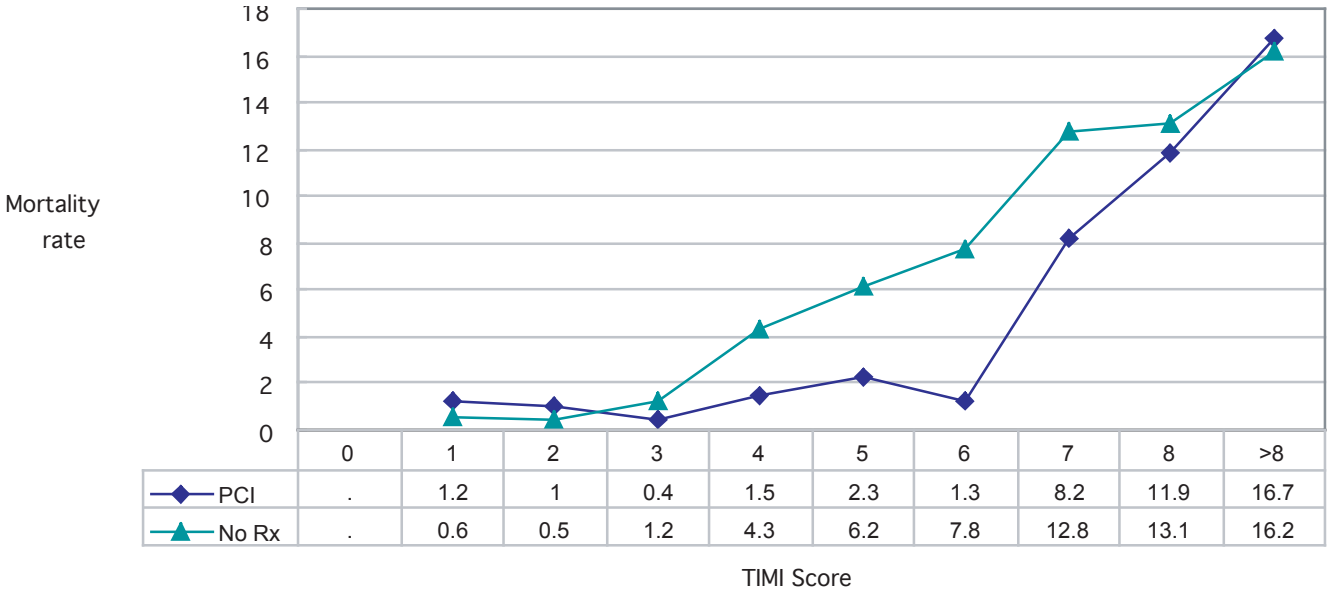


Figure C.7: Comparison of mortality rates among treatment groups

Treatment	Number of Records	AUC
no treatment	2480	0.745
thrombolysis	2470	0.82
PCI	3053	0.814

Table C.2: AUC-comparsion of different treatments on AMIS all stemi dataset

C.1.4 AMIS all non stemi dataset



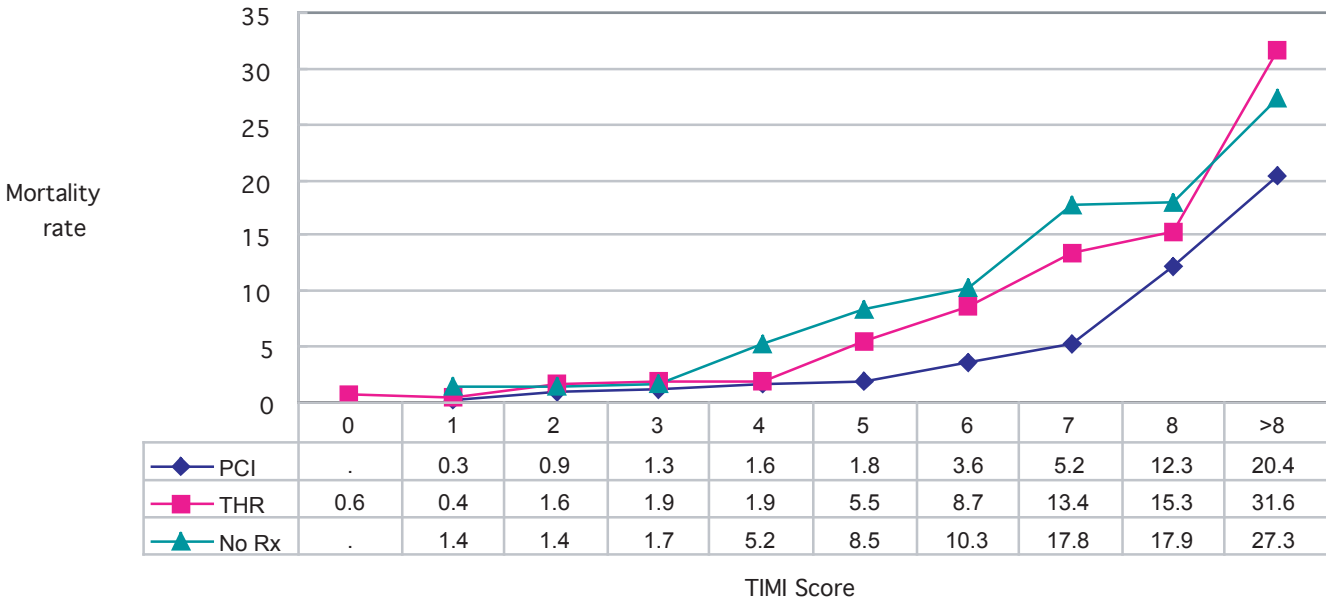
	0	1	2	3	4	5	6	7	8	>8
PCI	.	1.2	1	0.4	1.5	2.3	1.3	8.2	11.9	16.7
No Rx	.	0.6	0.5	1.2	4.3	6.2	7.8	12.8	13.1	16.2

Figure C.8: Comparison of mortality rates among treatment groups

Treatment	Number of Records	AUC
no treatment	2712	0.732
thrombolysis	176	0.743
PCI	1706	0.741

Table C.3: AUC-comparsion of different treatments on AMIS all non stemi dataset

C.1.5 AMIS all dataset



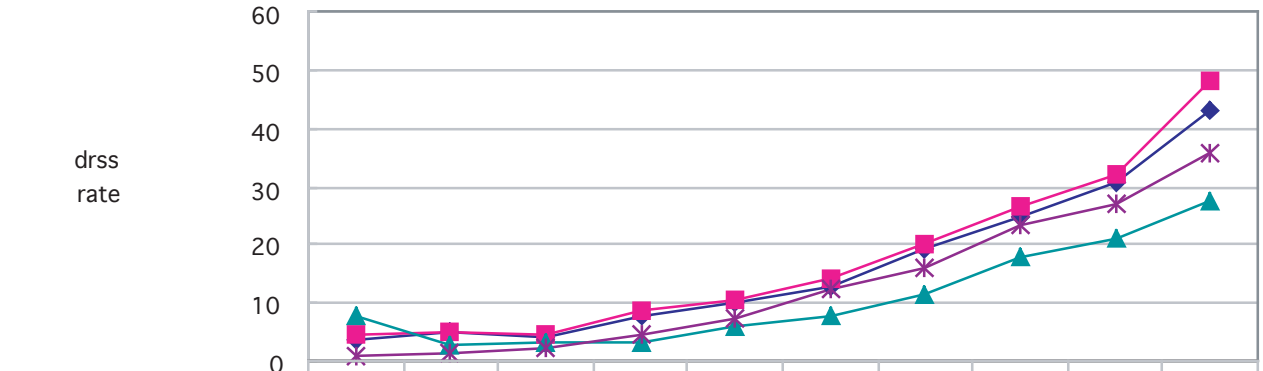
	0	1	2	3	4	5	6	7	8	>8
PCI	.	0.3	0.9	1.3	1.6	1.8	3.6	5.2	12.3	20.4
THR	0.6	0.4	1.6	1.9	1.9	5.5	8.7	13.4	15.3	31.6
No Rx	.	1.4	1.4	1.7	5.2	8.5	10.3	17.8	17.9	27.3

Figure C.9: Comparison of mortality rates among treatment groups

Treatment	Number of Records	AUC
no treatment	5223	0.745
thrombolysis	2660	0.814
PCI	4849	0.785

Table C.4: AUC-comparsion of different treatments on AMIS all dataset

C.1.6 AMIS MACE calculations



	0	1	2	3	4	5	6	7	8	>8
Amis nach TIMI	3.8	4.9	4	8	10.2	12.8	19.2	24.8	30.6	42.9
Amis all Stemi	4.4	5	4.7	8.5	10.4	14	20.3	26.5	32.2	48.2
Amis all non-Stemi	7.7	2.9	3.3	3.4	6	7.9	11.6	17.8	21	27.7
TIMI original	0.8	1.6	2.2	4.4	7.3	12.4	16.1	23.4	26.8	35.9

TIMI Score

Figure C.10: Original TIMI cohort vs. AMIS dataset and AMIS subsets - MACE (Death, Reinfarction, Stroke, Shock) Rates

D

Appendix: Medical Definitions

D.1 Indications for Fibrinolytic Therapy

The following section was copied out of the regulations published by the American Heart Association.
Class I

- In the absence of contraindications, fibrinolytic therapy should be administered to STEMI patients with symptom onset within the prior 12 hours and ST elevation greater than 0.1 mV in at least 2 contiguous precordial leads or at least 2 adjacent limb leads. (Level of Evidence: A)
- In the absence of contraindications, fibrinolytic therapy should be administered to STEMI patients with symptom onset within the prior 12 hours and new or presumably new LBBB. (Level of Evidence: A)

Class IIa

- In the absence of contraindications, it is reasonable to administer fibrinolytic therapy to STEMI patients with symptom onset within the prior 12 hours and 12-lead ECG findings consistent with a true posterior MI. (Level of Evidence: C)
- In the absence of contraindications, it is reasonable to administer fibrinolytic therapy to patients with symptoms of STEMI beginning within the prior 12 to 24 hours who have continuing ischemic symptoms and ST elevation greater than 0.1 mV in at least 2 contiguous precordial leads or at least 2 adjacent limb leads. (Level of Evidence: B)

Class III

- Fibrinolytic therapy should not be administered to asymptomatic patients whose initial symptoms of STEMI began more than 24 hours earlier. (Level of Evidence: C)

- Fibrinolytic therapy should not be administered to patients whose 12-lead ECG shows only ST-segment depression except if a true posterior MI is suspected. (Level of Evidence: A)

Contraindications and Cautions for Fibrinolysis in STElevation Myocardial Infarction[1]
Absolute contraindications

- Any prior intracranial hemorrhage
- Known structural cerebral vascular lesion (e.g., arteriovenous malformation)
- Known malignant intracranial neoplasm (primary or metastatic)
- Ischemic stroke within 3 months EXCEPT acute ischemic stroke within 3 hours
- Suspected aortic dissection
- Active bleeding or bleeding diathesis (excluding menses)
- Significant closed-head or facial trauma within 3 months

Relative contraindications

- History of chronic, severe, poorly controlled hypertension
- Severe uncontrolled hypertension on presentation (systolic blood pressure greater than 180 mm Hg or diastolic blood pressure greater than 110 mmHg)[2]
- History of prior ischemic stroke greater than 3 months, dementia, or known intracranial pathology not covered in contraindications
- Traumatic or prolonged (greater than 10 minutes) cardiopulmonary resuscitation or major surgery (less than 3 weeks)
- Recent (within 2-4 weeks) internal bleeding
- Noncompressible vascular punctures
- For streptokinase/anistreplase: prior exposure (more than 5 days ago) or prior allergic reaction to these agents
- Pregnancy
- Active peptic ulcer
- Current use of anticoagulants: the higher the international normalized ratio, the higher the risk of bleeding

[1] Viewed as advisory for clinical decision making and may not be all-inclusive or definitive.
[2] Could be an absolute contraindication in low-risk patients with MI.

E

Appendix: Decision Tree generated by C 4.5

cprarr = 1
| systbpBin = 1: 1 (0.0)
| systbpBin = 2: 1 (1.05/0.04)
| systbpBin = 3: 0 (22.24/7.32)
| systbpBin = 4
| | hrtrateBin = 1: 1 (0.0)
| | hrtrateBin = 2: 0 (5.35/2.46)
| | hrtrateBin = 3: 1 (39.18/14.33)
| | hrtrateBin = 4: 1 (24.7/10.97)
| | hrtrateBin = 5: 0 (19.7/6.1)
| | hrtrateBin = 6: 1 (0.0)
| | hrtrateBin = 7: 1 (0.0)
| | hrtrateBin = 8: 1 (0.0)
| | hrtrateBin = 9: 1 (0.0)
| | hrtrateBin = 10: 1 (0.0)
| systbpBin = 5
| | age = 1: 1 (1.0)
| | age = 2: 1 (1.4/0.4)
| | age = 3: 1 (5.26/2.0)
| | age = 4: 1 (12.6/2.73)
| | age = 5: 1 (20.27/4.74)
| | age = 6: 1 (17.37/5.0)
| | age = 7: 0 (21.34/9.12)
| | age = 8: 1 (9.17/3.36)
| | age = 9: 1 (0.26/0.06)
| | age = 10: 1 (0.0)

| systbpBin = 6: 1 (34.62/15.28)
| systbpBin = 7: 1 (7.0/3.32)
| systbpBin = 8: 1 (1.08/0.04)
| systbpBin = 9: 1 (0.0)
| systbpBin = 10: 1 (0.0)
cprarr = 0
| killip = 1
| | age = 1: 1 (2.0)
| | age = 2: 1 (61.94)
| | age = 3: 1 (308.44/1.0)
| | age = 4: 1 (887.82/5.0)
| | age = 5: 1 (1189.59/4.0)
| | age = 6: 1 (1211.59/19.97)
| | age = 7
| | | hrtrateBin = 1: 1 (0.0)
| | | hrtrateBin = 2: 1 (34.14/2.02)
| | | hrtrateBin = 3: 1 (773.65/26.42)
| | | hrtrateBin = 4
| | | | systbpBin = 1: 1 (0.0)
| | | | systbpBin = 2: 1 (0.0)
| | | | systbpBin = 3: 0 (1.0/0.0)
| | | | systbpBin = 4: 1 (17.28/3.94)
| | | | systbpBin = 5: 1 (111.77/9.92)
| | | | systbpBin = 6: 1 (130.41/3.0)
| | | | systbpBin = 7: 1 (31.63/1.0)
| | | | systbpBin = 8: 1 (3.01)
| | | | systbpBin = 9: 1 (0.0)
| | | | systbpBin = 10: 1 (0.0)
| | | hrtrateBin = 5: 1 (24.13/4.02)
| | | hrtrateBin = 6: 1 (2.01/1.0)
| | | hrtrateBin = 7: 1 (1.01/0.0)
| | | hrtrateBin = 8: 1 (0.0)
| | | hrtrateBin = 9: 1 (0.0)
| | | hrtrateBin = 10: 1 (0.0)
| | age = 8
| | | systbpBin = 1: 1 (0.0)

| | | systbpBin = 2: 1 (0.0)
| | | systbpBin = 3: 0 (4.01/0.01)
| | | systbpBin = 4
| | | | hrtrateBin = 1: 1 (0.0)
| | | | hrtrateBin = 2: 1 (5.0)
| | | | hrtrateBin = 3: 1 (43.49/9.97)
| | | | hrtrateBin = 4: 1 (22.06/5.97)
| | | | hrtrateBin = 5: 1 (4.97/0.97)
| | | | hrtrateBin = 6: 0 (1.0)
| | | | hrtrateBin = 7: 1 (1.0)
| | | | hrtrateBin = 8: 1 (0.0)
| | | | hrtrateBin = 9: 1 (0.0)
| | | | hrtrateBin = 10: 1 (0.0)
| | | systbpBin = 5: 1 (239.45/24.0)
| | | systbpBin = 6: 1 (212.57/13.0)
| | | systbpBin = 7: 1 (70.08/3.0)
| | | systbpBin = 8: 1 (8.01/2.0)
| | | systbpBin = 9: 1 (2.0)
| | | systbpBin = 10: 1 (0.0)
| | age = 9: 1 (115.4/19.97)
| | age = 10: 0 (3.0/1.0)
| killip = 2
| | age = 1: 1 (0.0)
| | age = 2: 1 (2.97/0.97)
| | age = 3: 1 (23.11)
| | age = 4: 1 (72.3)
| | age = 5: 1 (135.95/4.0)
| | age = 6
| | | hrtrateBin = 1: 1 (0.0)
| | | hrtrateBin = 2: 1 (5.14/1.0)
| | | hrtrateBin = 3: 1 (107.57/3.0)
| | | hrtrateBin = 4: 1 (84.84/5.0)
| | | hrtrateBin = 5: 1 (11.0/1.0)
| | | hrtrateBin = 6: 0 (1.0)
| | | hrtrateBin = 7: 1 (1.0)
| | | hrtrateBin = 8: 1 (0.0)

| | | hrtrateBin = 9: 1 (0.0)
| | | hrtrateBin = 10: 1 (0.0)
| | age = 7
| | | hrtrateBin = 1: 1 (0.0)
| | | hrtrateBin = 2: 1 (11.98/2.01)
| | | hrtrateBin = 3: 1 (177.28/10.06)
| | | hrtrateBin = 4
| | | | systbpBin = 1: 1 (0.0)
| | | | systbpBin = 2: 1 (0.0)
| | | | systbpBin = 3: 0 (1.96/0.97)
| | | | systbpBin = 4: 1 (17.38/1.15)
| | | | systbpBin = 5: 1 (47.07/9.61)
| | | | systbpBin = 6: 1 (36.6/3.32)
| | | | systbpBin = 7: 1 (9.46/0.08)
| | | | systbpBin = 8: 1 (3.04/0.03)
| | | | systbpBin = 9: 1 (0.0)
| | | | systbpBin = 10: 1 (0.0)
| | | hrtrateBin = 5: 1 (23.05/8.01)
| | | hrtrateBin = 6: 1 (3.01/0.0)
| | | hrtrateBin = 7: 1 (0.0)
| | | hrtrateBin = 8: 1 (1.0/0.0)
| | | hrtrateBin = 9: 1 (0.0)
| | | hrtrateBin = 10: 1 (0.0)
| | age = 8
| | | systbpBin = 1: 1 (0.0)
| | | systbpBin = 2: 1 (0.0)
| | | systbpBin = 3: 0 (3.0/1.0)
| | | systbpBin = 4
| | | | hrtrateBin = 1: 1 (0.0)
| | | | hrtrateBin = 2: 0 (1.0)
| | | | hrtrateBin = 3: 1 (27.25/12.97)
| | | | hrtrateBin = 4: 1 (19.94/3.97)
| | | | hrtrateBin = 5: 1 (4.0/2.0)
| | | | hrtrateBin = 6: 1 (0.0)
| | | | hrtrateBin = 7: 1 (1.0)
| | | | hrtrateBin = 8: 1 (0.0)

| | | | hrtrateBin = 9: 1 (0.0)
| | | | hrtrateBin = 10: 1 (0.0)
| | | systbpBin = 5: 1 (156.06/25.0)
| | | systbpBin = 6: 1 (127.32/17.0)
| | | systbpBin = 7: 1 (34.94/1.0)
| | | systbpBin = 8: 1 (4.0)
| | | systbpBin = 9: 1 (0.0)
| | | systbpBin = 10: 1 (0.0)
| | age = 9
| | | systbpBin = 1: 1 (0.0)
| | | systbpBin = 2: 1 (0.0)
| | | systbpBin = 3: 0 (2.0)
| | | systbpBin = 4: 0 (20.97/8.97)
| | | systbpBin = 5
| | | | hrtrateBin = 1: 1 (0.0)
| | | | hrtrateBin = 2: 0 (1.0)
| | | | hrtrateBin = 3: 1 (15.97/2.97)
| | | | hrtrateBin = 4: 1 (22.94/10.0)
| | | | hrtrateBin = 5: 0 (2.97/1.0)
| | | | hrtrateBin = 6: 1 (0.0)
| | | | hrtrateBin = 7: 1 (0.0)
| | | | hrtrateBin = 8: 1 (0.0)
| | | | hrtrateBin = 9: 1 (0.0)
| | | | hrtrateBin = 10: 1 (0.0)
| | | systbpBin = 6: 1 (35.14/4.0)
| | | systbpBin = 7: 1 (11.97/0.97)
| | | systbpBin = 8: 1 (2.0)
| | | systbpBin = 9: 1 (0.0)
| | | systbpBin = 10: 1 (0.0)
| | age = 10: 1 (3.0/1.0)
| killip = 3
| | systbpBin = 1: 1 (0.0)
| | systbpBin = 2: 1 (0.0)
| | systbpBin = 3: 0 (6.02/3.0)
| | systbpBin = 4: 1 (73.32/31.12)
| | systbpBin = 5: 1 (116.22/26.44)

| | systbpBin = 6: 1 (98.68/17.29)
| | systbpBin = 7: 1 (30.16/1.09)
| | systbpBin = 8: 1 (10.03/0.03)
| | systbpBin = 9: 1 (1.0/0.0)
| | systbpBin = 10: 1 (0.0)
| killip = 4
| | age = 1: 1 (0.0)
| | age = 2: 1 (0.0)
| | age = 3: 1 (3.02/1.0)
| | age = 4: 1 (16.06/2.97)
| | age = 5: 1 (27.92/7.94)
| | age = 6: 1 (24.0/5.94)
| | age = 7: 1 (37.28/15.04)
| | age = 8: 0 (29.09/8.09)
| | age = 9: 1 (6.99/3.0)
| | age = 10: 1 (0.0)

F

Appendix: Glossary

ACS	Acute Coronary Syndrome
AMI	Acute Myocardial Infarction
Anterior ST-Elevation	Ecg indication of anterior *stemi*
bedside	see at presentation
AODE	Averaged one-dependence estimators. Algorithm from the bayes family with less strong independence assumptions.
Cardiogenic shock	Disease state where the heart is too damaged to supply sufficient blood to the body.
Cerebrovascular disease	Stroke
Comorbidity	Coexisting or additional disease.
Cross-validation	Typical validation method in the field of data mining. The data is split into n folds of approximately the same size and distribution, then the model is trained n times, each time leaving out one part which is used for the validation.
ECG	Electrocardiogram
In hospital mortality	Mortality of patients while they are in hospital.
LBBB	Left bundle branch block (ecg indication)
MACE	Major adverse cardiac events
NONSTEMI	non ST-Elevation Myocardial Infarction
PCI	Percutaneous coronary intervention (PCI) is a treatment procedure that unblocks narrowed coronary arteries without performing surgery.
presentation, at	Information that can be gained at first presentation of the patient.
ROCCH	ROC Convex Hull
STEMI	ST-Elevation Myocardial Infarction, indicated by the ecg findings ST-Elevation and Left Bundle Branch Block
Thrombolysis	Thrombolytic therapy involves the use of drugs that break up or dissolve blood clots, which are the main cause of both heart attacks and stroke.
Time to treatment	Time to treatment stands for the time elapsed from symptom onset to treatment.
UA	Unstable Angina

References

Antman, E. M., Cohen, M., Bernink, P. J. L. M., McCabe, C. H., Horacek, T., Papuchis, G., et al. (2000, August). The timi risk score for unstable angina/non-st elevation mi. *JAMA*, *284*(7).

Breiman, L., Friedman, J., & Olsen, R. (1984). Classification and regression trees. *Wadsworth International Group, Belmont, CA*.

Fawcett, T. (2004). *Roc graphs: Notes and practical considerations for researchers* (Tech. Rep.). HPL-2003-4 HP Labs.

Hanley, J., & McNeil, B. (1982). The meaning and use off the area under a receiver coperating characteristic (roc) curve. *Radiology*, *143*.

James M. Brophy, M., & Lawrence Joseph, P. (1995, March). Placing trials in context using bayesian analysis. *Jama*, *273*(11).

Khot, U. N., Jia, G., Moliterno, D. J., Lincoff, A. M., Khot, M. B., Harrington, R. A., et al. (2003, October). Examination for heart failure in non-st-elevation acute coronary syndromes. *JAMA*, *290*(16).

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *In C. C. Mellish (Ed.), Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137-1143.

Lee, K., Woodlief, L., Topol, E., Weaver, W., Betriu, A., Col, J., et al. (1995, March). Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction. *Circulation*, *91*(6), 1659-68.

Morrow, D. A., Antman, E. M., Charlesworth, A., Cairns, R., Murphy, S. A., Lemos, J. A. de, et al. (2000, October). Timi risk score for st-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation. *Circulation*, *102*(17), 2031-7.

Morrow, D. A., Antman, E. M., Giugliano, R. P., Cairns, R., Charlesworth, A., Murphy, S. A., et al. (2001). A simple risk index for rapid initial triage of patients with st-elevation myocardial infarction: an intime ii substudy. *The Lancet*, *358*.

Ohman, E. M., Granger, C. B., Harrington, R. A., & LEE, K. L. (2000, August). Risk stratification and therapeutic decision making in acute coronary syndromes. *JAMA*, *284*(7).

Provost, F., & Domingos, P. (2003, September). Tree induction for probability-based ranking. *Machine Learning*, *52*(3), 199-203.

Provost, F., & Fawcett, T. (2001, March). Robust classification for imprecise environments. *Machine Learning*, *42*(3), 203-231.

Scott, M., Niranjan, M., Melvin, D., & Prager, R. (1998, April). Maximum realisable performance: a principled method for enhancing performance by using multiple classifiers. *Technical Report CUED/F-INFENG/TR.320, Department of Engineering, University of Cambridge, England*.

Webb, G. I., Boughton, J., & Wang, Z. (2002). Averaged one-dependence estimators: Preliminary results. *The Australasian Data Mining Workshop*.

Webb, G. I., Boughton, J., & Wang, Z. (2005, Jannuary). Not so naive bayes: Aggregating one-dependence estimators. *Machine Learning*, *58*(1), 5-24.

Witten, I. H., & Frank, E. (2005). *Data mining - practical machine learning tools and techniques* (2. ed.). Morgan Kaufmann Publishers.