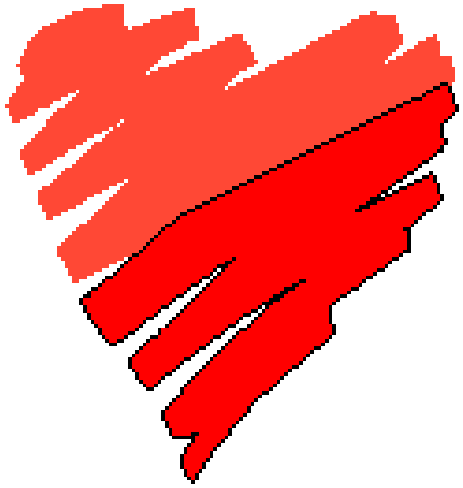
University of Zurich
Department of Informatics

Using a Bayesian Classifier for Probability Estimation: Analysis of the AMIS Score for Risk Stratification in Myocardial Infarction



DDIS Dynamic and Distributed Information Systems

Diploma Thesis 11th July 2006

Marcel Camporelli

Winterthur, Switzerland

Student-ID: 98-908-783

camporelli@gmx.ch

Advisor: **Abraham Bernstein**

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

Contents

Abstract

A recent publication has presented the AMIS model, a novel model for risk prediction in acute myocardial infarction (AMI). The model proposed therein is based on AODE, a probabilistic bayesian classifier. It outperforms TIMI, a widely accepted prediction model in the field, when classifying patients on their expected in-hospital survival or non-survival. It was hypothesised that the score which serves as a basis for the classification could be used as a probability estimator, allowing a more fine-graned stratification of patients into different mortality classes. An evaluation method for the fit of probabilistic models is developed and applied to the AMIS model. In the evaluation, the AMIS model clearly outperformes TIMI as a risk estimator.

Zusammenfassung

In einer vorangehenden Publikation wurde ein neuartiges Modell für die Risikovorhersage bei Herzinfarktpatienten (AMIS Modell) vorgestellt. Dieses basiert auf AODE, einem bayesianischen Klassifikator. Es ist bekannt, dass AMIS die Patienten besser bezüglich Überleben oder Nicht-Überleben im Spital klassifiziert als TIMI, ein in der medizinschen Fachwelt breit akzeptiertes Vorhersagemodell. Es wurde vermutet, dass sich die Werte, die zur Klassifikation in AMIS dienen, als Wahrscheinlichkeiten interpretieren lassen und so eine fein-körnigere Einteilung in verschiedene Risikoklassen möglich wäre. Eine Methodik für die Evaluation von Wahrscheinlichkeitsvorhersagen wird vorgestellt und auf das AMIS Modell angewendet. Es zeigt sich, dass AMIS auch hinsichtlich der Vorhersage von Wahrscheinlichkeiten bessere Resultate erzielt als TIMI.

Acknowledgments

This thesis has been elaborated in collaboration with Dr. David Kurz from the Cardiological Centre at the Triemli Hospital in Zurich. I would like to thank him for his time and dedication.

Katrin Hunt for giving me an understanding of her work at an early stage and who helped me with many questions.

Many thanks go to Peter Vorburger for countless inspiring and insightful discussions.

Last, but not least, I would like to thank Prof. Bernstein for his support and for valuable advice on many intricate matters that I encountered when writing this thesis.

Software packages which helped me a lot and deserve credits as well are R, LyX, LaTeX, Eclipse and Sun's JAVA platform.

1 Introduction

1.1 Myocardial Infarction for Computer Scientists

Before beginning the thesis, the most important medical terms will be shortly explained for the reader unfamiliar with it.

(A)MI (Acute) Myocardial Infarction – commonly called 'heart attack' – a heart condition where blood supply to some parts of the heart is interrupted, causing death and scarring of the local tissue.

STEMI Based on the electrocardiogram, two types of MIs can be distinguished: Those with ST-elevations (a particular shape in the ECG) and those without (Non-STEMI)

Non-STEMI MI which is not a STEMI

PCI Percutaneous Coronary Intervention. Therapy option for MIs. A device is inserted in the affected vessel in order to extend it. The most common technique is the well-known balloon angioplasty ('ballooning')

Thrombolysis Therapy option for MIs. A thrombolytic (blood diluting) medical drug is given to the patient so that the blood can more easily flow through the narrowed vessel.

AMIS Acute myocardial infarction in Switzerland; 1. name of the nation-wide registry which contains data about MI patients from most hospitals in Switzerland; 2. name of the novel risk stratification model based on the AMIS data.

TIMI A widely-used risk stratification model developed by the TIMI (Thrombolysis in Myocardial Infarction) study group.

1.2 Motivation

In the treatment of acute myocardial infarctions[1](MIs), it is vital for medical staff to assess the risk of their patients. For this purpose, various risk scores have been developed so far. Based on the levels of risk associated to a patient, therapy can be optimised in order to achieve the best possible outcome.

The existing risk scores have been developed mainly using traditional statistical methods such as multivariate regression techniques. Modern data mining/machine learning techniques have only rarely been applied. However, since extensive, integrated data bases about MI patients are incresingly available, it seems promising to apply these techniques in order to find new prediction models.

[1]Medical terms will be used without explication throughout this thesis. The reader not familiar with the related terminology is encouraged to consult the Glossary.

A master thesis previously published at our institute ([Hunt, 2006]) has presented the AMIS[2] model, a novel risk prediction score for MIs. As a classifyer, it has outperformed existing statistical scores significantly when evaluated on 2001-2005 data sets collected at the AMIS Plus national registry. This analysis ties up to these results, enhancing the model to reach additional findings.

1.3 Goals of this thesis

First, it aims to evaluate the capability of the AMIS model as a probability estimator rather than as a classifier. It is known that scores obtained by algorithms similar to AODE are, if successful in classifying, not generally good estimators. Technically speaking, this corresponds to the question wheter the score obtained by AODE is not only a good ranker , but also a good probability estimator.

Especially, AMIS shall be compared to the TIMI model which is one of the standard models for risk stratification for myocardial infarctions and which is widely applied in the field.

Second, additional attributes which are available for a subset of the AMIS patients shall be analysed as to whether they could serve as a basis for constructing a new model incorporating long-term mortality. To this point, the AMIS model takes into account hospital or short-term mortality only.

1.4 Structure of the Thesis

- The remainder of this section will introduce some notation which will be used consistently in the following.
- Section 2 elaborates on some fundamental properties of the AMIS and the TIMI model as well as the AMIS data set.
- Section 3 presents methods to evaluate probabilistic classifiers in terms of their accuracy to predict probabilities.
- Section 4 applies these methods to AMIS and TIMI.
- Section 5 considers techniques to improve the accuracy of predicted probabilities by AMIS.
- Section 7 analysis the Triemli data set with the prospect of using it for a long-term mortality model.

[2] 'AMIS Plus' (Acute Myocardial Infarction in Switzerland) is the nation-wide registry of myocardial infarction patients in Switzerland, hosted at the Institute of Social and Preventive Medicine of the University of Zurich.

1.5 Notation

Some standard notation will be used throughout the whole thesis:

- For each patient P, the AMIS model produces a prediction score s. When different scores belong to different patients 1, ..., n, s_i means the score calculated for the i-th patient P_i.
- The actual survival of a patient is denoted by m (m_i). It is 1 if the patient survives and 0 if the patient dies. To be consistent, m's will generally be used for 'effective outcomes', even when mortality is not directly addressed.
- The TIMI score of patient P_i is denoted by τ_i.
- The subset of all patients with a certain TIMI score is denoted by $T_x := \{P_i | \tau_i = x\}$, e.g. $T_0$is the set of all patients having TIMI score 0.
- a and b are normally used for lower and upper interval limits.
- $E(x)$ denotes the expected value of some random variable x.
- σ^2 denotes the variance, $\hat{\sigma}^2$ the variance of the sample of some random variable x.
- χ^2_d means a Chi-Square-Distribution with d degrees of freedom.

2 AMIS and TIMI

In this section, we will elaborate on some fundamental properties of the AMIS as well as the TIMI predictor. The latter will have to be considered when defining a suitable evaluation methodology.

2.1 Anatomy of the AMIS model

In this section, fundamental properties of the AMIS model and its underlying AODE algorithm will be discussed.

2.1.1 AMIS Score

The AMIS model bases on AODE. For each patient P_i having attributes a_1, ..., a_n, the AMIS model delivers a score $s_i = s(a_1, ..., a_n) \in [0, 1]$. In contrast to other methods (as an example, decision trees would have the predicted class as a result), the model delivers a *score*. It will therefore be called a *scoring* classifier.

Now, there are prinicipally two possible uses of that score:

- Using a threshold $\theta \in [0, 1]$, the score obtained from a patient can be used to classify the patient into the classes 'will probably survive' and 'will probably *not* survive'.
- The score s_i can also be interpreted as the *probability* of survival of a patient. Then, s_i corresponds to the relative frequency of survivals found among all patients with that particular score.[3]

Classifiers of the first kind will be referred to as *rankers*, those of the second kind as *probability estimators*.

Scoring classifiers can principally be used as either or both of those. Accordingly, if a score is used both as a ranker and a probability estimator, we need to distinguish between its ranking capabilities its probability accuracy (also called calibration). As will be shown in the course of evaluation, this distinction is crucial.

The analysis in [Hunt, 2006] (whose results could be consistently reproduced in the course of this study) found that the ranking capabilities of the AMIS model are very good.

2.1.2 Naïve Bayes and AODE

The basis of AODE is Naïve Bayes which is described in most machine learning or data mining books (e.g. [Witten/Frank, 2005]); in our context, it is noteworthy that the independence

[3] This is consistent with the frequentist notion of probability: When tossing toins, the probability of 'heads' is 50% because the relative frequency of heads approximates 50%.

assumptions which characterise Naïve Bayes approaches has impacts on the predictive capabilities of our score. Naïve Bayes makes the assumption that its input variables are conditionnally independent, i.e. that for two attributes a_i and a_j we can say $p(a_i, a_j) = p(a_i) \cdot p(a_j)$. This drastically simplifies the calculation of the score. However, since this assumption is not justified on many real-world data sets, this comes at the cost of inaccurate probability estimates.

On the other hand, despite being inaccurate as a probability estimator, Naïve Bayes has been shown to be powerful as a ranker. This is consistent with the good ranking performance of AMIS mentionned above.

This will be picked up again during the evaluation and the improvement of the score. For the moment, it shall be noticed that the score cannot be interpreted as an estimate for probabilities without further investigation.

2.1.3 AMIS Parameters

AMIS is based on the following parameters (copied from [Hunt, 2006], p. 27):

Variable	Description
`age`	Age
`systbp`	Systolic Blood Pressure
`hrtrate`	Heart rate
`killip`	Killip classification
`cprarr`	Cardiopulmary resuscitation prior to arrival to hospital
`cmcardin`	Comorbidity: Cardiac insufficiency
`cmcevdis`	Comorbidity: Cerebrovascular disease

Table 2.1: Parameters of the AMIS model

2.2 Data

2.2.1 Data sets

There are two datasets used: Firstly the original AMIS dataset (=: D), secondly the Triemli dataset (=: T), which is a subset of AMIS but contains additional information. In particular, they contain information about long-term rather than short-term survivability which will be of importance in Section 7.

As in [Hunt, 2006], records older than 12/2001 were discarded. The term 'training cohort' (=: D_T) will henceforth be used to denote this reduced dataset. The cleaning and preparation of the AMIS dataset were adopted from [Hunt, 2006] with minor enhancments (cf. Appendix B.1). The term 'the data set' will always refer to the cleaned and prepared data set.

2.2.2 Score Distribution

Irrespective of effective mortalities in specific risk classes, the prediction score itself has an underlying distribution: Basically, the scores are skewed toward a very high survival rate, which reflects the low mortality of MI patients in general.

The following figure shows histograms of the scores obtained from a 10-fold cross-validation on D_T for different intervals of the score:

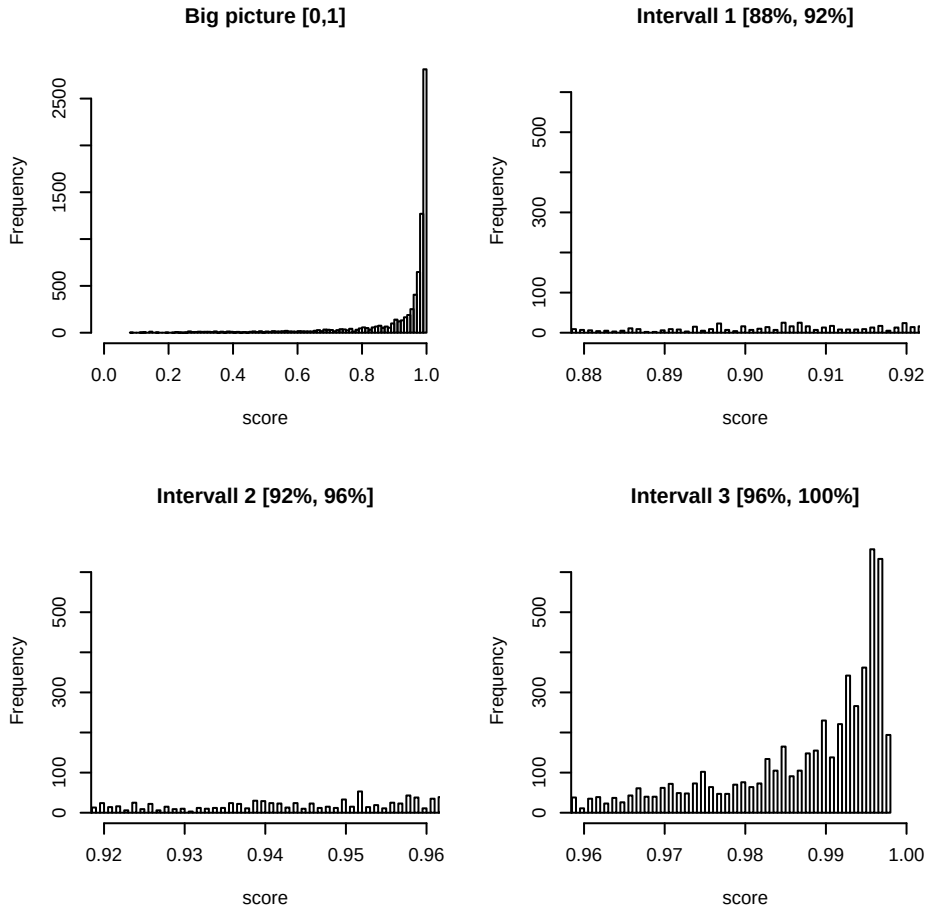


Figure 2.a: Histogram of the score (Frequency: number of patients which fall into the respective score bins).

In other words: We have a lot of patients with very high survival rates, but very few patients with low survival rates.

2.2.3 Context of the data

The data based on which the TIMI model was derived has a different context than the data used for developing the AMIS model. Most importantly, the clinical trials used in TIMI embrace

patients admissed between 1994 and 1998, where PCI was not commonly applied. However, this has changed during the past few years, as the following figure depicts:

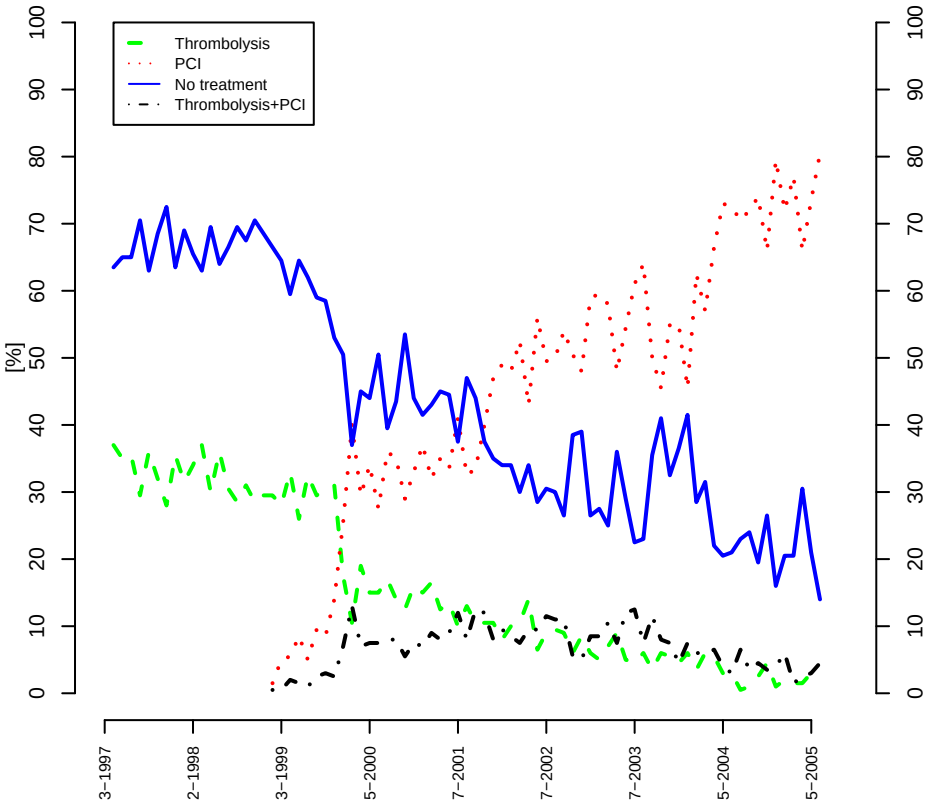


Figure 2.b: Proportions of patients treated with PCI, Thrombolysis, No Therapy or Both therapies, over time.

The amount of PCI patients has been steeply increasing since the late nineties (and the amount of Thrombolysis patients decreasing).

2.3 TIMI

TIMI for ST-elevated myocardial infarction is a widely used model for risk stratification for MI patients ([Morrow et al, 2000]).

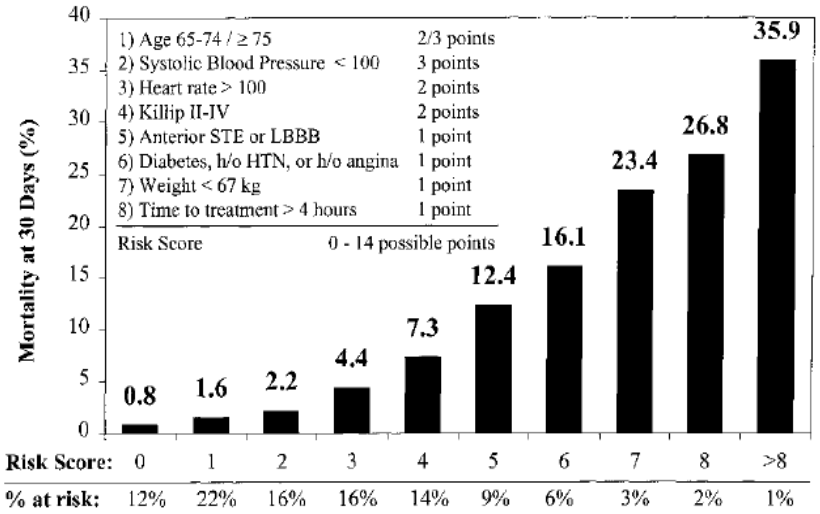


Figure 2.c: TIMI model

As can be seen in the figure, it can easily be calculated for a patient by summing up the points, depending on the values of the attributes. The sum of points corresponds to the TIMI class (sums greater than 8 are commonly summarised in the class '>8'). Each class has a mortality rate prediction assigned. For instance, if a patient is classified as TIMI 2, his/her corresponding mortality risk is assessed at 2.2%.

In contrast to the AMIS model, where a continous score between 0 and 1 is retrieved for each patient, the TIMI rules directly provide a discrete classification into different risk groups which have a discrete range of associated mortality rates (0.8%, 1.6%, 2.2%, etc.).

2.4 Summary

The main differences between AMIS and TIMI can be summarised as follows:

- AMIS is a continuous, probabilistic 2-class ranking classifier whose model is not explicitly available and which is trained on up-to-date data;
- TIMI is a discrete, multi-class classifier and probability estimator coming with explicit calculation rules (point system) readily to hand, which have been designed as to meet out-of-date fitting data.

These characteristics will have to be considered in the course of the evaluation.

3 Evaluation Methodology

In this section, a set of evaluation techniques, along with statistical procedures, will be presented in order to analyse the accuracy of probability estimates delivered by a continuous scoring classifier.

Firstly, diverse existing work which could be useful for our purpose, is discussed. We will then see that it is necessary to discretise continuous scores appropriately, before applying the methods. Also, some statsitical effects have to be considered. We then present the methods to be used.

3.1 Related Concepts and Literature

Various existing approaches found in similar problem domains shall be first presented. These can be found in different contexts.

3.1.1 Linear Regression Analysis

The correlation between effective and predicted probabilities can be analysed. Correlation plots can be used for visulisation. The strength of correlation can be quantified using PEARSONS' correlation coefficient.

3.1.2 Goodness-of-fit tests in Logistic Regression analysis

The traditionnal alternative to probabilistic classifiers is logistic regression analysis. There exist a couple of evaluation techniques for assessing the 'goodness-of-fit' of a given logistic regression model, i.e. the quality of predictions compared to actual outcome:

The BRIER score essentially[4] corresponds to the mean squared error of predicted probabilities against actual outcomes. [Murphy, 1973] provides a decomposition of the BRIER score into a discrimination (scoring capability) and calibration (probability estimating capability) component.

The HOSMER-LEMESHOW statistic ([Hosmer/Lemeshow, 1989]) is an overall measure for the goodness-of-fit of a model applied on some data.

3.1.3 Common Error Measures

[Witten/Frank, 2005] suggests to evaluate the performance of a classifier by calculating the *quadratic loss function* (QLF). Standard tools for measuring error are the well-known mean squared error (MSE) or mean absolute error (MAE) and its modifications (root mean squared error, relative squared error, relative absolute error).

[4] There exist different versions of the formula (one of the for instance compares against other predictive models rather than actual outcomes). We will adhere to [Vinterbo/Ohne, 1999] if not explicitly stated otherwise.

3.1.4 Common Datamining Evaluation Tools

In the area of data mining, diverse evaluation techniques are available: Precision/Recall-curves, ROC analysis, lift charts and many more.

3.1.5 Model Discrimination

Some work deals with methods enabling to discriminate models to optimally fit a given data set. [Aktinson, 1969] discusses different approaches, one of which is the WILLIAMS-KLOOT test, which tests wheter one model fits the data significantly better than does the other. A case-study similar to our problem, though in the area of cristallography, demonstrates the use of the test ([Prince, 1982]).

3.1.6 Discussion

Common accuracy or specificity measures based on true/false positive/negative rates will not help to evaluate the prediction capability in terms of probabilities: Those measures, among them the ROC curves, analyse the relationship between the false negative rate (instances *not* classified which actually *do* belong to the class) and the false positive rate (instances *classified* which acutally do *not* belong to the class) in a strictly discrete classification using a threshold θ, as explained earlier. Lift charts and similar techniques emphasise on cost/profit considerations, which are not applicable in this form, here. The common data mining evaluation tools will thus not help.

The BRIER score as well as error loss functions can all be subsumed under the category 'error measures' which will be considered in the evaluation, though using the more manifest terms mean squared error, mean absolute error and mean error.

The HOSMER-LEMESHOW tests will be considered, especially because they are particularly intuitive when used in conjunction with predicted versus effective mortality plots, which will be introduced later in this section.

3.1.7 Discrimination versus Evaluation

Evaluation, in this context, refers to analysing if predicted probabilities approximate the effective probabilities well or badly. Discrimination, on the other hand, means to analyse if a model A approximates those probabilities better or worse than a model B. In our setting, we are interested to evaluate AMIS and to discriminate between TIMI and AMIS.

With one exception (the WILLIAMS-KLOOT statistic), all methods can be used for both to *discriminate* between models and to *evaluate* models. Therefore, discrimination and evaluation methods are both covered side by side in this section. Similarly, if some results can be calculated for both models (such as error rates), we will do so.

3.2 Discretisation

This section first shows why discretisation is needed. The following aspects will be discussed:

- On which quantities shall we discretise?
- What is the predicted probability of a bin?
- How does the density (many instances in low-risk bins and few in high-risk bins) affect the discretisation and how can it be taken into account.
- Having the above in mind, what are appropriate discretisatisation approaches?

3.2.1 The Need to Discretise

The obvious would be to plot predicted mortalities against effective mortality. However, very few instances lead to exactly same prediction score, such that the group of patients having a specific score is often limited to 1 or 2 patients. This leads to bins with effective mortality rates of either 100% or 0%, depending on the effective survival of exactly those patients (or to a range of discrete rates, depending on the number of instances which happen to deliver the same score).

Hence, the choice of an approapriate discretisation is crucial.

3.2.2 The Advantage of Risk-Based Discretisation Criteria

As a thought experiment, consider the following approach: Theoretically, we could discretise patients into arbitrary groups and evaluate the model within these groups: As an example, patients might be binned according to their shoe sizes and we would evaluate the score within the group which has size 37-38, 39-40 and above 40. Intuitively, it is clear that such a discretisation is not very useful. Technically, within such bins, the high variance of the score alone will make the mean value uncertain as an estimate for the expected score. Consequently, such a 'non-sense' discretisation will never lead to statistically reliable statments about the score, let alone to comparisons based on it[5].

Therefore, we will favour a discretisation which leads to low score variances within its bins. Naturally, such discretisations will be based on either the score itself or any attribute which is reasonnably correlated to the score. Predicted probabilities of other models, such as the TIMI mortality in our case, fulfil this requirement.

We will therefore concentrate on the following types of discretisations:

- those based on the AMIS score;

[5]Of course, this tacitly assumes that shoe size is not actually correlated with risk in MI patients, which could not be analysed for lack of data, but seems very probable.

- those based on the TIMI class.

They will be explained in the remainder of this section.

3.2.3 Predictions Associated to Bins

Assume we had a discretisation: By default, what we have are prediction scores for instances, not for bins *containing* instances. What is the prediction score that we associate to bins? The question is not so trivial.

Theoretically, the assignment of a prediction score to a bin is somewhat arbitrary and could be any function of the scores falling into the bin. Actually, by discretising, we wrap a new prediction model around the score, by specifiying which scores are associated to which class (or bins). Put differently, by discretising the scores of a continous predictor, we are in fact discretising the predictor as such. Prediction values associated to these classes should then principally be free to choose.

Considering this, the initial question relates to the design of the discretised prediction model. In our context, the aim is to infer from evalution results which have been obtained from a discretisation of the orignial predictor. It is therefore desireable to have a discretised predictor which best represents the behaviour of the originial continous predictor.

For our bins, this means that the prediction value associated to them shall be most representative for the continuous scores to be found within the respective bins. Thus, if 1000 of the scores falling in a bin are around 99% and just a dozen around 98%, we better associate a prediction value of 99% to that bin than one of 98%.

When bins are not too big – which we will also ensure for other reasons – an appropriate measure for bin probability is the expected value of scores falling into the bin. These will be estimated using the arithmetic mean of the scores within the bin. Therefore, the expected TIMI prediction will be

$$\hat{E}(\tau_B) = \frac{1}{|B|} \sum_{P_i \in B} \tau(P_i) \tag{1}$$

for patients P_i in bin B (which contains $|B|$ elements), and accordingly for AMIS

$$\hat{E}(s_B) = \frac{1}{|B|} \sum_{P_i \in B} s(P_i) \tag{2}$$

3.2.4 Number of Patients within Bins

The following effects emerge when bins contain too few instances. They are closely related.

Bouncing Probabilities For the purpose of illustration, this shows a percentiles-based discretisation with 50 bins. The figures show the effective and predicted probabilities on D_T:

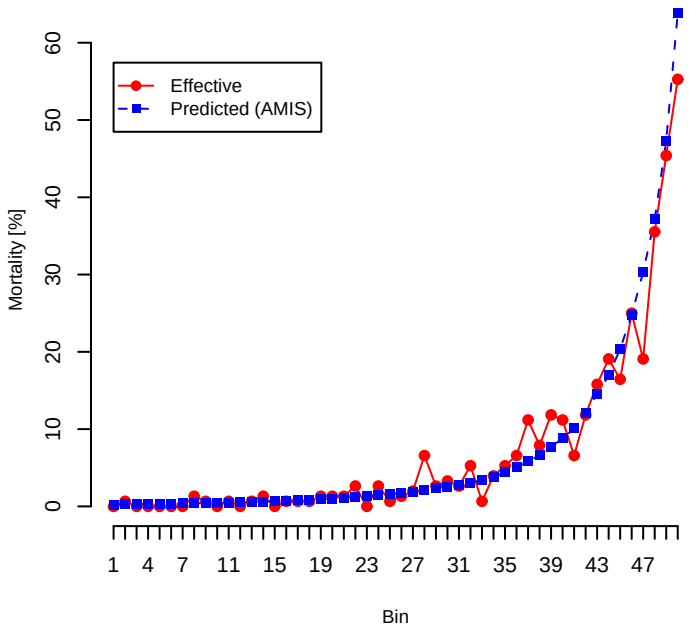


Figure 3.a: Bouncing mortalities in a 50-bins percentiles-based discretisation on AMIS (bins 1-50)

The bouncing probabilities appear not only at the high-risk side, as the following 'zoomed' version of the diagram shows for bins 1-38. There are adjacent bins which differ more than a couple of percentage points from each other:

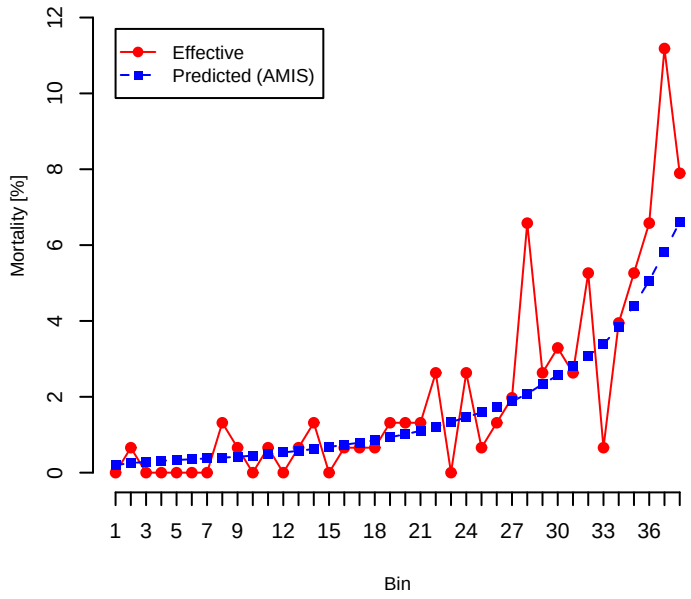


Figure 3.b: Bouncing mortalities in a 50-bins percentiles-based discretisation on AMIS (bins 1-38)

We can see that effective mortality rates within the bins 'bounce', due to two effects:

- Intervals contian few patients such that only a few negative cases can cause the probability to jump up (explains points well above the predicted mortalities).
- As most patients survive, negative outcomes are seldom and there exist many bins, especially in the low-risk region, in which only positive instances are present. In those, probability bounces back to 0.

Statistical Confidence Statistical confidence, be it in the form of confidence intervals for a measured quantity or in the form of significance levels when testing hypothesis about one or more of those quantities, depend on the number of observed instances available. The more instances that are considered in a sample, the more reliable are statements about some of their parameters. It is more certain to believe the expected mortality to be at 50% if 50 out 100 patients die than if 5 out of 10 did. This holds irrespecitve of the specific method used.

The number of patients falling into the individual bins of any discretisation is therefore crucial in order to enable statistically sound interpretations based on them.

On the other hand, statistical significance depends on the magnitude of the quantities themselves: it requires many more instances to be confident about a mean value of 0.005% than to be confident about one of 50%. In terms of variance, small deviations on small quantities are obviously more harmful as small deviations on bigger quantities.

3.2.5 Discretisation Approaches

There are many different approaches to discretise based on the AMIS score, which will be discussed in the following.

Fixed-sized intervals A practionner's approach would be to divide the score in, say, 10 intervals (possibly to be comparable to the TIMI score which goes from 0 to 8 and '>8'). Interval limits can be chosen equidistant (e.g. [100%, 99%],]99%, 98%], ... for patients with scores between 0.99 and 1, those between 0.98 and 0.98, etc.).

However, this leads to poor resuls: If distances are small, there are too few instances in the low score bins. If they are chosen too big, most of the data set's instances fall into the first couple of bins and the rest is virtually empty. As explained in Section 3.2.3, it is not viable to interpret the interval [100%, 99%] as containing predictions of at most 1% mortality, or of 99.5% mortality, or whatsoever. This interpretation is only viable in the case where the predictions within the bin are uniformly distributed, which is not the case for AMIS' low-risk bins and only approxiamtely so for the higher-risk bins.

Interpretation of interval limits More sophisticatedly, and for reasons of comparability to the TIMI model, intervals can be chosen as to represent the same predicted mortalities as the TIMI score. To that end, one might define intervals $I_i = [a_i, b_i]$ such that $(a_i + b_i)/2 = t_i$(where t_i is the TIMI prediction for class i).

Again, the interval limits $[a_i, b_i]$ cannot be interpreted as an approximation of the average prediction value associated to that interval. Therefore, interval arithmetic of any sort will not lead to a sensible association of predicted values to bins.

Parametric approach Considering the issues discussed above, one could take into account the score distribution by estimating parameters of an appropriate, common probability distribution $X(p_1, ..., p_n)$ (such as the normal distribution, beta distribution, or similar). Having estimated those parameters, the expected values could be easily obtained using the integrals of the respective density function (say $D(b)$), which are readily available for the common distribution families. Conversely, using some pre-defined mortality rates (say m_i), we could obtain the corresponding interval limits from $b = D^{-1}(m_i)$.

Non-Parametric approach The predictions can be sorted in descending order, and bins then chosen as to contain a fixed amount of n subsequent scores[6]. By holding counts constant, the density of predictions within an intervall is automatically considered.

[6]Of course, there will be some 'modulo problems' as the amount of scores available will in general not be dividable by n – but they can be overcome by cutting some scores off, by smart choice of n or similar.

3.2.6 Preferred Approach: Percentile-Based Discretisation

The above section demonstrated that the only useful approaches to discretisation are those who, either explicitly or implicitly, take into account the density of the score distribution (these are the last two explained above, entitled Parametric and Non-Parametric). To avoid density estimation, the preferred approach is a non-parametric one.

To discretise the set D into n bins, the procedure is thus as follows:

1. Lets denote the number of instances in D by $|D|$. The $|D|$ predictions are sorted in descending order according to their scores, which gives us a list (P_1, ..., P_2).
2. The first $\frac{|D|}{n}$ instances are associated to bin number 1, i.e. $B_1 = \{P_k | 1 \leq k < \frac{|D|}{n}\}$, the second $\frac{|D|}{n}$ to bin number 2, and so on, or in general: $B_b = \{P_k | k > (b-1)\frac{|D|}{n}+1 \leq k < \frac{|D|}{n}\}$ for bin number b.
3. The interval limits for the correspond to the scores delimiting the bins, i.e. the interval of bin number b can be written as $[\min_{P_i \in B_b}(s(P_i)), \max_{P_k \in B_b}(s(P_k))[$.

Discretisations obtained using this method will be referred to as percentile-based. This is more suggestive and reflects the fact that the upper limits of the bins are in fact the $1/n$ percentiles: for instance, in a 10-bins-discretisation, the first bin will go from 1 to the first decile, the second bin from just above the first decile to the second decile, and so on.

3.2.7 Discretisation Based on the TIMI class

Since the TIMI class already comes as a discrete quantity, discretisation based on TIMI is trivial: the bins simply contain all patients which have a specific value in their TIMI attribute. However, in accordance to most studies including the original TIMI paper ([Morrow et al, 2000]), we will add a slight modification to this by summarising all patients with $t > 8$ into a '$>$8' bin.

3.2.8 Significance Issues at the Extremes

Even when a suitable discretisation is chosen and the above effects were taken into account, there remain problems at the extremes of the score. As mentionned in Section 2, the distribution of scores is heavily skewed toward very low mortality rates. This has the following impact on significance:

- Intervals on the low-risk side of the score will contain many patients, but will assume very low expected mortality rates (both predicted and effective). Significance is more difficult to achieve when very small quantities are involved.[7]

[7] In the case of proportions, much bigger sample sizes are required to be confident about low means (e.g. to confirm a prevalence of a disease with $p = 0.0001$) than about higher means (say a disease with $p = 0.1$).

- Intervals on the high-risk side of the score will contain few patients, but will assume higher mortality rates.

This shows that it will be very difficult to assess the accuracy of the probabilistic models in those extreme regions of the score. For the Predicted-Effective Plots, this will be an issue.

3.3 Statistical Considerations

3.3.1 Population, Samples and Estimates

Statistics consists essentially of two things: 1. draw samples from populations; 2. calculate estimates to get an approximation of some real paramter from the population. The general problem when estimating parameters based on samples is that we do not know as to what extent their values are due to chance (variance) or to systematical bias determined by the selection of instances in the sample. We need therefore to consider the distributions of the random variables which lead to the parameter. To that end, we will analyse which samples are drawn from which populations, and which parameters are to be inferred by which estimates.

What are the Parameters? In the variouos evaluation approaches which will be carried out in the following, the parameter of interest will be:

- The effective probability of an outcome (e.g. dead or alive) within bins of the AMIS score, estimated by the arithmetic mean of the dichotomous variable $m_i$within the bin;
- The predicted probabilities assigned to a bin, i.e. the expected value of the scores within the bin (cf. Section 3.2.3, estimated by the arithmetic mean of scores s_i found in the bin;
- The loss functions used in the error measures for specific bins, estimated by the arithmetic mean of loss functions[8].

What is the Population? Theoretically, one could argue that the basic population we are interested in is the set of all possible myocardial infarction patients potentially assessed using the AMIS or TIMI model. The AMIS set D (full cohort) only represents a sample out of those. Obviously, this view is not viable. It would be difficult to define the population AMIS is based on. Furthermore, even after many exclusions[9], the size of the population would certainly be too big and AMIS too small a sample of it, such that statistically significant inference might be impossible.

[8]Statistically, mean absolute error, mean squared error and mean error are estimates for the absolute, squared or 'raw' deviation (='loss function'). For instance, mean squared error is defined as $\frac{1}{n}\sum_i(s_i - m_i)$, i.e. the arithmetic mean of the loss function $(s_i - m_i)$. Therefore, we need not estimate those error measures themselves, since they are already estimates (of the loss function) by definition.

[9]Otherwise, roughly 20% of the population would consist of Chinese people; as well we might exclude the San people from the Kahalari desert, people living on the Mongolian Plateau, and so on.

D_T is the Population! The problem outlined above is a general problem in data mining and in statistics. Actually, it has been solved when the AMIS model was developed: D_T was chosen as to best represent the characeteristics of patients which to whom the model will be applied. We will therefore define D_T as our population. The individual cross-validation runs are then where samples are drawn.

3.3.2 Normality Assumptions

For the statistics used in this thesis, it is important that the random variables approximately follow a normal distribution.

180 independent stratified cross-validation runs were performed on a 6-bins percentiles-based discretisation of the AMIS score. Histograms of the estimates caluculated on the individual bins look all reasonnably normal in shape (cf. Appendix A.2). Furthernore, P-values of KOLMOGOROF-SMIRNOV tests were sufficiently high, such that the null-hypothesis 'follows a normal distribution' cannot be rejected.

It can therefore be assumed that the quantities of interest approximately follow normal distributions.

3.4 Predicted/Effective Plots

Predictive/Effective plots show probabilities on the y-axis and bins on the x-axis. The red (solid) line shows the mean survival value, which is the estimate of the effective probabilities observed within the respective bins.

The other lines show the mean prediction scores of the individual models. Bin probabilities are calculated as explained in the discretisation section. A sample plot is shown in Figure 3.c, where predictor 1 tends to over-estimate and predictor 2 to unter-estimate the effective probability:

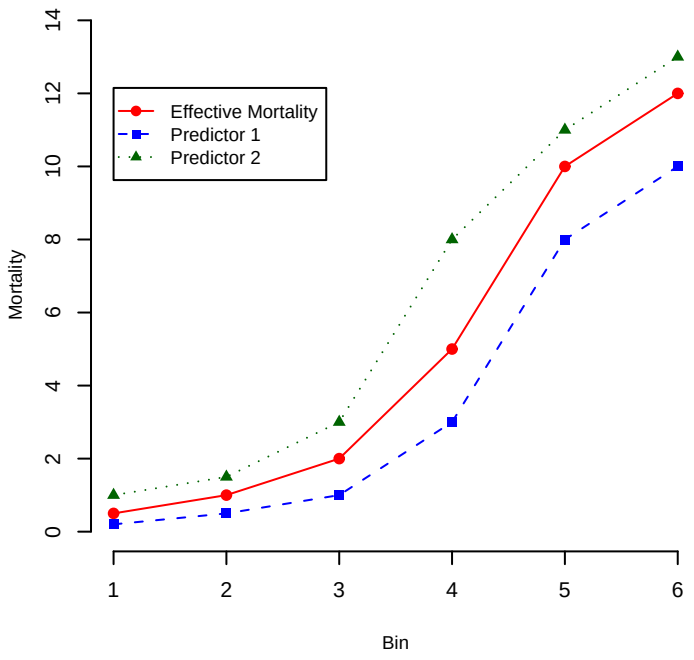


Figure 3.c: Sample Predicted/Effective Plot of two models

The plots are valuable for getting a visual impression of how accurate the probabilities of the respective models are.

A statistical test which underlines the bin-wise comparison of effective and predicted probabilities is the HOSMER-LEMESHOW test commonly used in logistic regression ([Hosmer/Lemeshow, 1989]).

The associated P-values can be interpreted as a measure for fit: the higher the P-value, the better the fit of the model. More can be found in Appendix A.1.

3.5 Correlation Plots

Correlation plots are scatterplots which consist of n points $(\bar{s}_i, \bar{m}_i)$ for each bin $i\,(1 \leq i \leq n)$. Therefore, the better a model's curve fits to the linear red (solid) curve representing the perfect correlation (consisting of points $(\bar{m}_i, \bar{m}_i)$), the better the predictions are.

The natural statistic that goes along with correlation plots is the correlation coefficient. The standard coefficient by PEARSON R will be used (see statistical appendix). An $R > 0.9$ is commonly assumed to indicate strong correlation.

3.6 Error-Based Evaluation

There are plenty of error measures which can be calculated:

- Mean Absolute Error
- (Root) Mean Squared Error
- Mean Error
- Relative Absolute Error
- Relative Squared Error

3.6.1 Measuring Correctness of Predictions

It is important to point out what an error means in our context and how the different error measures will have to be interpreted accordingly.

As an example, consider 10 patients which are delivered to a hospital and their mortality risk assessed by a model. Lets assume that the model predicts precisely 90% survivability for each of the patients, and that exactly one patient dies and the rest (9 of them) survive.

Two different interpretations of the error in this case are possible:

- One could argue that the model's predictions were perfect. The mean prediction within the bin consisting of those 10 patients corresponds to the effective survival proability of 90%. This is consistent with the approach chosen in the predicted-effective plots and the correlation plots discussed earlier, as in such a bin, the model's curve and the effective curve would be congruent.
- On the other hand, for the patient who obtained a 90% prediction and actually dies (that is, $y_{10} = 0$), the prediction is obviously catastrophic. It seems therefore reasonable to assign to it an error of $|y_{10} - s_{10}| = |0 - 90\%| = 90\%$ (or $(y_{10} - s_{10})^2$ for the squared error).

These two perspectives must be distinguished. They explain why the models show a good fit in the predicted-effective plots and correlation plots, but are doing badly in comparison, when absolute or squared errors are considered. This is consistent with the mean errors which are very low in comparison with absolute and squared error.

3.6.2 Error Measures

The following error measures will be calculated. For visualising, they are displayed in bar charts where a bar corresponds to a specific model's error within a specific bin. Overall error rates will be indicated as well.

Mean Absolute Error

$$\frac{1}{n}\sum_{i=1}^{n}|s_i - y_i| \tag{3}$$

Mean Squared Error

$$\frac{1}{n}\sum_{i=1}^{n}(s_i - y_i)^2 \tag{4}$$

Mean Error

$$\frac{1}{n}\sum_{i=1}^{n} s_i - y_i \tag{5}$$

As mentionned, this error corresponds to the differences of the mean values $\bar{s}$ and $\bar{y}$:

$$\frac{1}{n}\sum_{i=1}^{n} s_i - y_i = (\frac{1}{n}\sum_{i=1}^{n} s_i) - (\frac{1}{n}\sum_{i=1}^{n} y_i) = \bar{s} - \bar{y} \tag{6}$$

If the error is positive, the model tends to over-estimate, otherwise to under-estimate in its prediction. Of course, this does not hold in the case of mean absolute and mean squared error, where only positive values result.

3.6.3 Statistics

When comparing error rates of different models, the t-test for paired measurements (cf. Appendix A.1) will be used to show the significance of the differences.

The P-value of the t-statistic can be interpreted as follows: P quantifies the probability that the differences in error rates are due to chance alone. Conversely, if the P-value is very low (e.g. $P \approx 0$), we can conclude on a $1 - P$ confidence level that the differences are significant.

4 Evaluation

The methods presented in the previous section shall now be applied to the AMIS training cohort D_T. After some introductory comments about different subgroups of D_T and related to the mechanism used to produce the predictions, the following evaluations are carried out:

- An evaluation of AMIS and TIMI in an 8-bins percentile-based discretisation on the AMIS score;
- an evaluation of AMIS and TIMI in a discretisation based on the TIMI class.

4.1 Introduction

4.1.1 Data Cohorts

The evaluation techniques will be applied to a couple of different subsets of the AMIS data set, according to the following levels:

- STEMI / Non-STEMI patients
- Treatment subgroups (patients treated with PCI, Thrombolysis)

If not indicated otherwise, the charts and statistics are always based on the training cohort D_T. When subgroups are analysed, the criteria of the different levels are annotated (e.g. 'STEMI' means all STEMI patients within D_T were considered).

The levels can be combined, e.g. 'all PCI patients within the subgroup of STEMI patients'. This example would be denoted as 'STEMI, PCI'.

Whenever TIMI is directly compared to AMIS, only STEMI patients are considered, as the TIMI for STEMI model that we are interested in claims to be a predictor for those only.

4.1.2 Cross-validation

The training cohort D_T and 10-fold-crossvalidation were used similar as in [Hunt, 2006]. The scores are obtained programmatically[10], with one small extension to the standard WEKA approach. WEKA follows the approach outlined below:

1. The sequence of patients in D_T is randomly reordered.

[10]The corresponding JAVA classes along with their documentation is contained on the CD accompanying this thesis. Appendix [**] contains an overview of the code used.

2. Each fold[11] (say D_{F1}, ..., D_{F10}) is stratified such that the target variable has the same distribution in the folds as in the whole data set.

3. Should the instances in the training cohort contain attributes which are not part of the model, i.e. which are not used as input variables, they must be removed. Let $\hat{S}$ denote that fact that the irrelevant attributes have been removed from set S, then the cross-fold validation is performed on $\hat{D_T}$ rather than on D_T.

4. In each validation run (1..10), the training fold $\hat{D_{Fi}}$ and the test fold $\hat{D_{Ei}} := \hat{D_T} \backslash \hat{D_{Fi}}$ are selected from the training cohort.

5. For each training fold $\hat{D_{Fi}}$ the model is learned using and applied to the corresponding evaluation set $\hat{D_{Ei}}$. The result of the evaluation is a matrix M_{kl} containing the scores of each instance i for class j.

Since WEKA does not allow the inclusion of attributes in a training run which are not part of the model, they are removed in step 3. Yet, for our analysis, such attributes are required as well. For instance, we need to know what treatment a given patient has received in order to analyse predictions and outcomes given a certain initial therapy, even if the attribute treatment is not used as an input variable for the model.

The algorithm was therefore slightly modified: Step 3, which makes $\hat{D_T}$from D_T was performed on the level of individual folds rather than on the whole trainig cohort, i.e. the folds D_{Fi} and D_{Ei} were selected from D_T. Then, before the test, the D_{Ei}'s set's entries were copied (deep copy) into a separate data structure (`ComplexPrediction`[12]) unknown to the WEKA algorithms. Only then were irrelevant attributes removed and D_{Ei} and D_{Ti} turned into $\hat{D_{Ei}}$and $\hat{D_{Ti}}$ to perform the test. After the test, D_{Ei} will have the same sequence of patients as $\hat{D_{Ei}}$. $\hat{D_{Ei}}$ has the same sequence of patients as M_{kl}. Therefore, M_{kl} and D_{Ei} could be joined to a set $\widetilde{D_{Ei}}$ containing all patients of the training set along with their predictions.

The resulting ten disjoint test sets $\widetilde{D_{Ek}} \subset D_T$ $(1 \leq k \leq 10)$ were then assembled to the enhanced training set, i.e. $\widetilde{D_T} := \bigcup_k \widetilde{D_{Ek}}$, which will serve as the basis for the analysis provided below.

4.2 AMIS-based Discretisation

A couple of percentile-based discretisations has been evaluated. As an example, the 8-bin discretisation is presented here. More discretisations (for 4, 5, ..., 10 bins) can be found in Appendix D.2.

[11] The folds are essentially the 10 subsequent partitions of the training data, i.e. patients $P_1..P_{n/10}$, $P_{n/10+1}..P_{2(n/10)}$ and so on.

[12] An overview of the JAVA classes is given in Appendix [**].

4.2.1 8 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

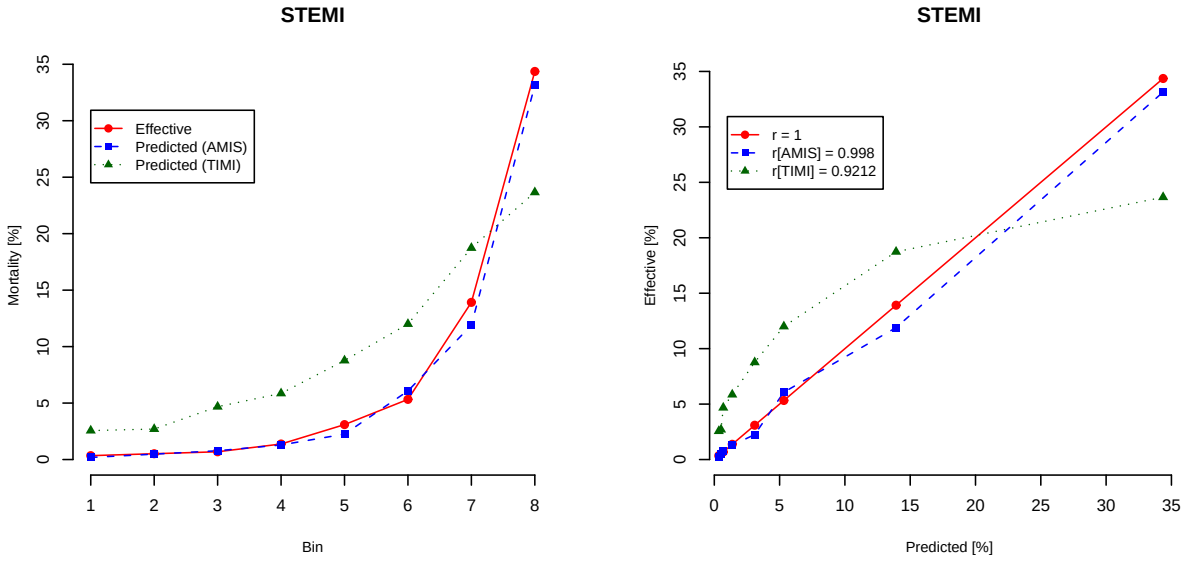


Figure 4.a: Predicted-Effective Plots for a percentile-based 8 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

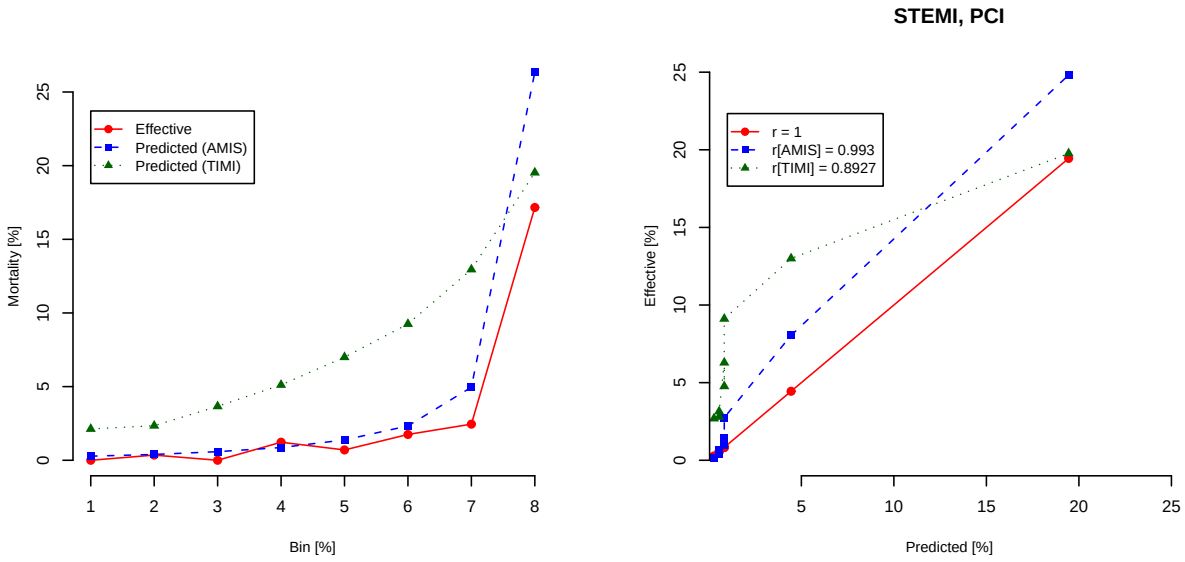


Figure 4.b: Predicted-Effective Plots for a percentile-based 8 bin discretisation based on the AMIS score.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T
1	99.85 %	99.64 %	582	0	0.29 %	2.05 %	0.0%	0.0 %	$\approx$ 0.0 %
2	99.64 %	99.43 %	582	2	0.44 %	2.43 %	0.34%	33.86 %	$\approx$ 0.0 %
3	99.43 %	99.08 %	582	5	0.72 %	3.91 %	0.86%	36.1 %	$\approx$ 0.0 %
4	99.08 %	98.39 %	582	11	1.23 %	6.38 %	1.89%	12.09 %	$\approx$ 0.0 %
5	98.38 %	97.14 %	582	19	2.18 %	8.85 %	3.26%	7.05 %	$\approx$ 0.0 %
6	97.13 %	93.85 %	582	30	4.25 %	12.63 %	5.15%	16.02 %	$\approx$ 0.0 %
7	93.84 %	81.44 %	582	83	10.94 %	18.58 %	14.26%	1.04 %	0.16 %
8	81.44 %	12.89 %	582	199	39.32 %	24.09 %	34.19%	0.38 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

$H_{TIMI} = 148.54840394177242$ (P-value 0.0 %)

$H_{AMIS} = 21.46766295980794$ (P-value 0.15 %)

Table 4.1: Statistics: Predicted/Effective mortalities for a percentile-based 8 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T
1	99.86 %	99.66 %	571	0	0.28 %	2.13 %	0.0%	0.0 %	$\approx$ 0.0 %
2	99.66 %	99.54 %	571	2	0.39 %	2.35 %	0.35%	42.94 %	$\approx$ 0.0 %
3	99.53 %	99.29 %	571	0	0.58 %	3.66 %	0.0%	0.0 %	$\approx$ 0.0 %
4	99.29 %	98.94 %	571	7	0.86 %	5.12 %	1.23%	21.59 %	$\approx$ 0.0 %
5	98.94 %	98.25 %	571	4	1.38 %	7.0 %	0.7%	2.59 %	$\approx$ 0.0 %
6	98.25 %	96.97 %	571	10	2.33 %	9.25 %	1.75%	14.62 %	$\approx$ 0.0 %
7	96.97 %	91.84 %	571	14	4.96 %	12.95 %	2.45%	0.01 %	$\approx$ 0.0 %
8	91.8 %	14.61 %	571	98	26.35 %	19.53 %	17.16%	$\approx$ 0.0 %	6.96 %

HOSMER-LEMESHOW-statistics:

$H_{TIMI} = 192.752964124033$ (P-value 0.0 %)

$H_{AMIS} = 41.06324578447627$ (P-value $\approx$ 0.0 %)

Table 4.2: Statistics: Predicted/Effective mortalities for a percentile-based 8 bin discretisation based on the AMIS score.

Error Plots

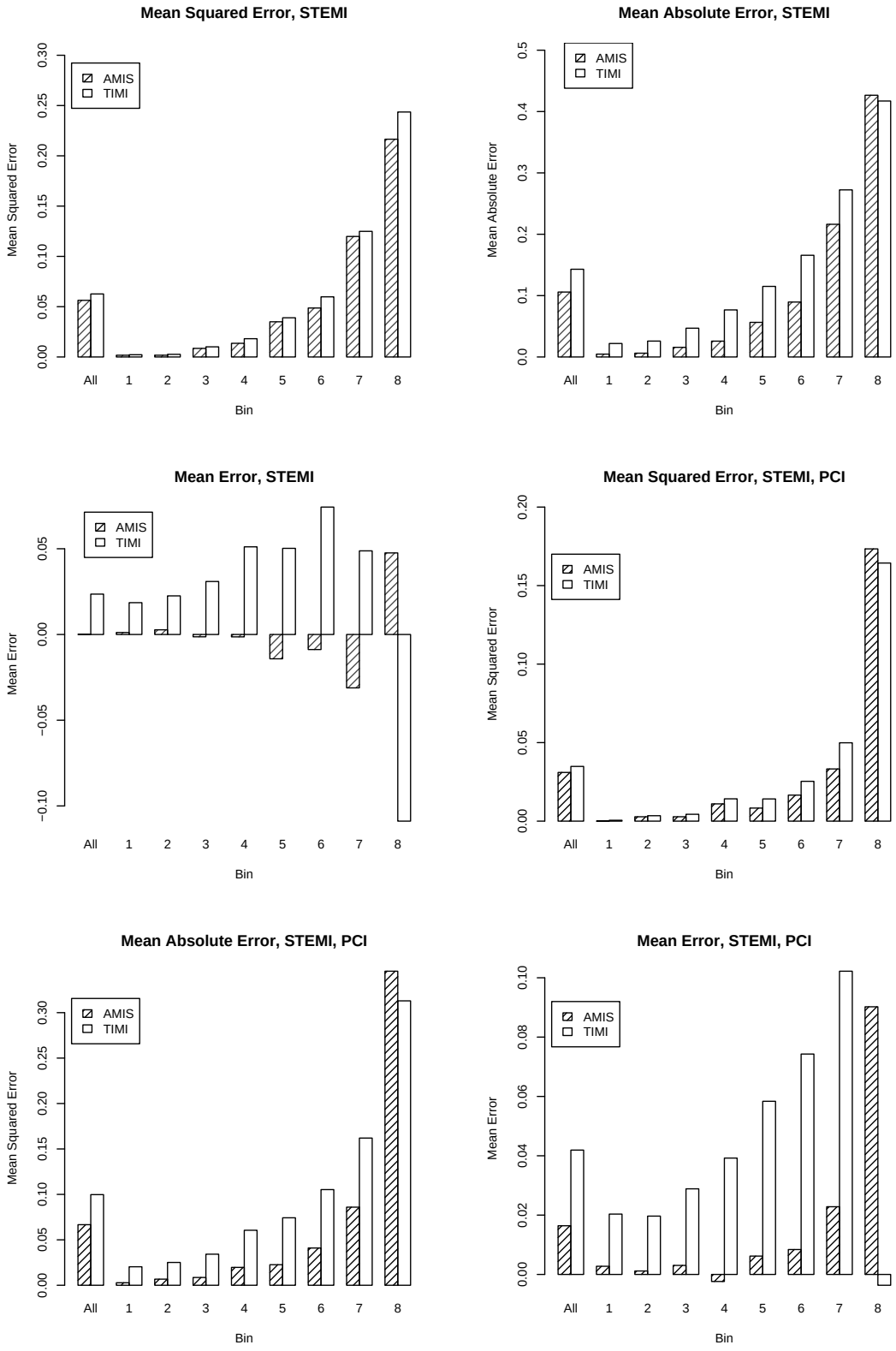


Figure 4.c: Error Plots for a percentile-based 8 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0564	0.0626	-0.0063	-4.6269	≈ 0.0 %	AMIS
1	582	0	0.0	$6.0 \cdot 10^{-4}$	$-6.0 \cdot 10^{-4}$	-18.0992	≈ 0.0 %	AMIS
2	582	2	0.0034	0.0042	$-8.0 \cdot 10^{-4}$	-7.9163	≈ 0.0 %	AMIS
3	582	5	0.0085	0.0104	-0.0019	-4.1107	≈ 0.0 %	AMIS
4	582	9	0.0152	0.0191	-0.0039	-4.5174	≈ 0.0 %	AMIS
5	582	20	0.0333	0.0373	-0.0040	-2.621	0.44 %	AMIS
6	582	30	0.0487	0.06	-0.0113	-5.8235	≈ 0.0 %	AMIS
7	582	84	0.122	0.127	-0.0050	-1.4832	6.9 %	?
8	582	199	0.2163	0.2411	-0.0248	-2.557	0.53 %	AMIS

Table 4.3: Mean Squared Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score.

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1057	0.143	-0.0373	-24.6364	≈ 0.0 %	AMIS
1	582	0	0.0029	0.0209	-0.018	-36.1174	≈ 0.0 %	AMIS
2	582	2	0.0078	0.0268	-0.019	-24.8612	≈ 0.0 %	AMIS
3	582	5	0.0156	0.047	-0.0314	-22.1946	≈ 0.0 %	AMIS
4	582	9	0.0273	0.0772	-0.0499	-25.4849	≈ 0.0 %	AMIS
5	582	20	0.0546	0.1147	-0.0601	-24.4144	≈ 0.0 %	AMIS
6	582	30	0.0891	0.1658	-0.0766	-25.1201	≈ 0.0 %	AMIS
7	582	84	0.2181	0.2735	-0.0554	-13.7118	≈ 0.0 %	AMIS
8	582	199	0.4259	0.4151	0.0108	1.1098	13.36 %	?

Table 4.4: Mean Absolute Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score.

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	$-2.0 \cdot 10^{-4}$	0.0235	-0.0237	15.0907	0.0 %	AMIS
1	582	0	0.0029	0.0209	-0.018	36.1174	0.0 %	AMIS
2	582	2	0.0010	0.02	-0.0191	25.081	0.0 %	AMIS
3	582	5	-0.0014	0.0307	-0.0321	23.1781	0.0 %	AMIS
4	582	9	-0.0032	0.0488	-0.052	27.9262	0.0 %	AMIS
5	582	20	-0.0126	0.0536	-0.0662	30.3832	0.0 %	AMIS
6	582	30	-0.0092	0.0746	-0.0838	30.9666	0.0 %	AMIS
7	582	84	-0.035	0.0437	-0.0788	23.8472	0.0 %	AMIS
8	582	199	0.052	-0.1031	0.1551	-21.3009	≈ 0.0 %	TIMI

Table 4.5: Mean Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score.

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0309	0.0348	-0.0039	-2.869	0.21 %	AMIS
1	360	0	0.0	$6.0 \cdot 10^{-4}$	$-6.0 \cdot 10^{-4}$	-14.5864	$\approx$ 0.0 %	AMIS
2	360	1	0.0028	0.0035	$-7.0 \cdot 10^{-4}$	-5.3782	$\approx$ 0.0 %	AMIS
3	360	1	0.0028	0.0044	-0.0016	-6.1429	$\approx$ 0.0 %	AMIS
4	360	3	0.0083	0.0112	-0.0029	-4.6832	$\approx$ 0.0 %	AMIS
5	360	5	0.0137	0.0197	-0.0060	-9.3558	$\approx$ 0.0 %	AMIS
6	360	5	0.0139	0.0229	-0.0090	-8.7393	$\approx$ 0.0 %	AMIS
7	360	13	0.0357	0.0512	-0.0155	-7.5407	$\approx$ 0.0 %	AMIS
8	360	71	0.1703	0.1629	0.0074	0.706	24.01 %	?

Table 4.6: Mean Squared Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score.

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0666	0.0997	-0.0331	-19.7353	$\approx$ 0.0 %	AMIS
1	360	0	0.0028	0.0213	-0.0186	-29.3401	$\approx$ 0.0 %	AMIS
2	360	1	0.0067	0.0249	-0.0182	-20.1559	$\approx$ 0.0 %	AMIS
3	360	1	0.0085	0.0345	-0.0259	-17.3627	$\approx$ 0.0 %	AMIS
4	360	3	0.0169	0.0556	-0.0387	-19.5404	$\approx$ 0.0 %	AMIS
5	360	5	0.0279	0.0806	-0.0526	-20.467	$\approx$ 0.0 %	AMIS
6	360	5	0.0382	0.1028	-0.0645	-22.6964	$\approx$ 0.0 %	AMIS
7	360	13	0.0883	0.1626	-0.0743	-19.6458	$\approx$ 0.0 %	AMIS
8	360	71	0.3428	0.3123	0.0305	2.7875	0.27 %	TIMI

Table 4.7: Mean Absolute Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score.

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0164	0.0419	-0.0255	14.7704	0.0 %	AMIS
1	360	0	0.0028	0.0213	-0.0186	29.3401	0.0 %	AMIS
2	360	1	0.0012	0.0195	-0.0183	20.3528	0.0 %	AMIS
3	360	1	0.0030	0.0291	-0.0261	17.5917	0.0 %	AMIS
4	360	3	$4.0 \cdot 10^{-4}$	0.0398	-0.0394	20.2621	0.0 %	AMIS
5	360	5	$6.0 \cdot 10^{-4}$	0.0539	-0.0533	21.0714	0.0 %	AMIS
6	360	5	0.0111	0.0771	-0.066	24.0266	0.0 %	AMIS
7	360	13	0.0199	0.0985	-0.0786	22.2344	0.0 %	AMIS
8	360	71	0.0935	$1.0 \cdot 10^{-4}$	0.0934	-9.4271	$\approx$ 0.0 %	TIMI

Table 4.8: Mean Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score.

4.3 TIMI-based Discretisation

Plots: All Patients (STEMI)

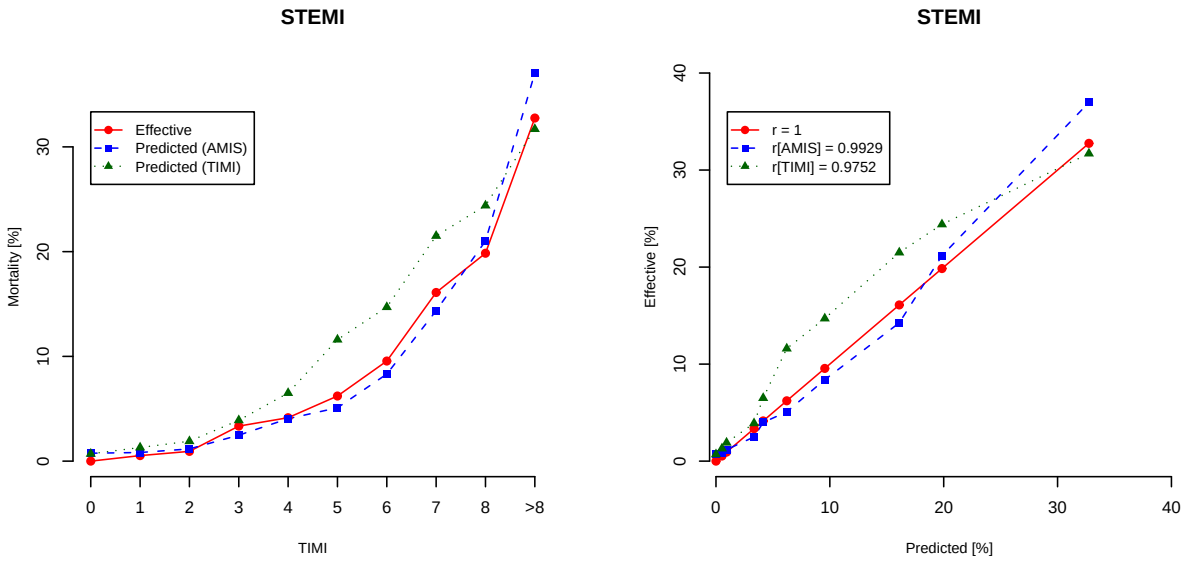


Figure 4.d: Predicted-Effective Plots for discretisation based on the TIMI score.

Plots: PCI Patients (STEMI)

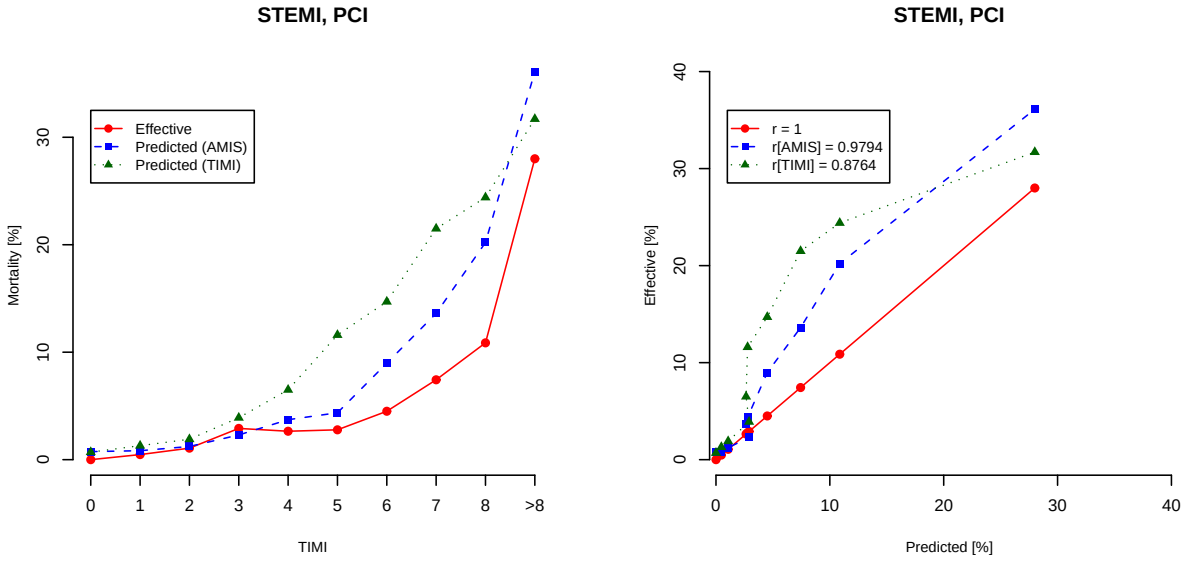


Figure 4.e: Predicted-Effective Plots for a discretisation based on the TIMI score.

Statistics: All Patients (STEMI)

TIMI	n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T
0	165	0	0.76 %	0.7 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %
1	570	3	0.82 %	1.3 %	0.53%	17.81 %	0.55 %
2	752	7	1.17 %	1.9 %	0.93%	25.65 %	0.29 %
3	628	21	2.48 %	3.9 %	3.34%	8.57 %	21.95 %
4	603	25	4.05 %	6.5 %	4.15%	45.09 %	0.2 %
5	563	35	5.09 %	11.6 %	6.22%	11.37 %	$\approx$ 0.0 %
6	450	43	8.31 %	14.7 %	9.56%	17.39 %	0.01 %
7	323	52	14.36 %	21.5 %	16.1%	18.41 %	0.44 %
8	262	52	21.02 %	24.4 %	19.85%	31.13 %	3.32 %
>8	345	113	37.05 %	31.7 %	32.75%	4.3 %	33.87 %

HOSMER-LEMESHOW-statistics:

$H_{TIMI} = 47.7402$ (P-value ≈ 0.0 %)

$H_{AMIS} = 10.3344$ (P-value 24.23 %)

Table 4.9: Statistics: Predicted/Effective mortalities for a discretisation based on the TIMI score.

Statistics: PCI Patients (STEMI)

TIMI	n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T
0	113	0	0.72 %	0.7 %	0.0%	$\approx$ 0.0 %	0.0 %
1	423	2	0.82 %	1.3 %	0.47%	16.38 %	0.68 %
2	561	6	1.2 %	1.9 %	1.07%	38.17 %	2.83 %
3	447	13	2.27 %	3.9 %	2.91%	18.24 %	10.66 %
4	416	11	3.79 %	6.5 %	2.64%	7.37 %	$\approx$ 0.0 %
5	360	10	4.34 %	11.6 %	2.78%	1.71 %	$\approx$ 0.0 %
6	222	10	8.97 %	14.7 %	4.5%	0.04 %	$\approx$ 0.0 %
7	148	11	13.84 %	21.5 %	7.43%	0.3 %	$\approx$ 0.0 %
8	92	10	20.26 %	24.4 %	10.87%	0.17 %	$\approx$ 0.0 %
>8	100	28	36.04 %	31.7 %	28.0%	4.59 %	20.71 %

HOSMER-LEMESHOW-statistics:

$H_{TIMI} = 89.322$ (P-value ≈ 0.0 %)

$H_{AMIS} = 24.3231$ (P-value 0.2 %)

Table 4.10: Statistics: Predicted/Effective mortalities for a discretisation based on the TIMI score.

Error Plots

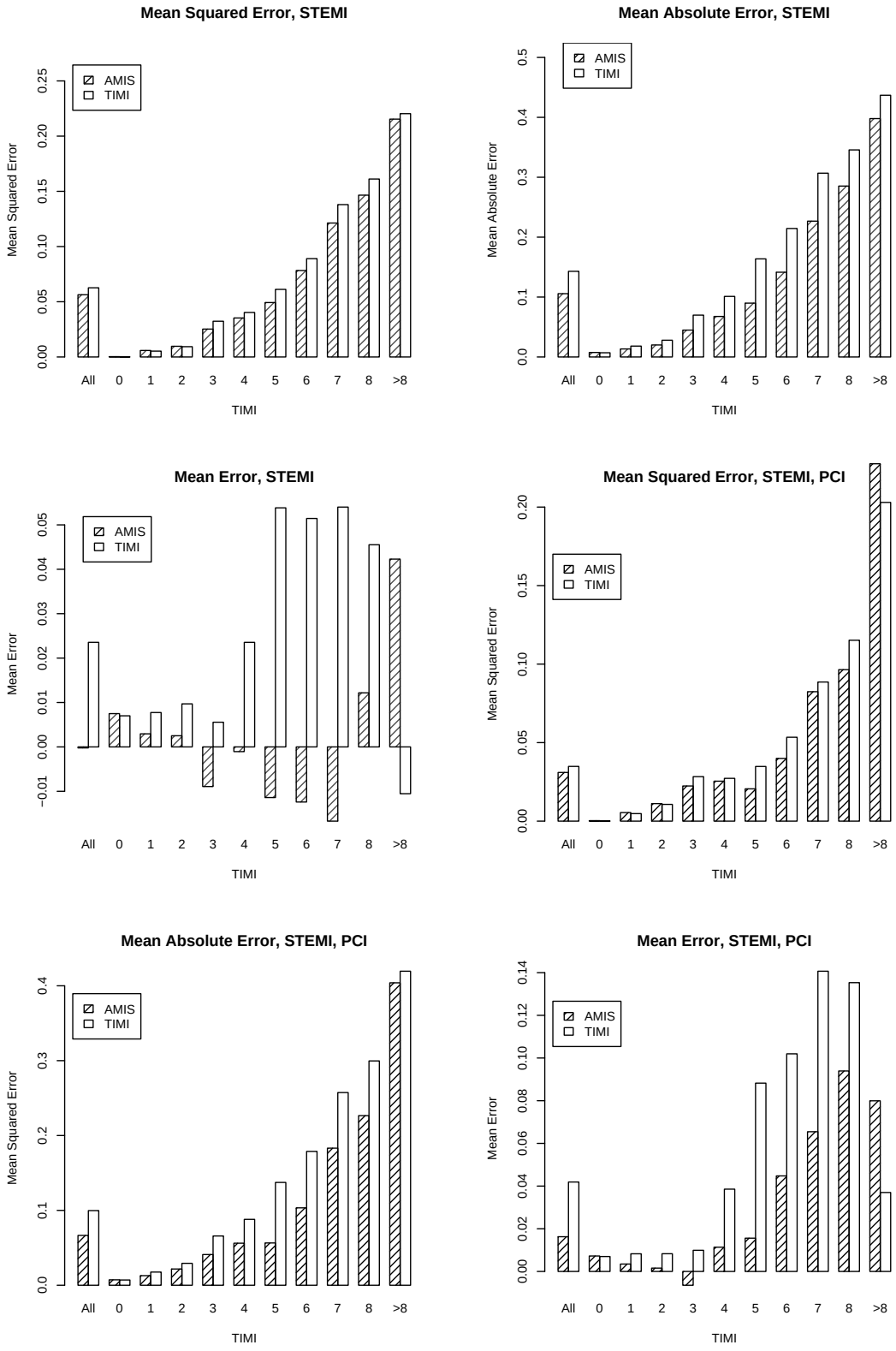


Figure 4.f: Error Plots for a discretisation based on the TIMI score.

Error Statistics

TIMI	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0564	0.0626	-0.0062	-4.6049	≈ 0.0 %	AMIS
0	165	0	$3.0 \cdot 10^{-4}$	≈ 0.0	$2.0 \cdot 10^{-4}$	1.3552	8.77 %	?
1	570	3	0.0059	0.0053	$6.0 \cdot 10^{-4}$	1.5953	5.54 %	?
2	752	7	0.0097	0.0093	$4.0 \cdot 10^{-4}$	0.4933	31.09 %	?
3	628	21	0.0252	0.0324	-0.0071	-2.5362	0.56 %	AMIS
4	603	25	0.0354	0.0403	-0.0049	-1.7146	4.32 %	?
5	563	35	0.0489	0.0612	-0.0123	-3.7139	0.01 %	AMIS
6	450	43	0.0787	0.0891	-0.0104	-2.2913	1.1 %	AMIS
7	323	52	0.1207	0.138	-0.0172	-2.8001	0.26 %	AMIS
8	262	52	0.1474	0.1612	-0.0138	-1.5875	5.62 %	?
9	345	113	0.216	0.2204	-0.0044	-0.3734	35.44 %	?

Table 4.11: Mean Squared Error Statistics for discretisation based on the TIMI score.

TIMI	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1057	0.143	-0.0372	-24.6225	≈ 0.0 %	AMIS
0	165	0	0.0076	0.0070	$6.0 \cdot 10^{-4}$	0.4757	31.72 %	?
1	570	3	0.0134	0.0181	-0.0047	-4.5178	≈ 0.0 %	AMIS
2	752	7	0.02	0.028	-0.0080	-6.2032	≈ 0.0 %	AMIS
3	628	21	0.045	0.0698	-0.0248	-9.8664	≈ 0.0 %	AMIS
4	603	25	0.0676	0.1011	-0.0335	-10.5	≈ 0.0 %	AMIS
5	563	35	0.0894	0.1637	-0.0743	-21.8639	≈ 0.0 %	AMIS
6	450	43	0.1419	0.2145	-0.0725	-14.139	≈ 0.0 %	AMIS
7	323	52	0.2261	0.3068	-0.0807	-11.3716	≈ 0.0 %	AMIS
8	262	52	0.2863	0.3456	-0.0593	-6.0973	≈ 0.0 %	AMIS
9	345	113	0.399	0.4369	-0.0379	-3.0936	0.1 %	AMIS

Table 4.11: Mean Absolute Error Statistics for discretisation based on the TIMI score.

TIMI	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	$2.0 \cdot 10^{-4}$	0.0235	-0.0234	14.7495	0.0 %	AMIS
0	165	0	0.0074	0.0070	$4.0 \cdot 10^{-4}$	-0.3405	36.67 %	?
1	570	3	0.0030	0.0077	-0.0048	4.4986	≈ 0.0 %	AMIS
2	752	7	0.0025	0.0097	-0.0071	5.4154	≈ 0.0 %	AMIS
3	628	21	-0.0084	0.0056	-0.014	5.0983	≈ 0.0 %	AMIS
4	603	25	-0.0012	0.0235	-0.0248	7.4077	≈ 0.0 %	AMIS
5	563	35	-0.0115	0.0538	-0.0653	17.8026	0.0 %	AMIS
6	450	43	-0.0117	0.0514	-0.0632	11.472	0.0 %	AMIS
7	323	52	-0.0171	0.054	-0.0711	9.5369	0.0 %	AMIS
8	262	52	0.014	0.0455	-0.0315	3.0932	0.1 %	AMIS
9	345	113	0.0453	-0.0105	0.0558	-4.5767	≈ 0.0 %	TIMI

Table 4.12: Mean Error Statistics for discretisation based on the TIMI score.

TIMI	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.031	0.0348	-0.0038	-2.7479	0.3 %	AMIS
0	113	0	$3.0 \cdot 10^{-4}$	≈ 0.0	$3.0 \cdot 10^{-4}$	1.0081	15.67 %	?
1	423	2	0.0055	0.0048	$7.0 \cdot 10^{-4}$	1.3908	8.22 %	?
2	561	6	0.0109	0.0106	$3.0 \cdot 10^{-4}$	0.2284	40.97 %	?
3	447	13	0.0219	0.0283	-0.0064	-2.1052	1.77 %	AMIS
4	416	11	0.0257	0.0272	-0.0015	-0.6251	26.6 %	?
5	360	10	0.0202	0.0348	-0.0146	-4.1372	≈ 0.0 %	AMIS
6	222	10	0.0405	0.0534	-0.0129	-1.8329	3.35 %	?
7	148	11	0.083	0.0886	-0.0056	-0.7122	23.82 %	?
8	92	10	0.0976	0.1152	-0.0176	-1.1721	12.06 %	?
9	100	28	0.2298	0.203	0.0268	1.2014	11.48 %	?

Table 4.13: Mean Squared Error Statistics for a discretisation based on the TIMI score (PCI).

TIMI	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0665	0.0997	-0.0332	-19.7422	≈ 0.0 %	AMIS
0	113	0	0.0075	0.0070	$5.0 \cdot 10^{-4}$	0.2968	38.33 %	?
1	423	2	0.013	0.0176	-0.0046	-3.3788	0.04 %	AMIS
2	561	6	0.0215	0.0293	-0.0078	-4.7747	≈ 0.0 %	AMIS
3	447	13	0.0407	0.0658	-0.0251	-9.223	≈ 0.0 %	AMIS
4	416	11	0.0562	0.088	-0.0318	-9.1167	≈ 0.0 %	AMIS
5	360	10	0.0562	0.1373	-0.0812	-23.1697	≈ 0.0 %	AMIS
6	222	10	0.1039	0.1788	-0.0749	-8.7876	≈ 0.0 %	AMIS
7	148	11	0.1819	0.2574	-0.0755	-6.6094	≈ 0.0 %	AMIS
8	92	10	0.2266	0.2997	-0.073	-4.114	≈ 0.0 %	AMIS
9	100	28	0.4065	0.4195	-0.013	-0.5486	29.17 %	?

Table 4.14: Mean Absolute Error Statistics for a discretisation based on the TIMI score (PCI).

TIMI	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0161	0.0419	-0.0258	14.9215	0.0 %	AMIS
0	113	0	0.0075	0.0070	$5.0 \cdot 10^{-4}$	-0.2968	38.33 %	?
1	423	2	0.0036	0.0083	-0.0047	3.3979	0.03 %	AMIS
2	561	6	0.0016	0.0083	-0.0067	4.1186	≈ 0.0 %	AMIS
3	447	13	-0.0061	0.0099	-0.016	5.5729	≈ 0.0 %	AMIS
4	416	11	0.0107	0.0386	-0.0278	7.7928	≈ 0.0 %	AMIS
5	360	10	0.0158	0.0882	-0.0724	18.0812	0.0 %	AMIS
6	222	10	0.0447	0.102	-0.0572	6.2669	≈ 0.0 %	AMIS
7	148	11	0.0618	0.1407	-0.0789	7.0077	≈ 0.0 %	AMIS
8	92	10	0.0934	0.1353	-0.0419	2.2267	1.3 %	AMIS
9	100	28	0.0811	0.037	0.0441	-1.8891	2.95 %	?

Table 4.15: Mean Error Statistics for a discretisation based on the TIMI score (PCI).

4.4 Summary

The aim of this section was to have an understanding about how the AMIS score performs, especially in comparison to the TIMI prediction. To that end, the 'raw'[13] AMIS score was used. Knowing its strenghts and weaknesses, the remainder of this work will concentrate on techniques which could help to ameliorate the AMIS score.

4.4.1 Strengths and Weaknesses

In general, we have found that the AMIS predictor is doing better at predicting probabilities in terms of subgroup mortality. This has been showed using plots and correlation analysis, as well as high values for the HOSMER-LEMESHOW statistics.

This represents a view where mean errors are considered. If absolute error is what counts, AMIS has considerable error rates as well. Nevertheless, it could be shown that even in on an absolute error level, AMIS significantly outperforms TIMI. This, in turn, is supported by the pairwise t-statistics calculated for the mean absolute error rates.

AMIS seems to clearly outperform TIMI. There are a few exceptions where the difference is less evident:

- In the very high risk region of the score, in the AMIS-based discretisations normally the last bin, TIMI has sometimes even significantly lower error rates for PCI patients.
- The difference between squared errors is generally not as clear as between absolute errors.

[13] I.e. the AMIS score as delivered by the standard AMIS model was considered; there were no modifications or adjustments made to it whatsoever.

5 Improvement

In order to improve and evaluate the utility of the score as a probability estimate, we will divide the score into discrete intervals, each of them representing specific expected values. The classification system obtained will be evaluated for different treatment subgroups.

5.1 Introduction

Starting point of this section are the AMIS score's weaknesses discovered in the course of evaluation (cf. Section 4.4). Methods are presented which could lead to an improvement of the score.

5.1.1 Raw Scores and Improved Scores

In this context, the original AMIS score s will occasianally be referred to as the 'raw' score. All the methods presented in the following take the latter as a basis, do something with it, and deliver an improved score, which hopefully performs better in some aspects. Those encompass the accuracy of the probability predictions as well as error rates.

5.2 Outlier Detection

The error analysis has shown substantial differences between the *mean* squared errors and the *absolute* squared errors, which is normally due to outliers, as those tend to preponderate when squared. It seems therefore promising to find common characteristics among them in order to discern them. Then, either their score can be adjusted, or they are excluded.

5.2.1 Analysis

To examine the distribution of the score including the outliers, the following figures show the actual outcome as a bold line on $y = 0$ (dying patients) and $y = 1$ (survivors), as well as the predicted scores sorted by outcome. This neatly shows that there is quite some variance among the scores. The farther the points are away from the bold lines representing the outcomes, the greater is the deviation:

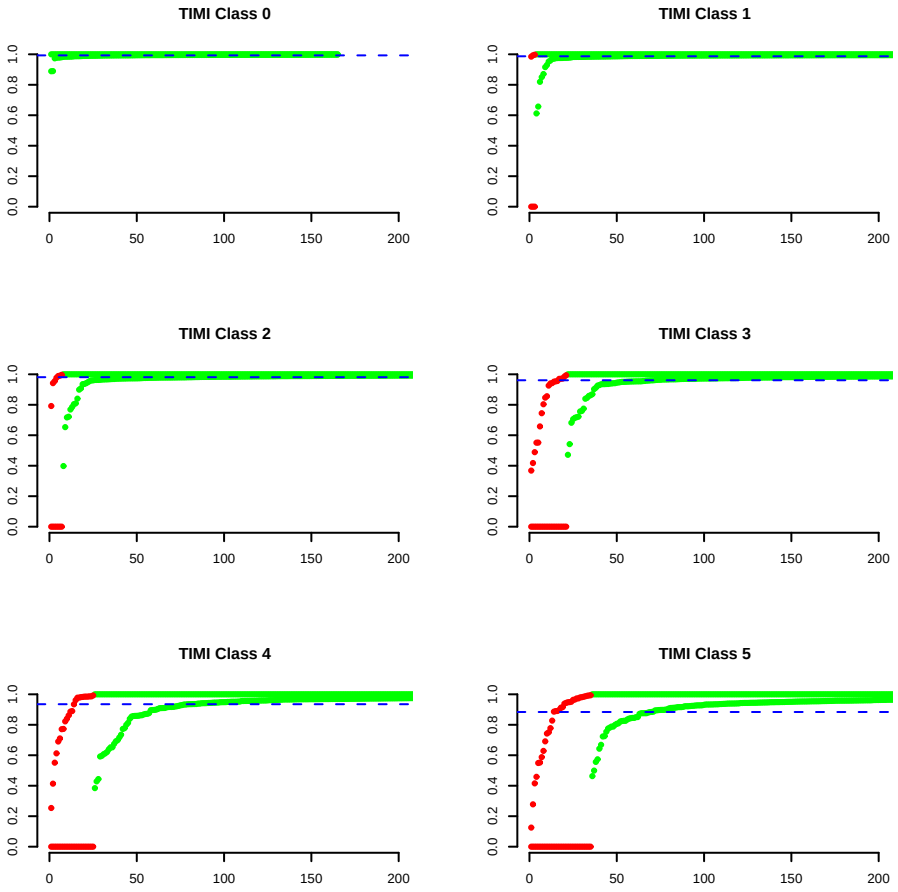


Figure 5.a: Outlier Visualisation in TIMI classes 0-5. The blue (dashed) line represents the TIMI prediction.

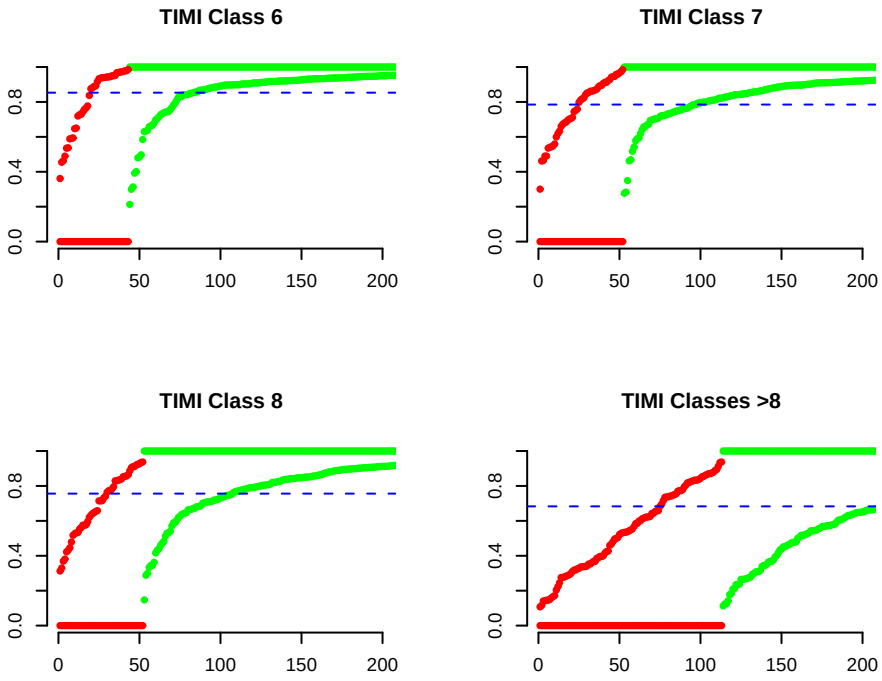


Fig 5.b: Outlier Visualisation in TIMI classes 6-9. The blue (dashed) line represents the TIMI prediction.

5.2.2 Possible Solution

Some attempts have been made to learn which are possible outliers. For this purpose, the top outliers (the 5% highest scores among the false positives and the 5% lowest scores of false negative) have been isolated and given a value 'outlier' in a newly created attribute 'isoutlier'. The results were not convincing. From a theoretical perspective though, it is questionnable at all if this approach is appropriate: if outliers could be learned, why are they not incoroporated by the model in the first place.

5.3 Recalibration

A possible source of the observed errors could be a bad calibration of the score. The calibration issue will be introduced and analysed with respect to the AMIS score.

5.3.1 Introduction

We do not generally know if the probability estimates obtained by the algorithms are well-calibrated, or systematically distorted: For instance, Simple Naive Bayes classifiers have been shown to consistently over- and under-estimate probabilities (e.g. in [Zadrozny/Elkan, 2001]).

Therefore, even if they are good rankers (which is reflected in the good results obtained from ROC curves), they can be poor probability estimators.

As our model is based on a bayesian model, it will be analysed wheter bad calibration is a problem in AMIS or wheter it is not.

The idea is then to adjust the score in a way such that it is more accurate as a probability predictor, without compromising its good ranking capability.

5.3.2 Calibration of the AMIS score

We already know that differences between effective and predicted mortality exist. The remaining question is therefore wheter those are systematic. To find out, the results obtained in the evaluation can be analysed:

When the different discretisations shown in Appendix D.2 are compared visually[14], some degree of regularity seems to exist in the over-/under-estimation of the effective mortality: in the left bins (until somewhere where the bin assumes an $\bar{s}$ of roughly below 2%), the score seems to over-estimate mortality. Then, for a couple of bins, it seems to under-estimate. Only for the last bin does it again over-estimate the mortality.

5.3.3 Related Work and Available Strategies for Recalibration

[Provost/Domingos, 2003] has analysed the probability estimates delivered by probability estimation trees (PETs) and shown that they do not result in accurate probability estimates. Avoiding pruning and adjusting the estimates by the LAPLACE correction are proposed for improving the estimates.[15]

[Domingos/Pazzani, 1996] analysed simple Naïve Bayes estimates and provide theoretical optimality conditions for the estimates.

[Zadrozny/Elkan, 2001] proposes a discretisation strategy to obtain more accurate probability estimates by binning the scores[16].

It could also be considered to improve the score by some sort of *ex post* transformation. The difference between effective and predicted probabilities could be approximated using appropriate methods (e.g. polynomial regression, B-Splines or similar) and the score of the AMIS model adjusted by that difference.

[14] The reader is encouraged to observe the behaviour of the 10, 9, ..., 4-bins-discretisation curves like a flip-book!

[15] The LAPLACE correction adjust the calculation of conditional probabilities within the algorithm ($p = \frac{k+1}{n+2}$ instead of $p = \frac{k}{n}$), such that probabilities are smoothed toward 50%, which works well for balanced data sets ([Provost/Domingos, 2003]). This has been generalised in the 'm-estimation' method for use in non-balanced data sets, where p is set to $\frac{k+b \cdot m}{n+m}$, where b is the average probability of the positive class and m a paramter which controls how many scores are smoothed and which is normally set heuristically ([Cestnik, 1990]).

[16] A 'histogram method' is proposed which sorts predictions by score and then splits them into bins of equal size. The corrected probability of an instance belonging to a specific bin is then given by the actual relative frequency of true positives within that bin.

5.3.4 Discussion

Again, as discussed earlier, we are relying on discretisations in order to analyse the prediction behaviour of the AMIS score. It is therefore almost infeasible to extrapolate the deviations leading to the observed under- and over-estimation of effective mortality, since such an extrapolation, of whatever form, will look differently depending on the discretisation it is based on. An *ex ante* transformation is therefore a bad choice.

Given our use of an algorithm 'out of the box', intrinsic adjustments of the predictor (LAPLACE corrections and the like) (e.g. by using corrections on conditional probabilities) are out of scope.

A discretisation strategy as described in [Zadrozny/Elkan, 2001] seems promising. Note the difference to the discretisation issues discussed during the evaluation: Here, discretisation is used to get a new *model* based on the scores, not directly relying on the 'raw' scores delivered by the original model (AODE in our case).

5.3.5 Discretisation

Mechanism The discretised adjustments of the bins was carried out as follows:

1. In the ν-fold cross-validation, each test fold's (say F_1, ..., F_ν) results were sorted by their AMIS score s_i, in descending order, i.e. the highest score first.
2. The folds were then binned in to n bins (let $B_i^{(F_j)}$ denote the i-th bin of test fold F_j), with a pre-defined amount of instances within one bin ($|B_i^{(F_j)}| = n_i(B_i)$, i.e. in all folds, the same bin structure is applied). The greater n, the more fine-graned are the adjusted predictions.
3. Within each test fold F_j, and for each bin, an adjusted score $s_{B_i}^{F_j}$ was determined, which corresponds to the average mortality observed within that bin i.e. $s_{B_i}^{F_j} = \frac{1}{n_i} \sum_{Patients\, in\, Bin\, B_i\, of\, F_j} y_i$. This quantity was taken as the predicted probability.
4. All evaluations were performed on the joint test sets, i.e. considering scores $[s_{B_i}^{F_j}]_{j=1..\nu, i=1..n}$, just as explained earlier for the standard tenfold cross-validation.

The following matrix looks more complicated than it actually is and helps to understand:

5.3.6 The n and ν Trade-off

As n, the number of bins for the adjusted scores, is concerned, they are more fine-graned when n is big. Obviously, there is a trade-off between high n's for fine-graned adjusted scores on the one hand, and low enough n's to ensure that a sufficient number of instances in individual bins are available, on the other hand: too small bins lead to bouncing average mortality rates, as

there are too few dead patients available in most of those bins[17]. As a consequence, the adjusted prediction would, not surprisingly, lead to bad results.

Of course, another quantity determining the size of the resulting bins is the number of folds in the cross-fold validation: If the number is kept low, more instances are available for a bin. However, choosing ν too low decreases the quality of the evaluation.

As an example, if the AMIS training cohort ($|D| = 7648$) is used in conjunction with 10 equally-sized bins (i.e. 10 differnt levels for the adjusted score), the usual ten-fold cross-validation ($\nu = 10$) would lead to bin sizes of $|B_i^{(F_j)}| = 76$, if each bin has the same number of instances.

5.3.7 Threefold cross-validation Results

As a proof of concept, some discretisations were evaluated in a 3-fold cross-validation. $\nu = 3$ was chosen to ensure bins are big enough even when adjusted scores are sufficiently fine-graned. Therefore, the results are more of a 'proof of concept' than a genuine evaluation. However, the method can easily be deployed under different test scenarios. For example, one might use the whole training cohort to determine the adjusted scores and evaluate them on a independent test set when more data will be available in the future. Doing so, an even more fine-graned adjusted score could be obtained.

Discretisation choice To reflect the skewd distribution of the score, the bin sizes have been chosen decreasing rather than equally-sized. The following table shows the results obtained by AUC:

Bins	Instances in Bins ($\Sigma = 2048 = \lvert D\rvert/\nu$)	AUC
7	$500 + 500 + 250 + 250 + 250 + 250 + 48 = 2048$	0.808
6	$1000 + 500 + 500 + 250 + 250 + 48 = 2048$	0.864
8	$1000 + 250 + 250 + 250 + 250 + 250 + 250 + 48 = 2048$	0.871
9	$500 + 250 + 250 + 250 + 250 + 250 + 250 + 250 + 48 = 2048$	0.874

Table 5.1: AUCs of adjusted scores-based predictions with different discretisations

Percentile-based plots When the percentile-based evaluation method presented in Section 3.6.2 is applied on the last discretisation in Table 5.1, a considerable improvement in accuracy of predicted probabilities can be observed. Apart from the visual impression found when comparing the raw and adjusted curves, most of the P-values for the adjusted score are even higher than the raw ones.

The following figure illustrates the effects of recalibration on the D_T STEMI level, i.e. involving all therapies, but only STEMI patients:

[17] This is exactly the same effect as explained in [**]

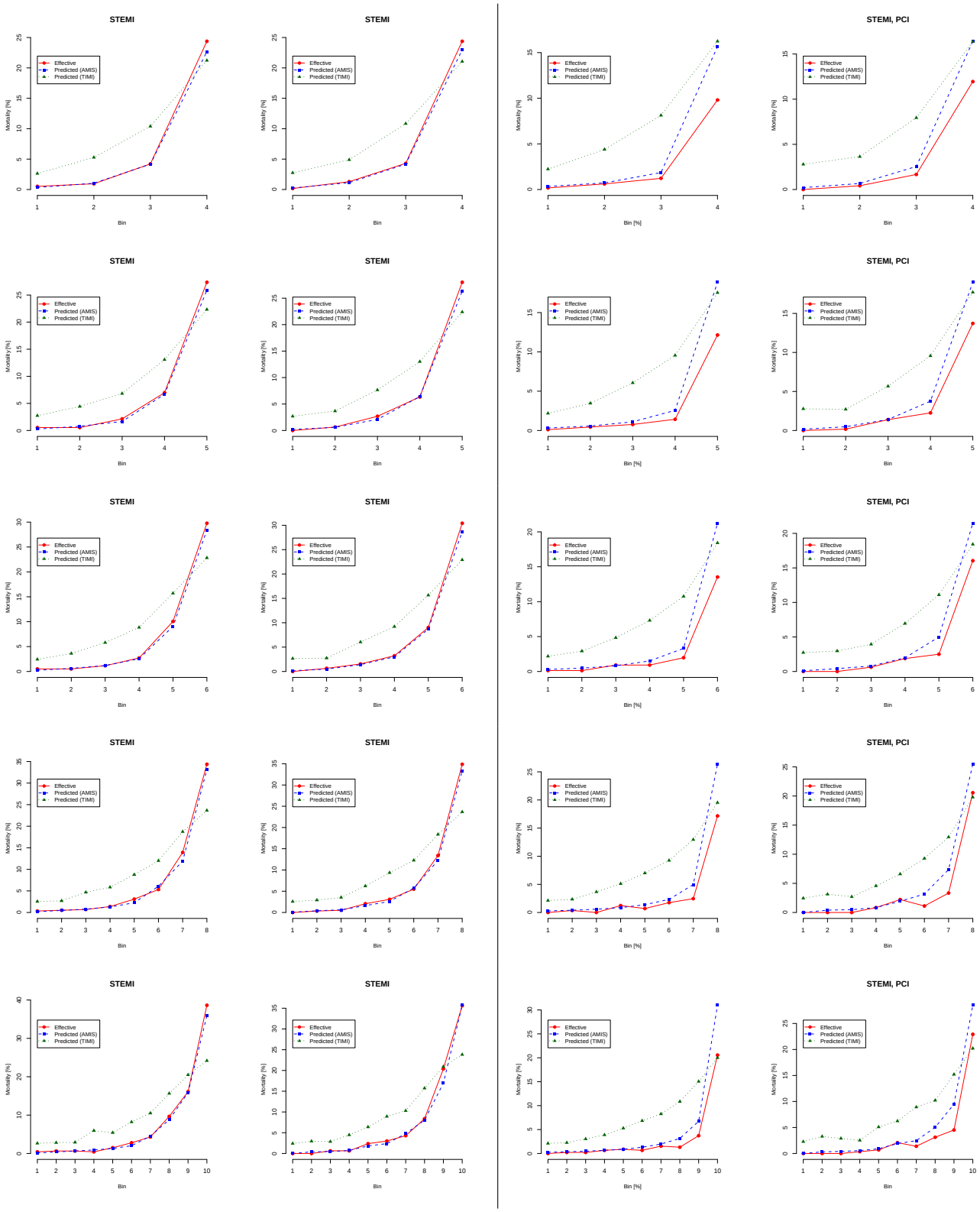


Figure 5.c: The optical comparison of the raw vs the adjustes score. The legends can be ignored at this point – the diagrams can be found in bigger size in the appendix.

As an example, the 8 bins discretisation is shown in the figure below:

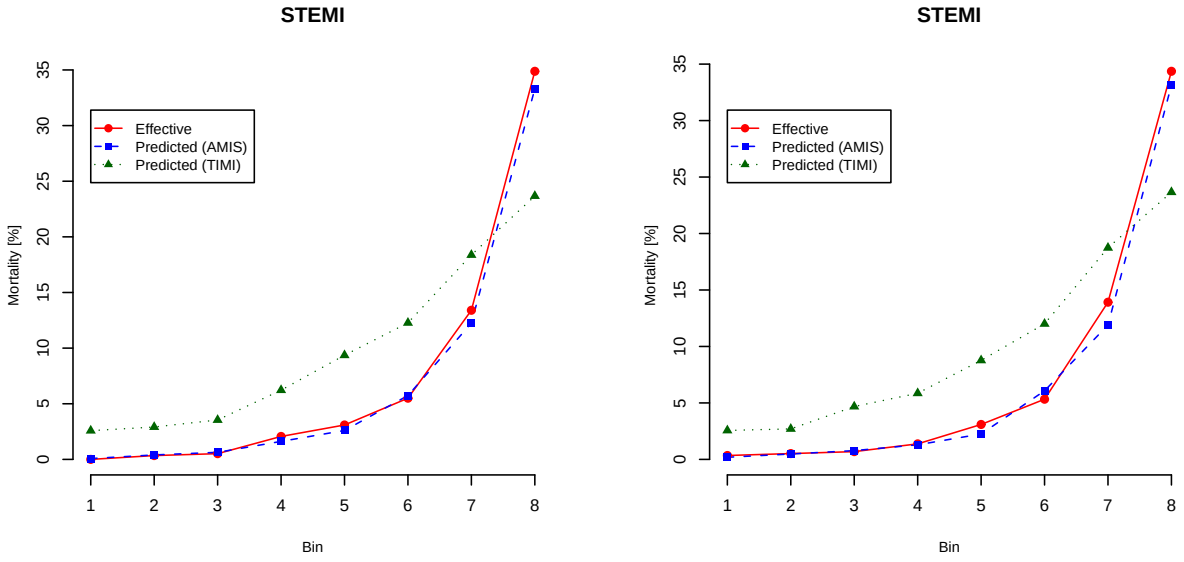


Figure 5.d: Adjusted score on the left side; Raw score on the right side.

The HOSMER-LEMESHOW test's P-values could be drastically improved on the STEMI training cohort:

n	STEMI		STEMI∩PCI	
	H_s	$H_{\tilde{s}}$	H_s	$H_{\tilde{s}}$
4	2.41%	39.7%	0.01%	0.06%
5	5.7%	18.83%	0.03%	0.11%
6	1.81%	54.13%	0.03%	0.16%
7	1.98%	56.42%	0.13%	0.19%
8	1.97%	76.54%	0.08%	0.15%
9	0.58%	52.92%	0.13%	≈0%
10	0.32%	≈0%	0.5%	≈0%

Table 5.2: Different P-Values of the HOSMER-LEMESHOW test applied to raw and adjusted AMIS score

5.3.8 Interpretation of the Results

The results are encouraging for the STEMI trainig cohort, where the fit of predicted means could be visibly improved. They were not so encouraging for the PCI cohort and even less so, when squared and mean errors were compared.

The mean error and the fit could be quickly improved with an ad-hoc choice of parameters. In future, and with more training data (or independent test data) available, the recalibration bins could be selected systematically. The evaluation parameter ν could be increased if more data were available.

5.4 Model Refinement

The analysis of error rates brought to light that the AMIS score is doing rather badly in the very high risk regions.

5.5 Including the Treatment Option

It has been presumed that the inclusion of the treatment attribute[18] could ameliorate the prediction capability of the model. Indeed, it does so in terms of ranking, the ROC analysis suggests:

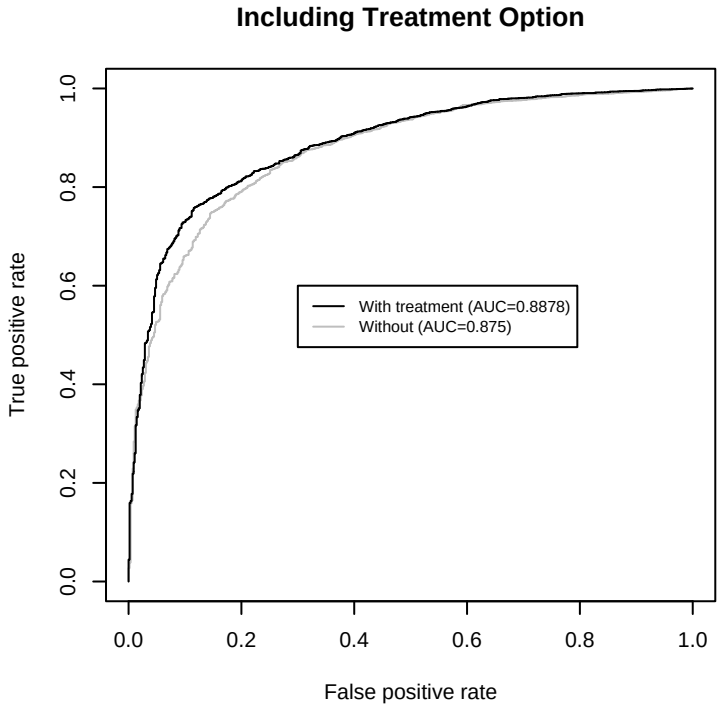


Figure 5.e: ROC curves of the model without (standard) and with the treatment attribute.

[18] which consists of the 4 possible treatment options: 0=Thrombolysis, 1=PCI, 2=Both, 3=None

6 AMIS[L] Model of Long-term Survivability

The AMIS Model is based on the in-hospital or short-term mortality rate.[19] The aim of this part of the thesis is to develop a similar model to predict long-term mortality.

To this end, an extended data set containing additionnal attributes for some of the AMIS patients, collected at the Triemli hospital in Zurich, was at disposal. As the Triemli data set contains information not already covered by the AMIS data set, it has been decided to go through the whole data mining process anew, independent from the existing AMIS model.

6.1 Data Understanding

6.1.1 General considerations

The Triemli dataset contains information about patients delivered to the Triemli Hospital. Reducing the set to patients admissed after 1997 (=: T) , at which point AMIS begins is (theoretically) a subset of the AMIS dataset (D), i.e. for each record $P_i \in T$ there exists a $P_j \in D$ and P_i and P_j represent the same patient.

900 patients in T contain follow-up information collected in phone calls to them made within different periods after their initial admission. That information includes the date of decease as well as the cause of death, if the patient has died previous to the call, and details about the general condition of the patient if he/she is still alive.

6.2 Data Preparation

6.2.1 Matching

Not all attributes needed as inputs for the AMIS model are present in the T. The 900 instances had first to be assigned to the according AMIS instances.

Method For each instance of the Triemli set, an instance in the AMIS set[20] was iteratively searched. As a combined natural key, the date of birth (`DAT_GEBURT` in T, `birthdat` in D) and the date of hospitalisation was used.

All instances as well as in T as in D provide values for the date of birth.

Not so for the hospitalisation date: Whereas in D, the attributes `admis_day`, `admis_month` and `admis_year` could be used, the corresponding attribute in T (`DAT_HOSP`) had 103 missing values. On the other hand, dates found in `DAT_ANKUNFT_TRIEMLI`, `DOOR_IN_TRIEMLI` and `DOOR_IN_HKL` are consistently identical to `DAT_HOSP` for instances where both are provided (with the exemption

[19] According to expert knowledge, the short-term (e.g. 30 days) mortality does not significantly differ from the in-hospital mortality (cf. assumption in [Hunt, 2006], Section 3.2.1).

[20] The basis was the cleaned, prepared AMIS set ($n = 16605$), i.e. the full cohort.

of 2 or 3 instances, where they differed only by 1 day). The hospitalisation date was therefore taken from `DAT_HOSP` or from any of the three others, whereever the information was available.

Result Unfortunately, there were only a couple ($n = 23$) of instances which matched perferctly. It was however known that the Triemli admission date could slightly differ from the AMIS admission date, though only a few days. But even when allowing such a difference between the AMIS and the TRIEMLI hospitalisation dates, only few perfect matches were possible. The following table shows the outcomes for different tolerances for matching[21]:

	Perfect match	$\Delta h \leq 5$	$\Delta h \leq 10$	$\Delta h \leq 30$
Not found at all	110	110	110	110
Date of birth matched	768	644	531	176
Perfectly Matched	22	146	259	614
Sum	900	900	900	900

Table 6.1: Matches found in the AMIS set, using cleaned and prepared data.

Theoretically (assuming integrity and correctness of both data sets), this could only be explained if those not matching are incidentally the same patients which have been removed from the data set during the cleaning stage. To test this hypothesis, the matching procedure was applied to the raw data:

	Perfect match	$\Delta h \leq 5$	$\Delta h \leq 10$	$\Delta h \leq 30$
Not found at all	104	104	104	104
Date of birth matched	234	187	184	180
Perfectly Matched	562	609	612	616
Sum	900	900	900	900

Table 6.2: Matches found in the AMIS set, using the raw data set.

Practically though, it is not within our means to resolve this problem. A tool which has been developed to analyse and match the data can be found on the CD accompanying this thesis. Details are described in Appendix C.1.

6.3 Conclusion

This inconsistency makes the integration of the two data sets, impossible. As the latter would be a prerequisite for any further steps in the data mining process, the idea of evaluating the long-term model has been abandoned altogether.

[21] 'Perfect match' indicates that both date of birth and hospitalisation date must be identical. Using $\Delta h \leq C$, date of birth must be identical, whereas the hospitalisation dates can have differences of up to C days.

7 Conclusion

7.1 What has been done

We have seen that the evaluation of the probabilities delivered by the AMIS score requires discretisation. For discretising, we needed to take into account various aspects. Considering these, we chose TIMI-based and AMIS-percentiles-based discretisations to evaluate the AMIS predictions and to compare them to TIMI predictions.

The analysis has shown that, with the exception of high-risk PCI patients, AMIS significantly outperforms TIMI in terms of risk estimation.

Furthermore, methods for improving the AMIS score have been presented. The recalibration method achieved a better fit of the model.

Additionnaly, the TRIEMLI data set was analysed. It turned out that the data in the present form could not be used as basis for a new model for long-term mortality prediction.

7.2 What has been achieved

Most importantly, a sound evaluation of the AMIS model's prediction values showed that it outperforms TIMI. This is an important enhancement to the results found by [Hunt, 2006] which concentrated on ROC analysis of the two-class problem 'dead' or 'alive' rather than on predictive power for mortality rates.

The discussion of viable approaches to evaluation and, for this purpose, discretisation, contributed to the understanding of the AMIS model and its interpretation. Discretisation is not only crucial for evaluation purposes. For a future clinical use, the AMIS score will certainly have to be discretised (a couple of distinct risk classes might be specified). It has been well documented how this can be sensibly done, and how it should not be done.

Furthermore, a collection of JAVA programs has been developed which helps to evaluate and analyse the score.

7.3 Outlook

As time is always constrained, there remain many possible enhancements to this work.

- Techniques to improve the score should be revisited in a more systematical way. These include recalibration and model refinement as discussed in Section 6. A code base for implementing the refinement procedure exists.
- The design of a practical risk classification model based on a discretisation of the AMIS score should be discussed. Prototypes calculating the respective AMIS risk predictions can easily be built based on the WEKA API and using the enhancements provided by this thesis.

A Appendix A: Statistics

A.1 Statistical Methods

A.1.1 Paired t-Test

When $x_i$and y_i are measured values belonging to the same instance (e.g. patient), the difference $d_i := x_i - y_i$ is used to form the test statistic

$$\hat{t} = \frac{\bar{d}}{s_{\bar{d}}} = \frac{\frac{1}{n}\sum_{i=1}^{n} d_i}{\sqrt{\frac{\sum_{i=1}^{n} {d_i}^2 - (\sum_{i=1}^{n} d_i)^2/n}{n(n-1)}}} \tag{7}$$

The test setup is then as follows:

H0: $\mu_d = 0$

H1: $\mu_d \neq 0$

If $\hat{t} > t_{n-1;1-\alpha/2}$ or $\hat{t} < -t_{n-1;1-\alpha/2}$, H_0 can be rejected at a confidence level of α, i.e. when $\hat{t}$ is 'extreme' enough, we can say that the probability of the observed difference being due to chance alone (this is the P value) is very low (1).

Accordingly, low P values prevent us from rejecting H_0 and can therefore be interpreted as a hint of equality (2) (strictly speaking, this is not equivalent to rejecting inequality – however, a test allowing this could not be found).

In our domain, we use the test properly (by significantly rejecting H_0)when we show that error measures are different from each other (difference is good in terms of discriminating models). We refer to the weaker interpretation (2) as when predicted and effective mortality rates are compared (difference is bad in terms of accurately estimating probabilities).

A.1.2 Hosmer-Lemeshow statstic

The HOSMER-LEMESHOW statistic H is similar to the χ^2 statistic used for a χ^2 test for categorical, grouped data. H follows a χ^2_{n-2} distribution. If the fit of the model is good enough, the statistic is sufficiently high and H_0 (predicted probabilities = effective probabilities) cannot be rejected. This corresponds to a high P-value, since the P-value is the probability to be wrong about rejecting H_0, i.e. the probability to be wrong about rejecting the good fit.

A.1.3 WILLIAMS-KLOOT Statistic

[Prince, 1982] applies the WILLIAMS-KLOOT statistic, following a version provided by [Himmelblau, 1970]. The following hypothesis relating to the estimate $\hat{\lambda}$ for the regression coefficient between $X_i$and Z_i is tested

$H_0 : \lambda = 0$ (no significant positive or negative slope of the regression line)

$H_1 : \lambda \neq 0$ (either negative or positive slope of the regression line)

by the calculation of the confidence interval $CI = \hat{\lambda} \pm \hat{\sigma_\lambda} T_{1-\alpha/2;n-1}$, where $T_{x;N}$ denotes the quantile with $\Phi(T_{x;N}) = x$. If the confidence interval does not contain 0, $\hat{\lambda}$ significantly differs from 0 and thus H_1 is confirmed on the confidence level α.

To be consistent with the other tests performed here, the procedure is slightly modified by using rejection areas rather than confidence intervals. It will thus be tested whether the transformed number $\frac{\hat{\lambda}}{\hat{\sigma_\lambda}}$ lies outside the rejection points $T_{1-\alpha/2;n-1}$, i.e. H_0 will be rejected if

$$\frac{|\hat{\lambda}|}{\hat{\sigma_\lambda}} > T_{1-\alpha/2;n-1}$$

A.2 Normality Assumptions

A.2.1 Histograms of the Mean AMIS Score in a 6-Bins Discretisation

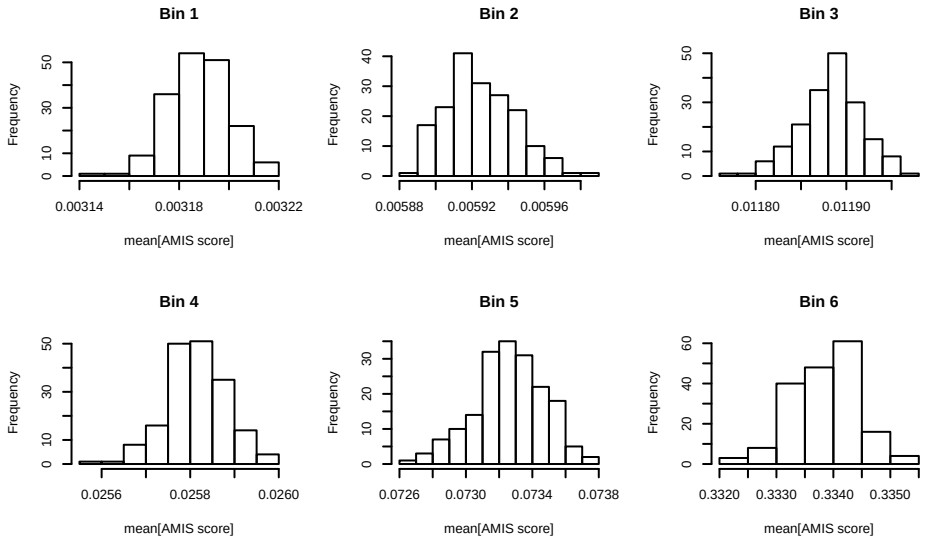


Figure 1: Abcd

Figure A.1: Histograms of mean scores obtained in 180 tenfold cross-validation runs

A.2.2 Histograms of the Mean Survival in a 6-Bins Discretisation

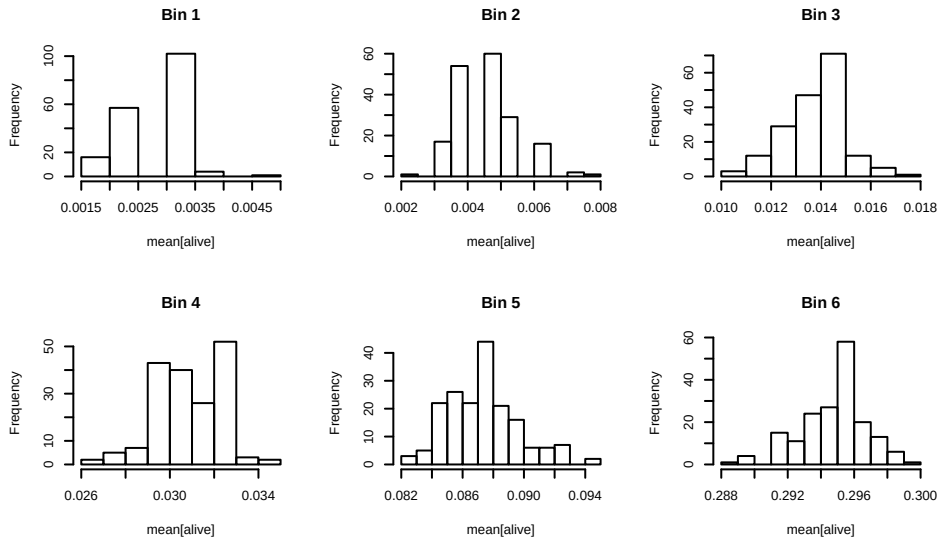


Figure A.2: Histograms of mean survivals a obtained in 180 tenfold cross-validation runs

A.2.3 Histograms of the Loss Functions in a 6-Bins Discretisation

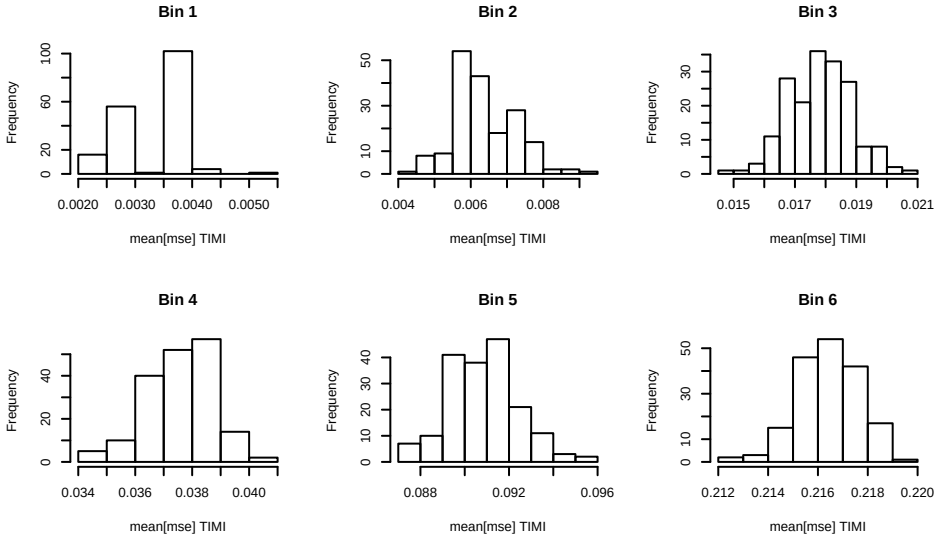


Figure A.3: Histograms of mean squared TIMI error obtained in 180 tenfold cross-validation runs

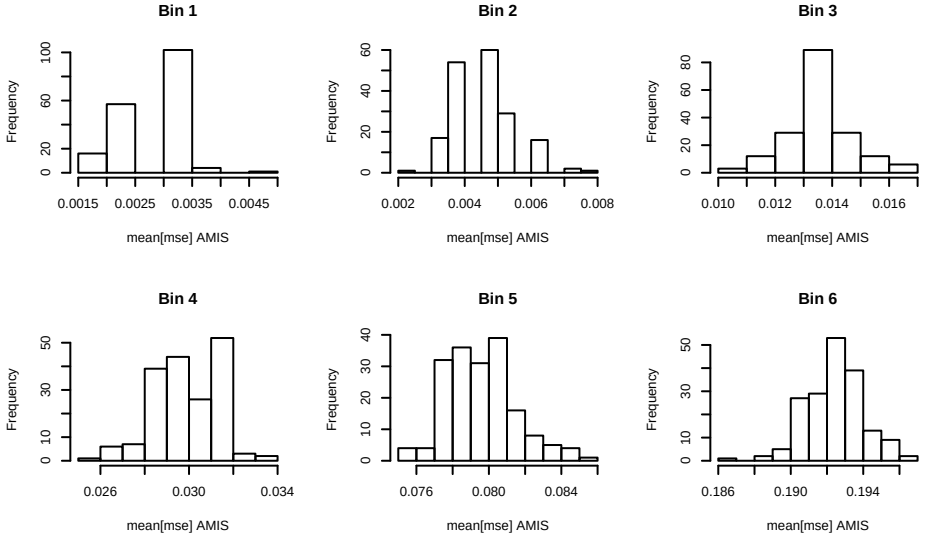


Figure A.4: Histograms of mean squared AMIS error obtained in 180 tenfold cross-validation runs

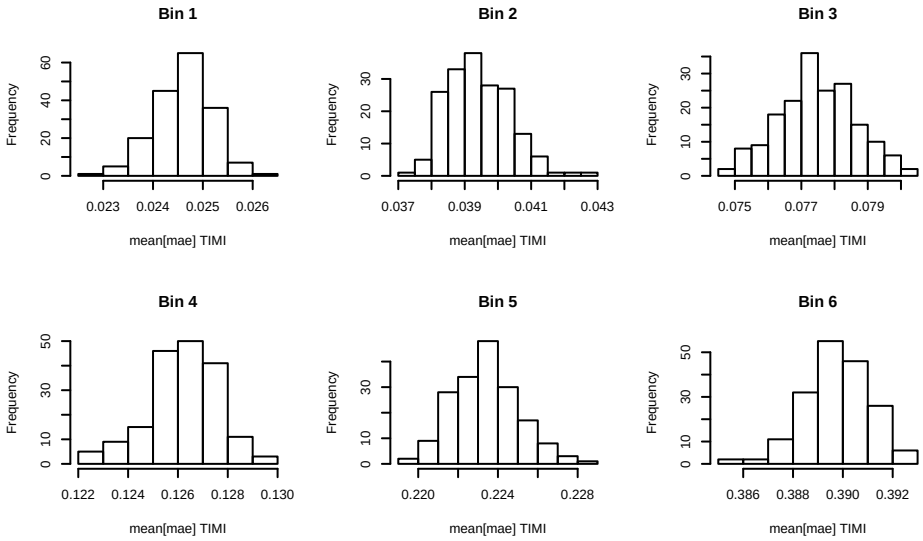


Figure A.5: Histograms of mean absolute TIMI error obtained in 180 tenfold cross-validation runs

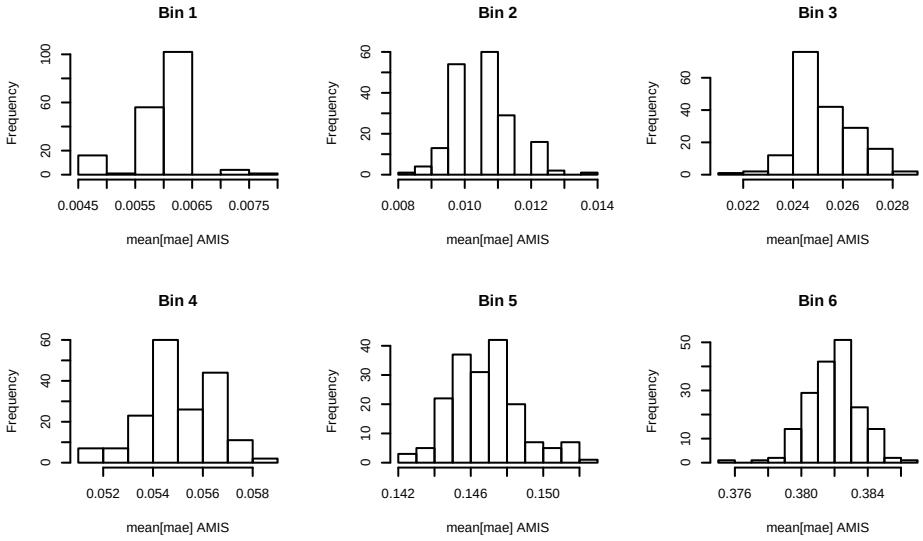


Figure A.6: Histograms of mean absolute AMIS error obtained in 180 tenfold cross-validation runs

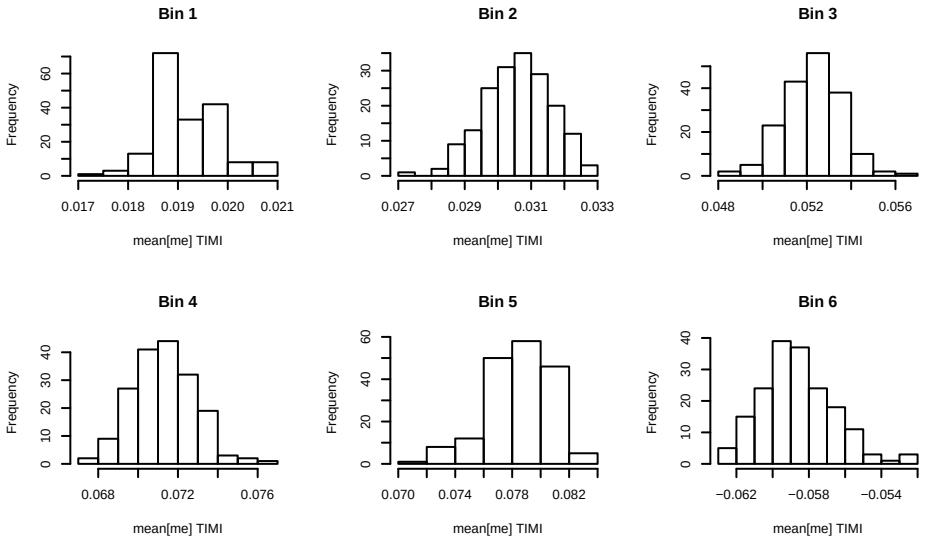


Figure A.7: Histograms of mean TIMI error obtained in 180 tenfold cross-validation runs

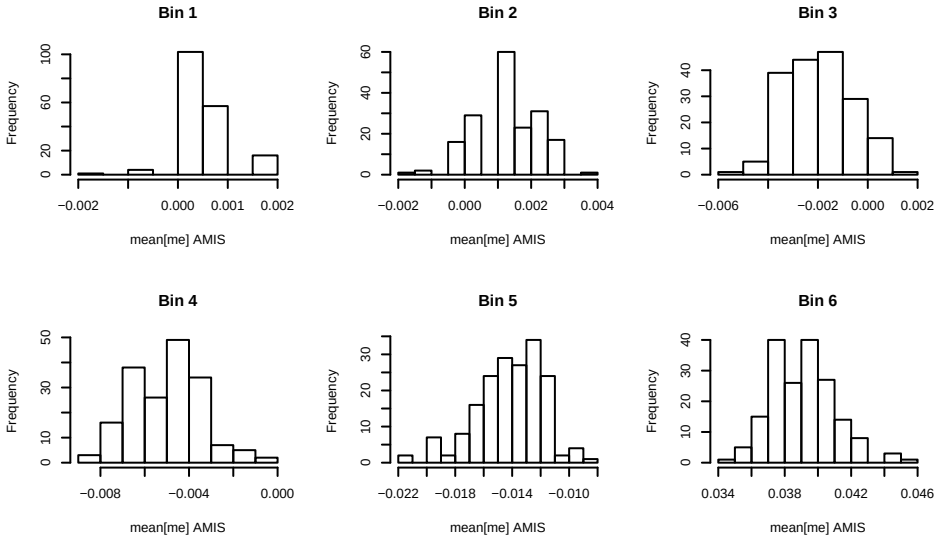


Figure A.8: Histograms of mean AMIS error obtained in 180 tenfold cross-validation runs

B Data

B.1 Data Cleaning and Preparation

In addition to the cleaning and preparation steps performed in [Hunt, 2006], the following adjustments have been made. Clementine streams are available on the CD.

Enhanced Preparation In addition to the steps explained above, enhanced cleaning and preparation of data was undertaken:

1. Roughly 50 records had `thrmblys` missing. Most of them were pre 1999. They were discarded.

C Software Tools

The following sections contain a more detailed description of the mining process and enable the reconstruction of the results provided in this thesis.

C.1 Triemli Matcher

The TriemliMatcher program iterates through the Triemli data set. For each Triemli record, the whole AMIS data set is sequentially searched through.

If date of birth and date of hospitalisation in the two sets match, all fields are written to `matched.csv`. The matching criteria of the hospitalisation date can be softened using the `APPROXIMATE_MATCH_DIFF` parameter: When set to 0, hospitalisation dates must perfectly match (equal day, month and year), when set to d, a difference of up to d days is tolerated for a 'perfect' or 'unique' match.

If the date of birth only matches, the record is written to `notunique.csv`, along with all the weak matches found (i.e. all AMIS records having the same date of birth). `notunique.csv` has thus the following structure:

[TRIEMLI i Attributes], [AMIS i_1 Attributes]
[TRIEMLI i Attributes], [AMIS i_2 Attributes]
...
[TRIEMLI i Attributes], [AMIS $i_{n(i)}$ Attributes]

where [TRIEMLI j Attributes] is the comma-separated list of patient number j's values from the Triemli data set, [AMIS k Attributes] the comma-separated list of patient number k's values from the AMIS data set. Triemli i data are thus replicated for each AMIS record (there are $n(i)$ of those, depending on the Triemli record).

`notfound.csv` contains the comma-separated values of all Triemli records for which no corresponding date of birth could be found within AMIS.

`notuniqueselection.csv` is a summary useful for cross-checking, containing the type of match (either `UNIQUE` (perfect), `BIRTHDAY_MATCHED` or `NOT_FOUND`) for each Triemli record, along with the relevant attributes for the match (id's from both sets, dates of birth, dates of hospitalisation, hospitalisation difference).

D Evaluation Results

D.1 Symbols used in Statistical Summary Tables

For convenience, symbols used in the statistical summary tables are collected below:

Symbol	Meaning
n	Number of instances in the bin
$n_\dagger$	Number of dead patients in the bin (alive= 0)
$\bar{s}$	Arithmetic Mean of the AMIS score in the bin
$\bar{\tau}$	Arithmetic Mean of the TIMI scores in the bin
$\bar{m}$	Arithmetic Mean of the Effective Mortality in the bin
P_A	P-Value for H_0: $\bar{s} = \bar{m}$ in the paired t-test of the predicted/effective means (low value indicates that differences are not to chance alone)
P_T	P-Value for H_0: $\bar{\tau} = \bar{m}$ in the paired t-test of the predicted/effective means (low value indicates that differences are not to chance alone)
e^{AMIS}	Error measured for AMIS (type of error depending on chart)
e^{TIMI}	Error measured for TIMI (type of error depending on chart)
Δe	$= e^{AMIS} - e^{TIMI}$
t_p	t-statistic for $d := e^{AMIS} - e^{TIMI}$ (if high enough, $H_0 : e^{AMIS} = e^{TIMI}$ can be rejected)
P	P-value for t_p (corresponds to the probability of falsly assuming difference in errors)
H_{Model}	HOSMER-LEMESHOW-statistic of the model. The related P-value can be interpreted as the goodness of fit of the model (the higher the better).

Table: Meaning of symbols used in the statistical summary tables

D.2 Evaluation of Percentile-based Discretisations on the raw AMIS score

D.2.1 4 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

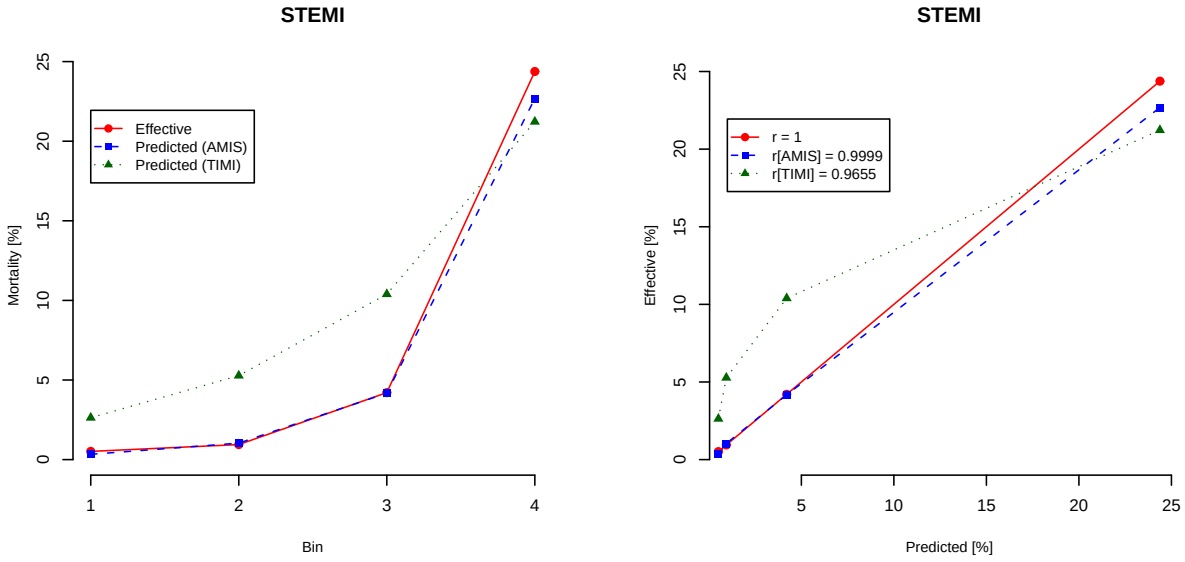


Figure: Predicted-Effective Plots for a percentile-based 4 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

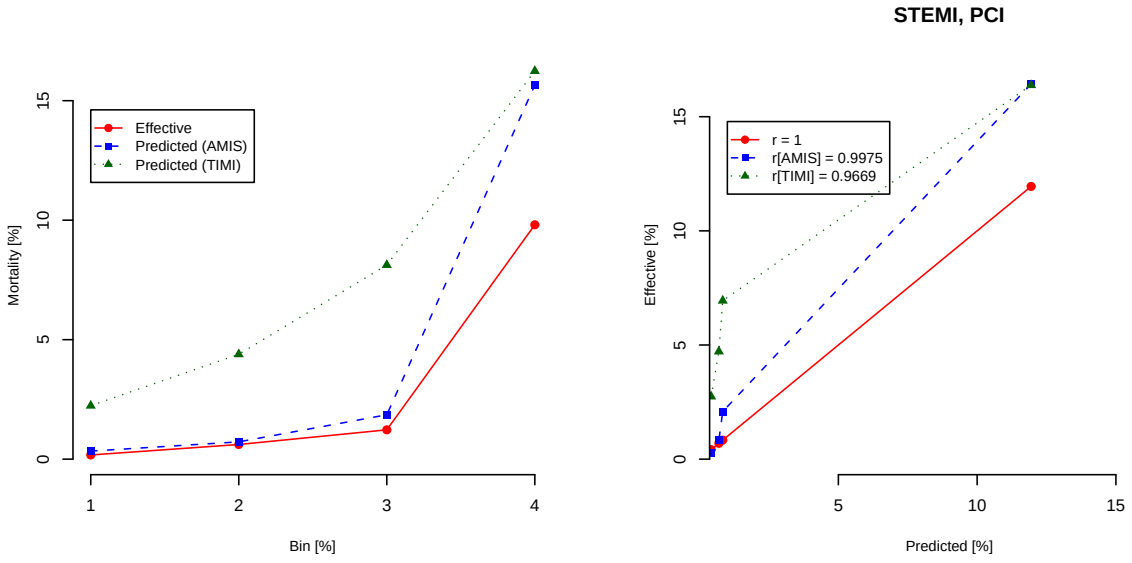


Figure: Predicted-Effective Plots for a percentile-based 4 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	99.41 %	1165	2	0.21 %	2.78 %	0.17%	38.91 %	$\approx$ 0.0 %	0.0 %
2	99.41 %	97.63 %	1165	15	1.13 %	4.91 %	1.29%	31.71 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	97.63 %	92.31 %	1165	49	4.05 %	10.69 %	4.21%	39.75 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	92.31 %	46.75 %	1165	284	23.0 %	21.14 %	24.38%	12.48 %	0.51 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =120.6594 (P-value 0.0 %)

H_{AMIS} =1.645 (P-value 43.93 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 4 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	99.41 %	720	1	0.12 %	2.58 %	0.14%	45.77 %	$\approx$ 0.0 %	0.0 %
2	99.41 %	98.82 %	720	4	0.88 %	4.29 %	0.56%	12.06 %	$\approx$ 0.0 %	0.0 %
3	98.82 %	96.45 %	720	12	2.48 %	7.25 %	1.67%	4.45 %	$\approx$ 0.0 %	0.0 %
4	96.45 %	46.75 %	720	82	16.39 %	16.62 %	11.39%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.02 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =89.1807 (P-value 0.0 %)

H_{AMIS} =15.9677 (P-value 0.03 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 4 bin discretisation based on the AMIS score.

Error Plots

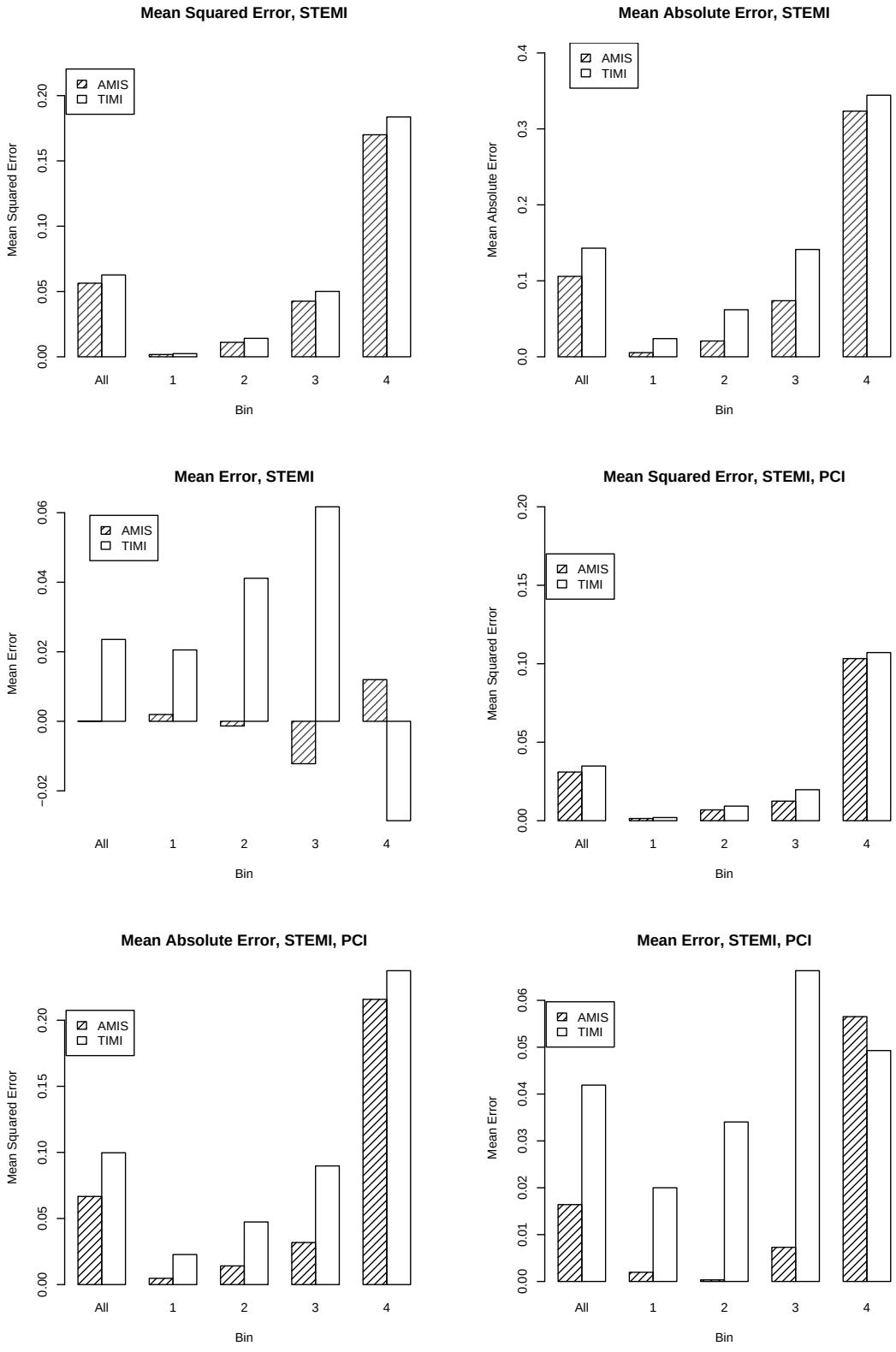


Figure: ErrorPlots for a percentile-based 4 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	1165	2	0.0017	0.0028	-0.0011	-4.5421	≈ 0.0 %	AMIS
2	1165	15	0.0127	0.0151	-0.0024	-4.1163	≈ 0.0 %	AMIS
3	1165	49	0.0401	0.0474	-0.0073	-5.6015	≈ 0.0 %	AMIS
4	1165	284	0.1672	0.1849	-0.0178	-4.4263	≈ 0.0 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 4 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	1165	2	0.0038	0.0292	-0.0255	-34.3894	≈ 0.0 %	AMIS
2	1165	15	0.0239	0.0599	-0.036	-28.0283	≈ 0.0 %	AMIS
3	1165	49	0.0787	0.1385	-0.0598	-29.6109	≈ 0.0 %	AMIS
4	1165	284	0.3243	0.3438	-0.0194	-4.6238	≈ 0.0 %	AMIS

Table: Mean Absolute Error Statistics for a percentile-based 4 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	1165	2	$3.0 \cdot 10^{-4}$	0.0261	-0.0257	35.1221	0.0 %	AMIS
2	1165	15	-0.0016	0.0362	-0.0378	30.3935	0.0 %	AMIS
3	1165	49	-0.0015	0.0649	-0.0664	36.2441	0.0 %	AMIS
4	1165	284	-0.0138	-0.0323	0.0186	-4.4112	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 4 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	720	1	0.0014	0.0025	-0.0011	-10.823	≈ 0.0 %	AMIS
2	720	4	0.0055	0.0087	-0.0032	-11.8387	≈ 0.0 %	AMIS
3	720	12	0.0164	0.022	-0.0056	-7.4545	≈ 0.0 %	AMIS
4	720	82	0.0937	0.1044	-0.0107	-2.5936	0.48 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 4 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	720	1	0.0026	0.0271	-0.0245	-30.1577	≈ 0.0 %	AMIS
2	720	4	0.0142	0.0482	-0.034	-22.8135	≈ 0.0 %	AMIS
3	720	12	0.0405	0.0867	-0.0462	-23.9309	≈ 0.0 %	AMIS
4	720	82	0.2129	0.235	-0.0221	-4.4713	≈ 0.0 %	AMIS

Table: Mean Absolute Error Statistics for a percentile-based 4 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	720	1	$-1.0 \cdot 10^{-4}$	0.0244	-0.0245	30.2596	0.0 %	AMIS
2	720	4	0.0032	0.0374	-0.0341	23.0169	0.0 %	AMIS
3	720	12	0.0081	0.0559	-0.0478	25.4536	0.0 %	AMIS
4	720	82	0.05	0.0523	-0.0024	0.4735	31.8 %	?

Table: Mean Error Statistics for a percentile-based 4 bin discretisation based on the AMIS score. STEMI, PCI

D.2.2 5 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

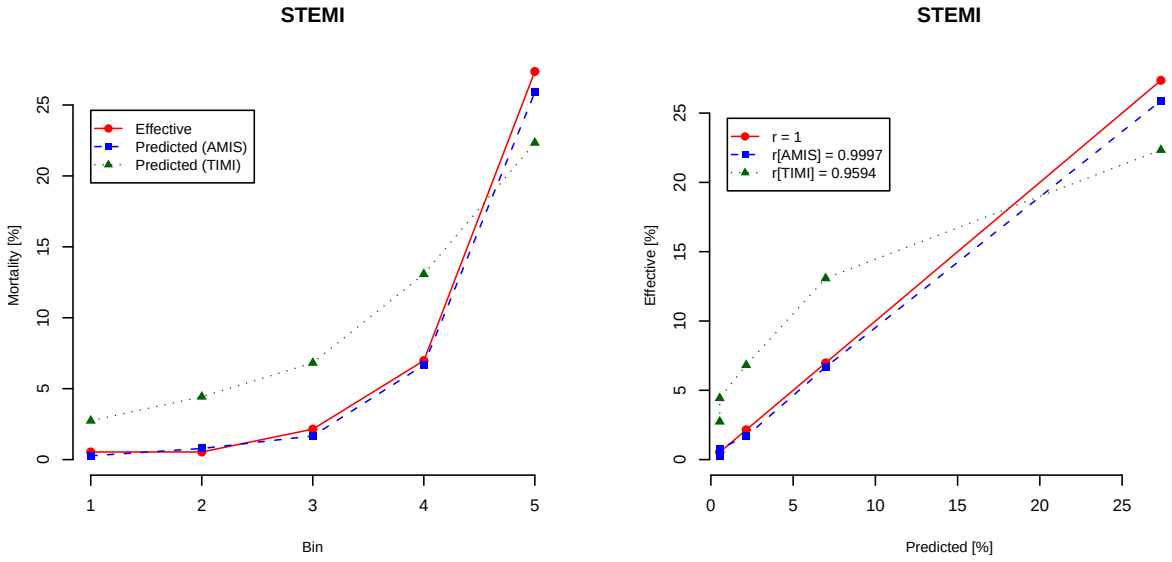


Figure: Predicted-Effective Plots for a percentile-based 5 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

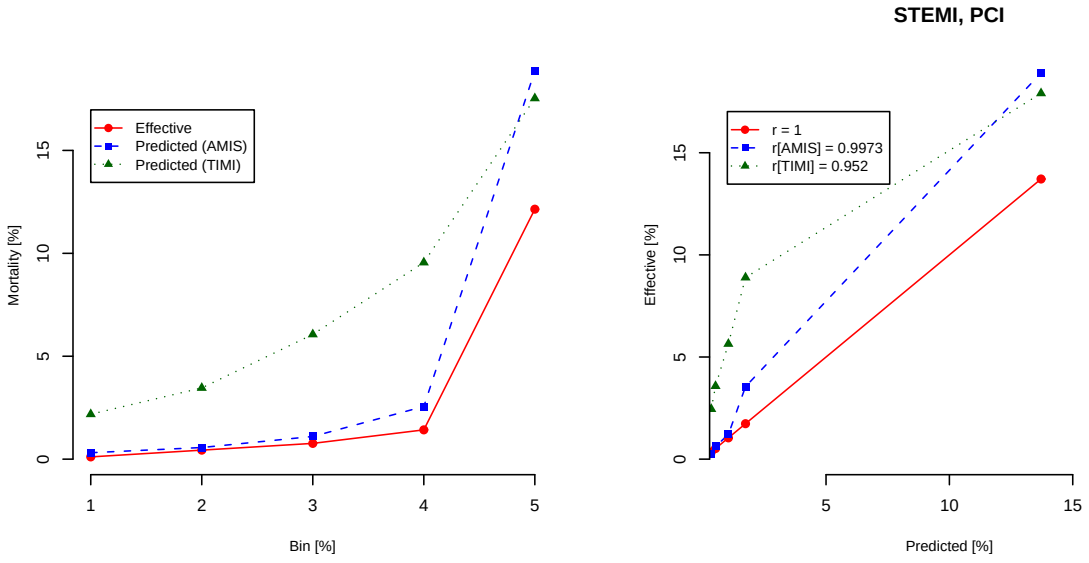


Figure: Predicted-Effective Plots for a percentile-based 5 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	99.41 %	932	0	0.11 %	2.5 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	99.41 %	98.82 %	932	10	0.83 %	4.38 %	1.07%	23.43 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	98.82 %	97.04 %	932	24	2.08 %	7.26 %	2.58%	16.91 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	97.04 %	88.17 %	932	56	6.24 %	13.02 %	6.01%	38.14 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	88.17 %	46.75 %	932	260	26.23 %	22.24 %	27.9%	11.83 %	0.01 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =136.3284 (P-value 0.0 %)

H_{AMIS} =4.2644 (P-value 23.43 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 5 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	99.41 %	576	0	0.01 %	2.62 %	0.0%	0.23 %	$\approx$ 0.0 %	0.0 %
2	99.41 %	98.82 %	576	2	0.66 %	3.56 %	0.35%	10.37 %	$\approx$ 0.0 %	0.0 %
3	98.82 %	97.63 %	576	5	1.42 %	5.58 %	0.87%	7.55 %	$\approx$ 0.0 %	0.0 %
4	97.63 %	93.53 %	576	14	3.47 %	9.07 %	2.43%	5.38 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	93.53 %	46.75 %	576	78	19.28 %	17.61 %	13.54%	$\approx$ 0.0 %	0.24 %	0.46 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =94.426 (P-value 0.0 %)

H_{AMIS} =16.1964 (P-value 0.1 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 5 bin discretisation based on the AMIS score.

Error Plots

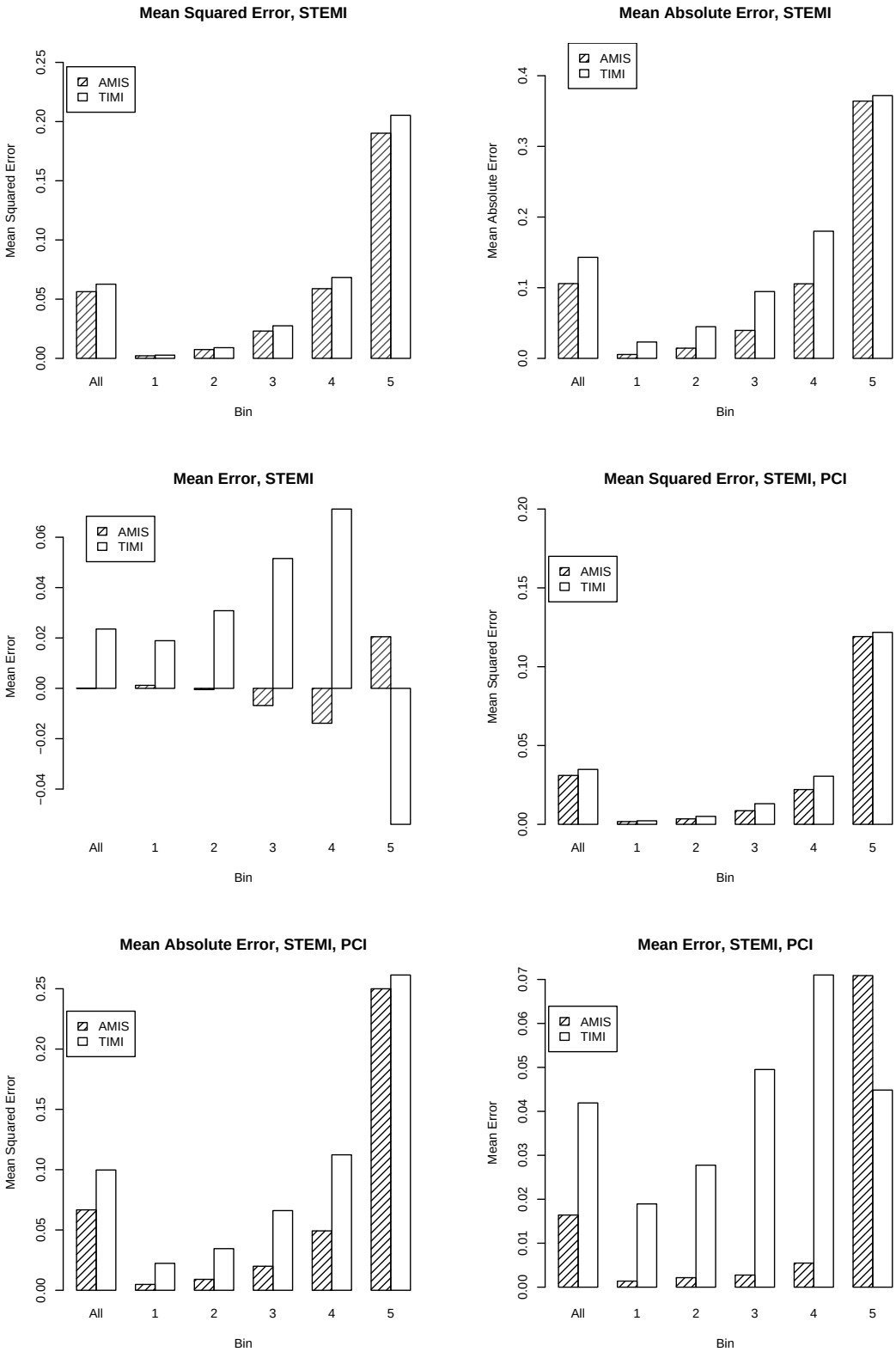


Figure: ErrorPlots for a percentile-based 5 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	932	0	≈ 0.0	0.0010	-0.0010	-13.8304	≈ 0.0 %	AMIS
2	932	10	0.0106	0.0129	-0.0023	-3.8098	0.01 %	AMIS
3	932	24	0.025	0.0288	-0.0038	-3.8904	0.01 %	AMIS
4	932	56	0.0558	0.0636	-0.0078	-4.3502	≈ 0.0 %	AMIS
5	932	260	0.1857	0.2065	-0.0208	-4.2583	≈ 0.0 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 5 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	932	0	0.0011	0.025	-0.0239	-35.5798	≈ 0.0 %	AMIS
2	932	10	0.0188	0.0531	-0.0343	-25.1233	≈ 0.0 %	AMIS
3	932	24	0.0453	0.0932	-0.0479	-26.2838	≈ 0.0 %	AMIS
4	932	56	0.1136	0.1713	-0.0577	-22.4462	≈ 0.0 %	AMIS
5	932	260	0.3595	0.3716	-0.0121	-2.4173	0.78 %	AMIS

Table: Mean Absolute Error Statistics for a percentile-based 5 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	932	0	0.0011	0.025	-0.0239	35.5798	0.0 %	AMIS
2	932	10	-0.0024	0.0331	-0.0355	26.7451	0.0 %	AMIS
3	932	24	-0.0050	0.0468	-0.0518	30.3765	0.0 %	AMIS
4	932	56	0.0023	0.0701	-0.0678	29.5851	0.0 %	AMIS
5	932	260	-0.0167	-0.0566	0.0399	-8.2259	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 5 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	576	0	≈ 0.0	0.0012	-0.0012	-11.0492	≈ 0.0 %	AMIS
2	576	2	0.0035	0.0059	-0.0024	-8.4442	≈ 0.0 %	AMIS
3	576	5	0.0087	0.0128	-0.0042	-6.0278	≈ 0.0 %	AMIS
4	576	14	0.0239	0.0318	-0.0079	-7.3932	≈ 0.0 %	AMIS
5	576	78	0.1102	0.1204	-0.0102	-1.9902	2.33 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 5 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	576	0	$1.0 \cdot 10^{-4}$	0.0262	-0.0261	-28.4937	≈ 0.0 %	AMIS
2	576	2	0.01	0.0389	-0.0289	-19.0542	≈ 0.0 %	AMIS
3	576	5	0.0227	0.0634	-0.0407	-21.4545	≈ 0.0 %	AMIS
4	576	14	0.0573	0.1108	-0.0535	-21.801	≈ 0.0 %	AMIS
5	576	78	0.2478	0.257	-0.0092	-1.5652	5.88 %	?

Table: Mean Absolute Error Statistics for a percentile-based 5 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	576	0	$1.0 \cdot 10^{-4}$	0.0262	-0.0261	28.4937	0.0 %	AMIS
2	576	2	0.0031	0.0321	-0.029	19.2057	0.0 %	AMIS
3	576	5	0.0056	0.0471	-0.0415	22.245	0.0 %	AMIS
4	576	14	0.0104	0.0664	-0.056	23.8151	0.0 %	AMIS
5	576	78	0.0574	0.0407	0.0167	-2.8577	0.21 %	TIMI

Table: Mean Error Statistics for a percentile-based 5 bin discretisation based on the AMIS score. STEMI, PCI

D.2.3 6 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

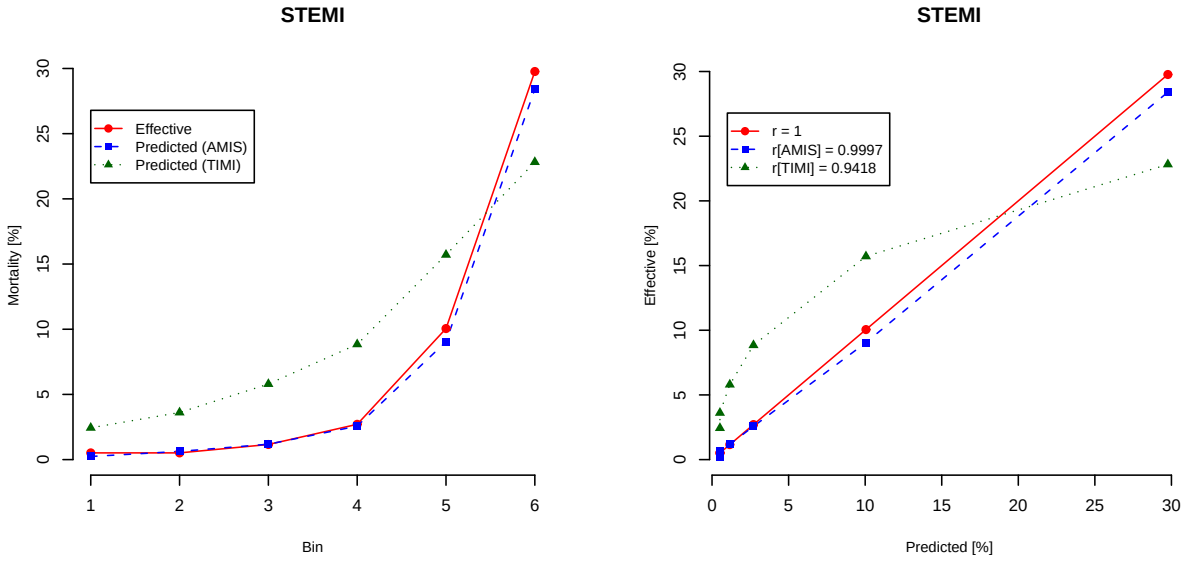


Figure: Predicted-Effective Plots for a percentile-based 6 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

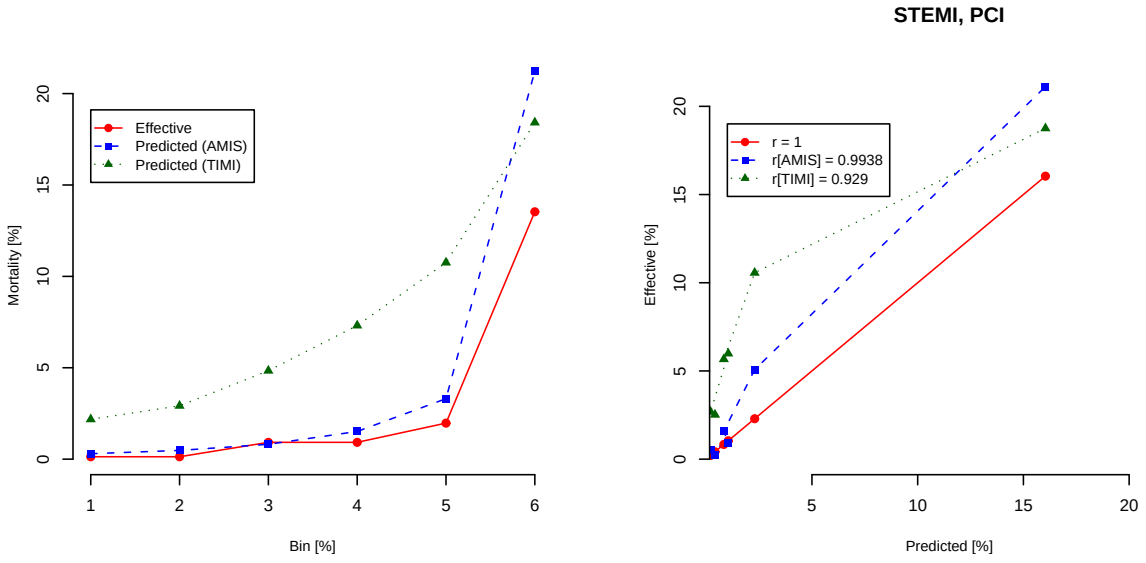


Figure: Predicted-Effective Plots for a percentile-based 6 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	99.41 %	776	0	0.01 %	2.54 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	99.41 %	98.82 %	776	5	0.64 %	3.56 %	0.64%	49.21 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	98.82 %	97.63 %	776	12	1.35 %	5.41 %	1.55%	32.7 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	97.63 %	95.86 %	776	26	3.04 %	9.05 %	3.35%	31.81 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	95.86 %	86.98 %	776	71	8.58 %	15.7 %	9.15%	28.89 %	$\approx$ 0.0 %	0.01 %
6	86.98 %	46.75 %	776	232	28.72 %	22.96 %	29.9%	23.0 %	$\approx$ 0.0 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =138.9909 (P-value 0.0 %)

H_{AMIS} =1.4304 (P-value 83.89 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 6 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	480	0	0.0 %	2.63 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	480	1	0.48 %	3.51 %	0.21%	9.52 %	$\approx$ 0.0 %	0.0 %
3	99.41 %	98.82 %	480	4	1.02 %	4.16 %	0.83%	32.24 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	98.82 %	97.06 %	480	5	2.06 %	7.1 %	1.04%	1.41 %	$\approx$ 0.0 %	0.0 %
5	97.06 %	92.31 %	480	14	4.53 %	10.03 %	2.92%	1.89 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	92.31 %	46.75 %	480	75	21.71 %	18.69 %	15.63%	0.01 %	3.39 %	3.05 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =98.3596 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 6 bin discretisation based on the AMIS score.

Error Plots

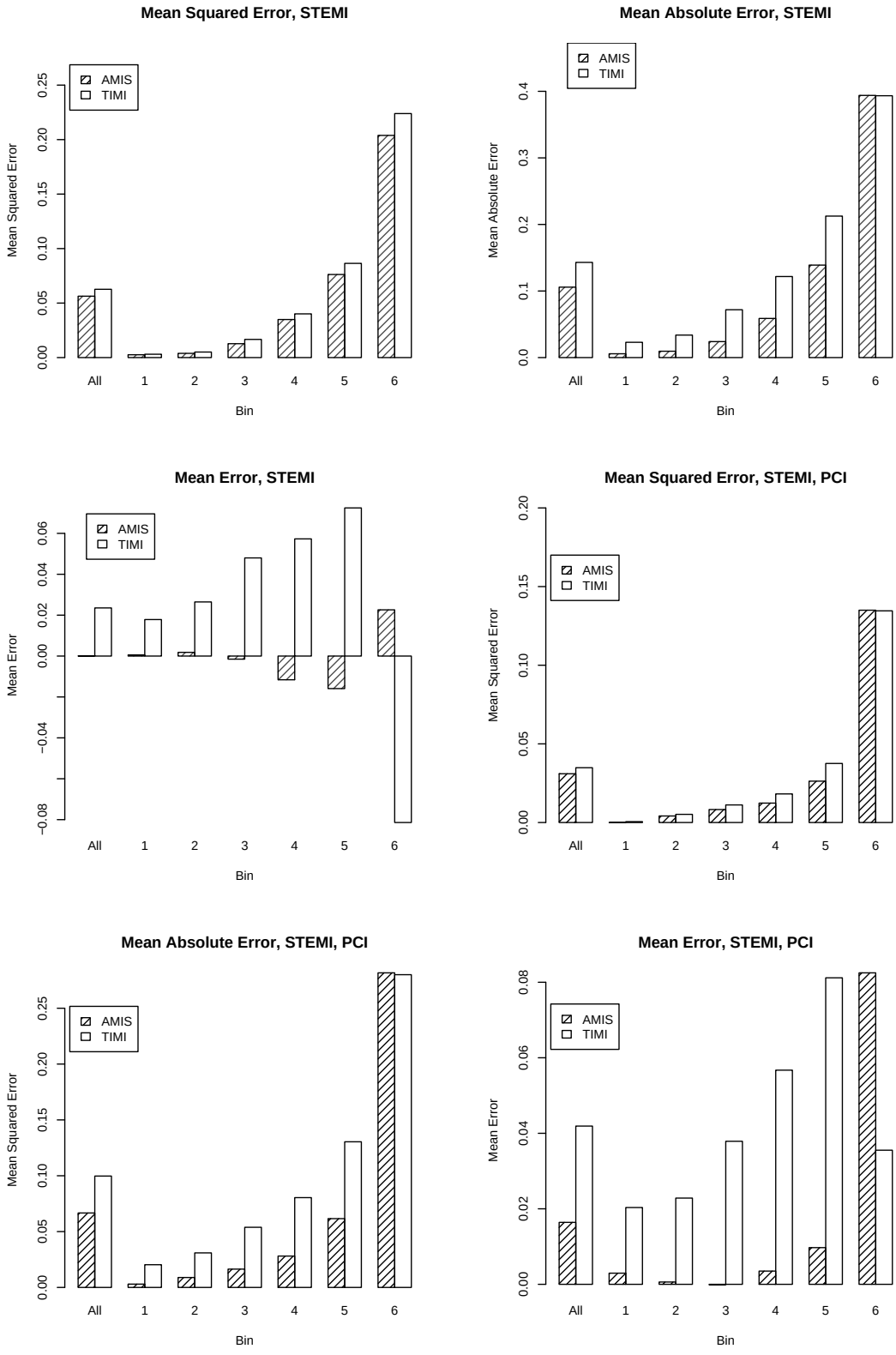


Figure: ErrorPlots for a percentile-based 6 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	776	0	≈ 0.0	0.0011	-0.0011	-12.9853	≈ 0.0 %	AMIS
2	776	5	0.0064	0.0082	-0.0018	-3.6431	0.01 %	AMIS
3	776	12	0.0152	0.0176	-0.0024	-2.9715	0.15 %	AMIS
4	776	26	0.0324	0.0383	-0.0059	-5.0549	≈ 0.0 %	AMIS
5	776	71	0.0821	0.0901	-0.0080	-3.3796	0.04 %	AMIS
6	776	232	0.1956	0.2173	-0.0217	-3.8326	0.01 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 6 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	776	0	$1.0 \cdot 10^{-4}$	0.0254	-0.0253	-34.161	≈ 0.0 %	AMIS
2	776	5	0.0127	0.0412	-0.0284	-21.3821	≈ 0.0 %	AMIS
3	776	12	0.0285	0.0669	-0.0384	-23.7095	≈ 0.0 %	AMIS
4	776	26	0.0619	0.1174	-0.0555	-25.2892	≈ 0.0 %	AMIS
5	776	71	0.1596	0.2166	-0.057	-18.1366	≈ 0.0 %	AMIS
6	776	232	0.3813	0.3864	-0.0051	-0.8842	18.83 %	?

Table: Mean Absolute Error Statistics for a percentile-based 6 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	776	0	$1.0 \cdot 10^{-4}$	0.0254	-0.0253	34.161	0.0 %	AMIS
2	776	5	$-1.0 \cdot 10^{-4}$	0.0292	-0.0292	22.325	0.0 %	AMIS
3	776	12	-0.0020	0.0386	-0.0406	26.2342	0.0 %	AMIS
4	776	26	-0.0031	0.057	-0.06	29.5039	0.0 %	AMIS
5	776	71	-0.0057	0.0655	-0.0712	26.0049	0.0 %	AMIS
6	776	232	-0.0117	-0.0694	0.0577	-10.7234	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 6 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	480	0	0.0	0.0012	-0.0012	-9.9888	≈ 0.0 %	AMIS
2	480	1	0.0021	0.0045	-0.0024	-8.3203	≈ 0.0 %	AMIS
3	480	4	0.0082	0.011	-0.0028	-9.2711	≈ 0.0 %	AMIS
4	480	5	0.0104	0.0162	-0.0059	-6.1893	≈ 0.0 %	AMIS
5	480	14	0.0289	0.0385	-0.0096	-7.7031	≈ 0.0 %	AMIS
6	480	75	0.1259	0.1349	-0.0091	-1.4825	6.92 %	?

Table: Mean Squared Error Statistics for a percentile-based 6 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	480	0	0.0	0.0263	-0.0263	-25.8664	≈ 0.0 %	AMIS
2	480	1	0.0069	0.0371	-0.0302	-18.4468	≈ 0.0 %	AMIS
3	480	4	0.0184	0.0495	-0.0311	-18.3807	≈ 0.0 %	AMIS
4	480	5	0.0305	0.0795	-0.049	-20.7124	≈ 0.0 %	AMIS
5	480	14	0.0718	0.1247	-0.0529	-18.6937	≈ 0.0 %	AMIS
6	480	75	0.2778	0.2783	$-6.0 \cdot 10^{-4}$	-0.0808	46.78 %	?

Table: Mean Absolute Error Statistics for a percentile-based 6 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	480	0	0.0	0.0263	-0.0263	25.8664	0.0 %	AMIS
2	480	1	0.0027	0.033	-0.0303	18.504	0.0 %	AMIS
3	480	4	0.0019	0.0333	-0.0314	18.6468	0.0 %	AMIS
4	480	5	0.0102	0.0606	-0.0504	21.8802	0.0 %	AMIS
5	480	14	0.0161	0.0712	-0.0551	20.0957	0.0 %	AMIS
6	480	75	0.0608	0.0306	0.0302	-4.501	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 6 bin discretisation based on the AMIS score. STEMI, PCI

D.2.4 7 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

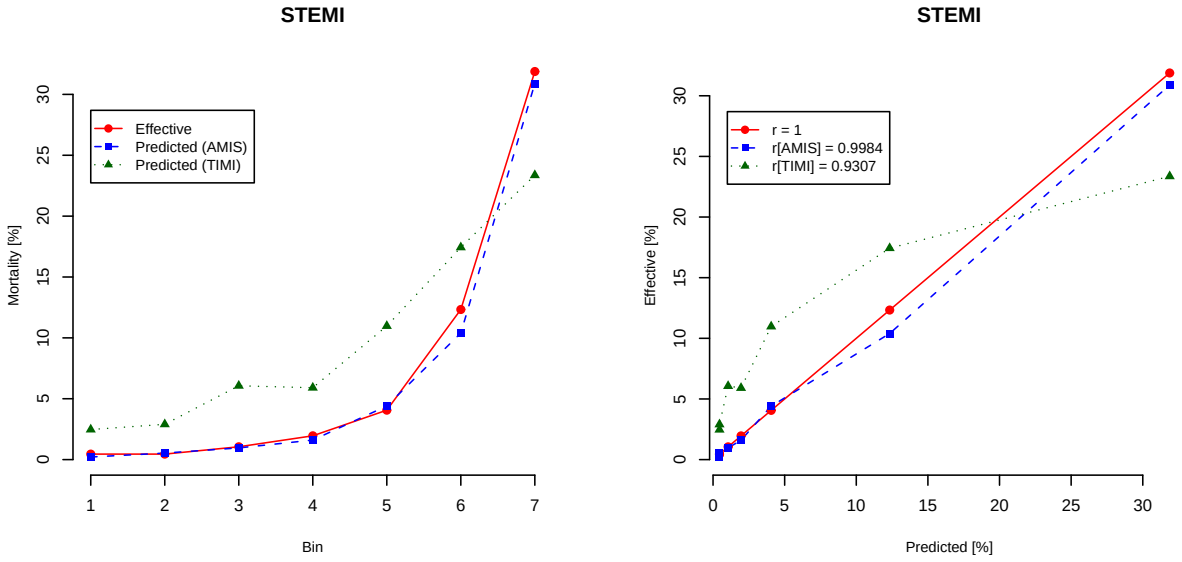


Figure: Predicted-Effective Plots for a percentile-based 7 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

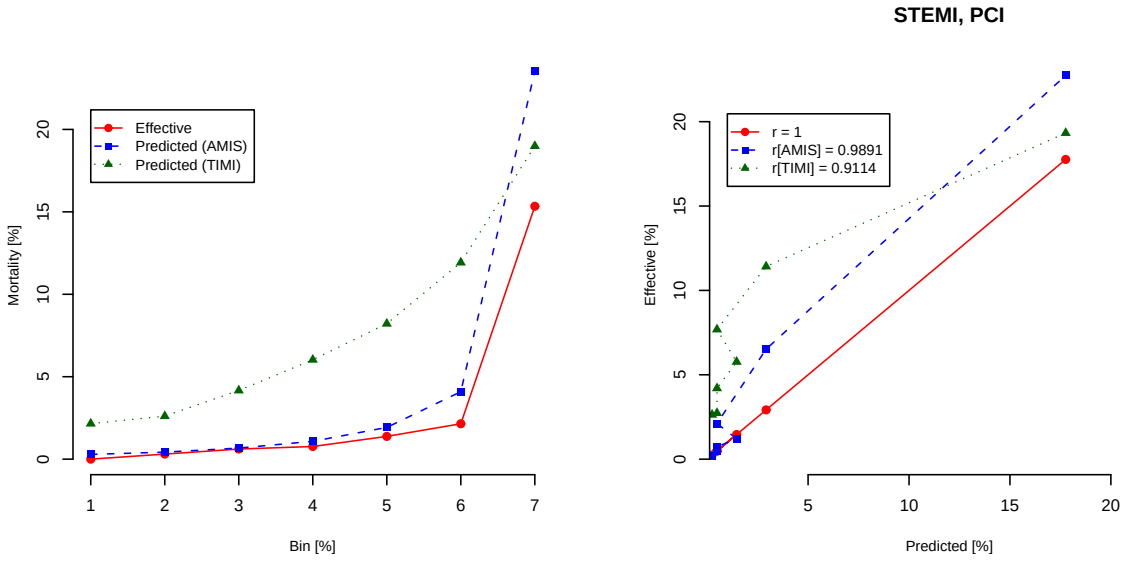


Figure: Predicted-Effective Plots for a percentile-based 7 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	665	0	0.0 %	2.54 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	665	4	0.51 %	3.51 %	0.6%	37.68 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	99.41 %	98.82 %	665	7	1.04 %	4.5 %	1.05%	48.8 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	98.82 %	97.06 %	665	17	2.08 %	7.44 %	2.56%	21.69 %	$\approx$ 0.0 %	0.02 %
5	97.06 %	93.53 %	665	28	3.98 %	10.7 %	4.21%	38.51 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	93.53 %	85.8 %	665	77	10.56 %	17.44 %	11.58%	20.41 %	$\approx$ 0.0 %	0.36 %
7	85.8 %	46.75 %	665	212	31.17 %	22.93 %	31.88%	34.21 %	$\approx$ 0.0 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =150.6289 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 7 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	411	0	0.0 %	2.63 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	411	1	0.36 %	2.71 %	0.24%	31.04 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	99.41 %	98.82 %	411	3	0.8 %	4.38 %	0.73%	43.42 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	98.82 %	98.22 %	411	2	1.34 %	5.03 %	0.49%	0.7 %	$\approx$ 0.0 %	0.0 %
5	98.22 %	96.47 %	411	9	2.69 %	8.08 %	2.19%	24.27 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	96.47 %	89.94 %	411	15	5.71 %	12.09 %	3.65%	1.28 %	$\approx$ 0.0 %	$\approx$ 0.0 %
7	89.94 %	46.75 %	411	66	23.51 %	18.71 %	16.06%	$\approx$ 0.0 %	7.43 %	11.96 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =100.0846 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 7 bin discretisation based on the AMIS score.

Error Plots

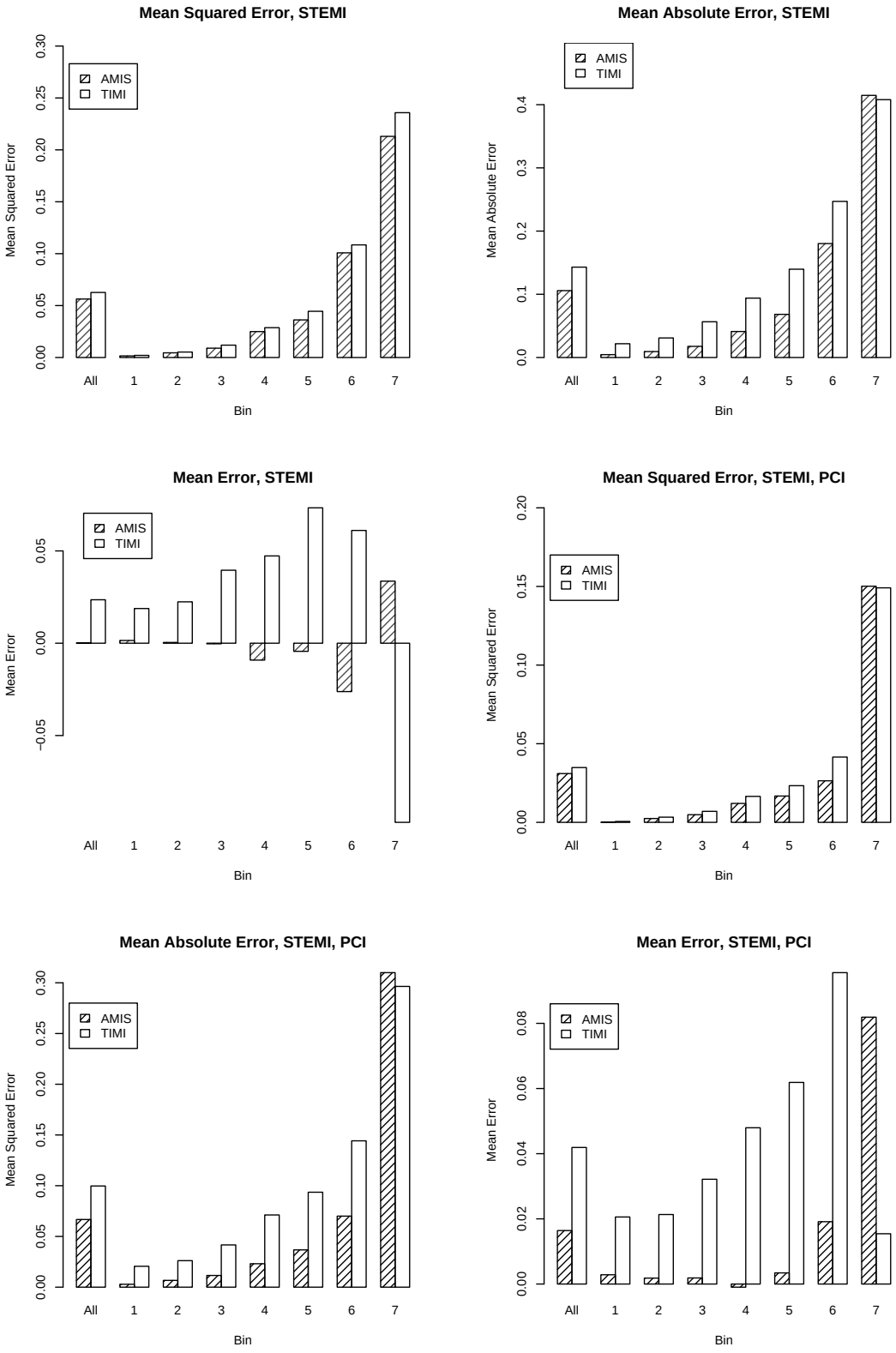


Figure: ErrorPlots for a percentile-based 7 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	665	0	0.0	0.0011	-0.0011	-11.7655	≈ 0.0 %	AMIS
2	665	4	0.0060	0.0076	-0.0016	-2.8891	0.19 %	AMIS
3	665	7	0.0104	0.0129	-0.0025	-3.8206	0.01 %	AMIS
4	665	17	0.0249	0.0281	-0.0032	-2.4634	0.69 %	AMIS
5	665	28	0.0401	0.0473	-0.0072	-4.1278	≈ 0.0 %	AMIS
6	665	77	0.1014	0.1084	-0.0069	-2.5345	0.56 %	AMIS
7	665	212	0.204	0.2287	-0.0247	-3.8501	0.01 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 7 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	665	0	0.0	0.0254	-0.0254	-31.1339	≈ 0.0 %	AMIS
2	665	4	0.011	0.0402	-0.0292	-21.3115	≈ 0.0 %	AMIS
3	665	7	0.0207	0.0542	-0.0335	-20.8526	≈ 0.0 %	AMIS
4	665	17	0.0452	0.094	-0.0488	-22.006	≈ 0.0 %	AMIS
5	665	28	0.0782	0.1385	-0.0602	-22.4556	≈ 0.0 %	AMIS
6	665	77	0.1952	0.2456	-0.0505	-14.2535	≈ 0.0 %	AMIS
7	665	212	0.4007	0.3973	0.0034	0.5268	29.92 %	?

Table: Mean Absolute Error Statistics for a percentile-based 7 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	665	0	0.0	0.0254	-0.0254	31.1339	0.0 %	AMIS
2	665	4	$-9.0 \cdot 10^{-4}$	0.0291	-0.03	22.3436	0.0 %	AMIS
3	665	7	$-1.0 \cdot 10^{-4}$	0.0345	-0.0346	22.0471	0.0 %	AMIS
4	665	17	-0.0048	0.0488	-0.0536	26.2027	0.0 %	AMIS
5	665	28	-0.0023	0.0649	-0.0671	27.7008	0.0 %	AMIS
6	665	77	-0.0102	0.0586	-0.0688	22.6557	0.0 %	AMIS
7	665	212	-0.0071	-0.0895	0.0823	-14.7333	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 7 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	411	0	0.0	0.0012	-0.0012	-9.1698	≈ 0.0 %	AMIS
2	411	1	0.0024	0.0036	-0.0012	-6.894	≈ 0.0 %	AMIS
3	411	3	0.0072	0.0107	-0.0035	-8.8953	≈ 0.0 %	AMIS
4	411	2	0.0049	0.0088	-0.0039	-11.534	≈ 0.0 %	AMIS
5	411	9	0.0214	0.0279	-0.0066	-5.3649	≈ 0.0 %	AMIS
6	411	15	0.0352	0.0469	-0.0117	-5.663	≈ 0.0 %	AMIS
7	411	66	0.1322	0.1386	-0.0064	-0.9317	17.58 %	?

Table: Mean Squared Error Statistics for a percentile-based 7 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	411	0	0.0	0.0263	-0.0263	-23.704	≈ 0.0 %	AMIS
2	411	1	0.0060	0.0295	-0.0234	-20.3972	≈ 0.0 %	AMIS
3	411	3	0.0151	0.0507	-0.0356	-17.1525	≈ 0.0 %	AMIS
4	411	2	0.0181	0.0549	-0.0367	-18.2403	≈ 0.0 %	AMIS
5	411	9	0.0475	0.0993	-0.0517	-18.5471	≈ 0.0 %	AMIS
6	411	15	0.0887	0.1479	-0.0592	-16.882	≈ 0.0 %	AMIS
7	411	66	0.2945	0.2819	0.0126	1.6735	4.72 %	?

Table: Mean Absolute Error Statistics for a percentile-based 7 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	411	0	0.0	0.0263	-0.0263	23.704	0.0 %	AMIS
2	411	1	0.0012	0.0247	-0.0235	20.5097	0.0 %	AMIS
3	411	3	$7.0 \cdot 10^{-4}$	0.0365	-0.0358	17.297	0.0 %	AMIS
4	411	2	0.0085	0.0454	-0.0369	18.3914	0.0 %	AMIS
5	411	9	0.0050	0.0589	-0.0539	20.0661	0.0 %	AMIS
6	411	15	0.0206	0.0844	-0.0638	19.2586	0.0 %	AMIS
7	411	66	0.0745	0.0265	0.048	-6.6637	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 7 bin discretisation based on the AMIS score. STEMI, PCI

D.2.5 8 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

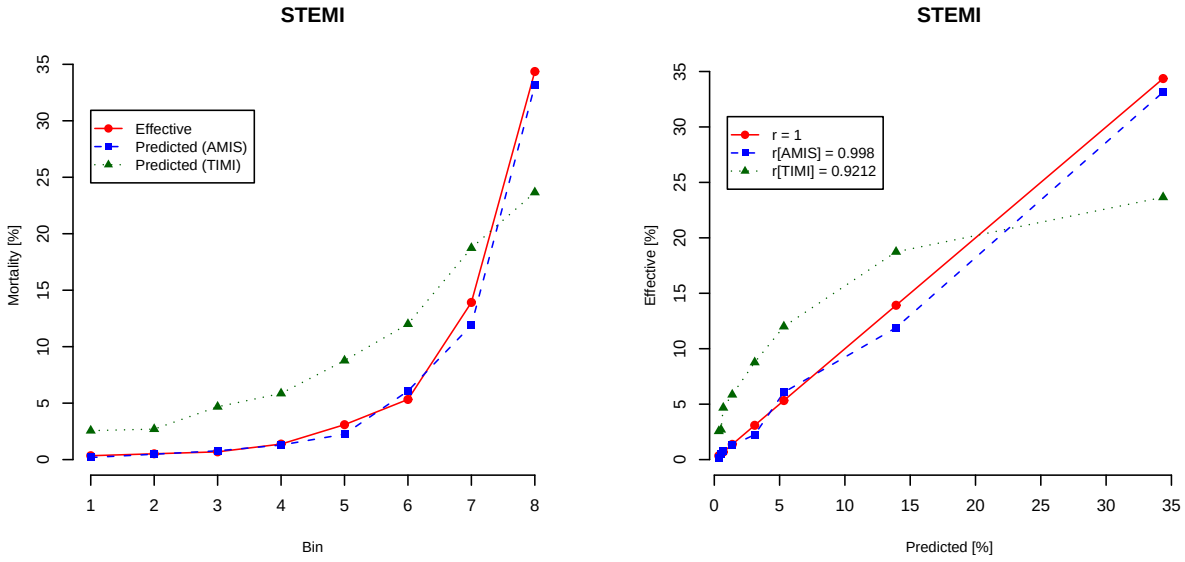


Figure: Predicted-Effective Plots for a percentile-based 8 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

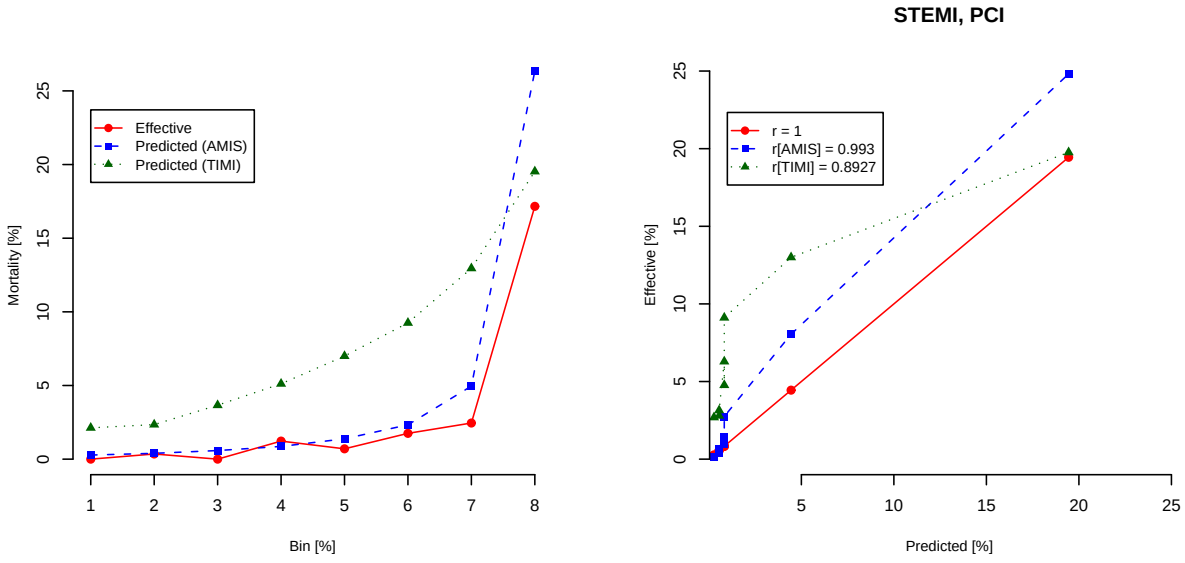


Figure: Predicted-Effective Plots for a percentile-based 8 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	582	0	0.0 %	2.56 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	582	2	0.41 %	2.99 %	0.34%	39.08 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	99.41 %	98.82 %	582	6	0.85 %	4.54 %	1.03%	33.42 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	98.82 %	97.63 %	582	9	1.4 %	5.25 %	1.55%	39.05 %	$\approx$ 0.0 %	0.01 %
5	97.63 %	96.47 %	582	19	2.82 %	8.92 %	3.26%	27.38 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	96.47 %	92.31 %	582	30	5.26 %	12.46 %	5.15%	45.35 %	$\approx$ 0.0 %	$\approx$ 0.0 %
7	92.31 %	84.02 %	582	86	12.22 %	18.97 %	14.78%	4.09 %	0.25 %	8.8 %
8	84.02 %	46.75 %	582	194	33.49 %	23.26 %	33.33%	46.67 %	$\approx$ 0.0 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =153.0996 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 8 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	360	0	0.0 %	2.66 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	360	1	0.25 %	2.5 %	0.28%	45.78 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	99.41 %	99.41 %	360	0	0.59 %	3.71 %	0.0%	0.0 %	$\approx$ 0.0 %	0.0 %
4	99.41 %	98.82 %	360	4	1.17 %	4.88 %	1.11%	45.83 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	98.82 %	97.63 %	360	3	1.81 %	6.77 %	0.83%	2.14 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	97.63 %	96.45 %	360	9	3.15 %	7.74 %	2.5%	21.62 %	$\approx$ 0.0 %	$\approx$ 0.0 %
7	96.45 %	88.17 %	360	14	7.2 %	13.72 %	3.89%	0.07 %	$\approx$ 0.0 %	$\approx$ 0.0 %
8	88.17 %	46.75 %	360	68	25.57 %	19.53 %	18.89%	0.05 %	38.01 %	10.37 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =105.4142 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 8 bin discretisation based on the AMIS score.

Error Plots

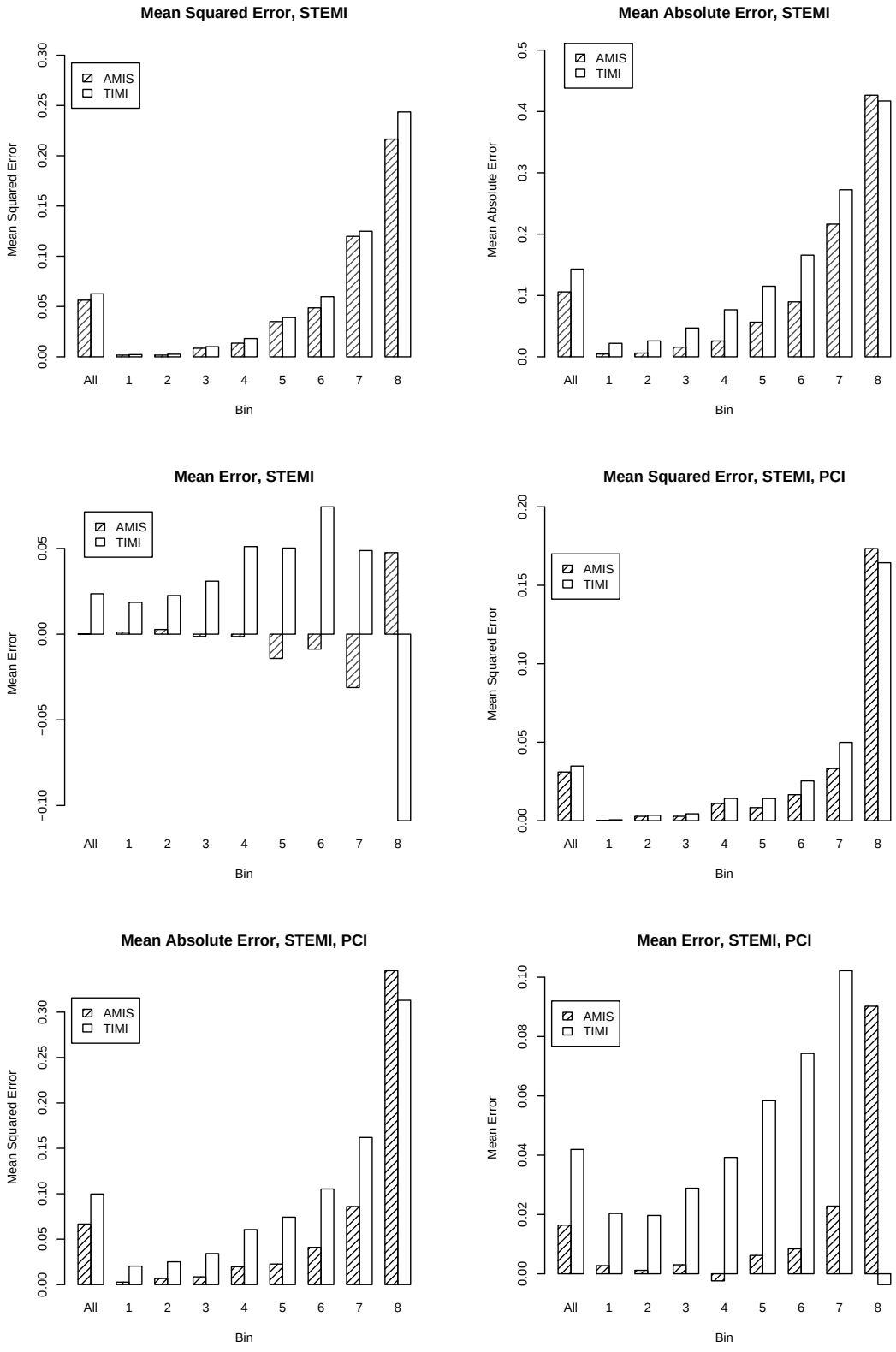


Figure: ErrorPlots for a percentile-based 8 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	582	0	0.0	0.0011	-0.0011	-11.0684	≈ 0.0 %	AMIS
2	582	2	0.0034	0.0046	-0.0011	-2.3503	0.94 %	AMIS
3	582	6	0.0102	0.0128	-0.0026	-3.3186	0.05 %	AMIS
4	582	9	0.0152	0.0174	-0.0022	-2.5105	0.6 %	AMIS
5	582	19	0.0316	0.0377	-0.0060	-4.4683	≈ 0.0 %	AMIS
6	582	30	0.0487	0.0571	-0.0084	-3.8222	0.01 %	AMIS
7	582	86	0.1258	0.13	-0.0042	-1.2917	9.83 %	?
8	582	194	0.2074	0.2362	-0.0288	-4.0018	≈ 0.0 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	582	0	0.0	0.0256	-0.0256	-28.9832	≈ 0.0 %	AMIS
2	582	2	0.0075	0.0328	-0.0253	-21.2372	≈ 0.0 %	AMIS
3	582	6	0.0186	0.0543	-0.0357	-19.8241	≈ 0.0 %	AMIS
4	582	9	0.0291	0.0653	-0.0363	-19.7343	≈ 0.0 %	AMIS
5	582	19	0.059	0.1157	-0.0567	-22.3445	≈ 0.0 %	AMIS
6	582	30	0.0983	0.1613	-0.063	-20.0552	≈ 0.0 %	AMIS
7	582	86	0.2326	0.277	-0.0444	-11.1398	≈ 0.0 %	AMIS
8	582	194	0.4138	0.4065	0.0074	1.0233	15.31 %	?

Table: Mean Absolute Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	582	0	0.0	0.0256	-0.0256	28.9832	0.0 %	AMIS
2	582	2	$7.0 \cdot 10^{-4}$	0.0265	-0.0258	22.0488	0.0 %	AMIS
3	582	6	-0.0018	0.0351	-0.0369	21.0061	0.0 %	AMIS
4	582	9	-0.0014	0.0371	-0.0385	21.8977	0.0 %	AMIS
5	582	19	-0.0044	0.0565	-0.061	25.8384	0.0 %	AMIS
6	582	30	0.0011	0.073	-0.072	25.8076	0.0 %	AMIS
7	582	86	-0.0256	0.0419	-0.0675	19.9497	0.0 %	AMIS
8	582	194	0.0016	-0.1008	0.1023	-17.616	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	360	0	0.0	0.0012	-0.0012	-8.6654	≈ 0.0 %	AMIS
2	360	1	0.0028	0.0037	-0.0010	-6.6702	≈ 0.0 %	AMIS
3	360	0	≈ 0.0	0.0029	-0.0029	-7.684	≈ 0.0 %	AMIS
4	360	4	0.011	0.0145	-0.0035	-9.0358	≈ 0.0 %	AMIS
5	360	3	0.0084	0.0138	-0.0054	-4.863	≈ 0.0 %	AMIS
6	360	9	0.0244	0.0302	-0.0058	-5.7214	≈ 0.0 %	AMIS
7	360	14	0.0389	0.0521	-0.0132	-5.5102	≈ 0.0 %	AMIS
8	360	68	0.1485	0.1568	-0.0083	-1.042	14.88 %	?

Table: Mean Squared Error Statistics for a percentile-based 8 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	360	0	0.0	0.0266	-0.0266	-22.3794	≈ 0.0 %	AMIS
2	360	1	0.0052	0.0276	-0.0224	-20.3911	≈ 0.0 %	AMIS
3	360	0	0.0059	0.0371	-0.0311	-15.179	≈ 0.0 %	AMIS
4	360	4	0.0225	0.0593	-0.0368	-17.1059	≈ 0.0 %	AMIS
5	360	3	0.0261	0.0743	-0.0483	-18.2233	≈ 0.0 %	AMIS
6	360	9	0.0549	0.099	-0.0441	-15.7122	≈ 0.0 %	AMIS
7	360	14	0.1051	0.1651	-0.06	-14.624	≈ 0.0 %	AMIS
8	360	68	0.3208	0.3049	0.0158	1.8583	3.16 %	?

Table: Mean Absolute Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	360	0	0.0	0.0266	-0.0266	22.3794	0.0 %	AMIS
2	360	1	$-3.0 \cdot 10^{-4}$	0.0222	-0.0225	20.535	0.0 %	AMIS
3	360	0	0.0059	0.0371	-0.0311	15.179	0.0 %	AMIS
4	360	4	$6.0 \cdot 10^{-4}$	0.0377	-0.0371	17.4044	0.0 %	AMIS
5	360	3	0.0097	0.0593	-0.0496	19.2309	0.0 %	AMIS
6	360	9	0.0065	0.0524	-0.0459	16.8455	0.0 %	AMIS
7	360	14	0.0332	0.0983	-0.0652	16.8147	0.0 %	AMIS
8	360	68	0.0668	0.0064	0.0604	-7.5971	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 8 bin discretisation based on the AMIS score. STEMI, PCI

D.2.6 9 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

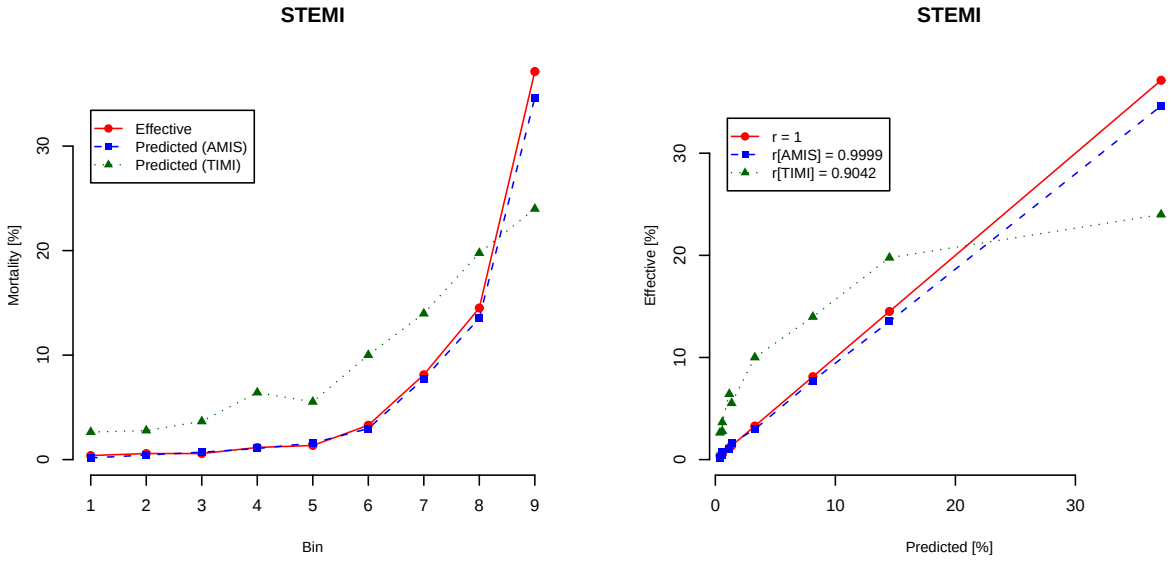


Figure: Predicted-Effective Plots for a percentile-based 9 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

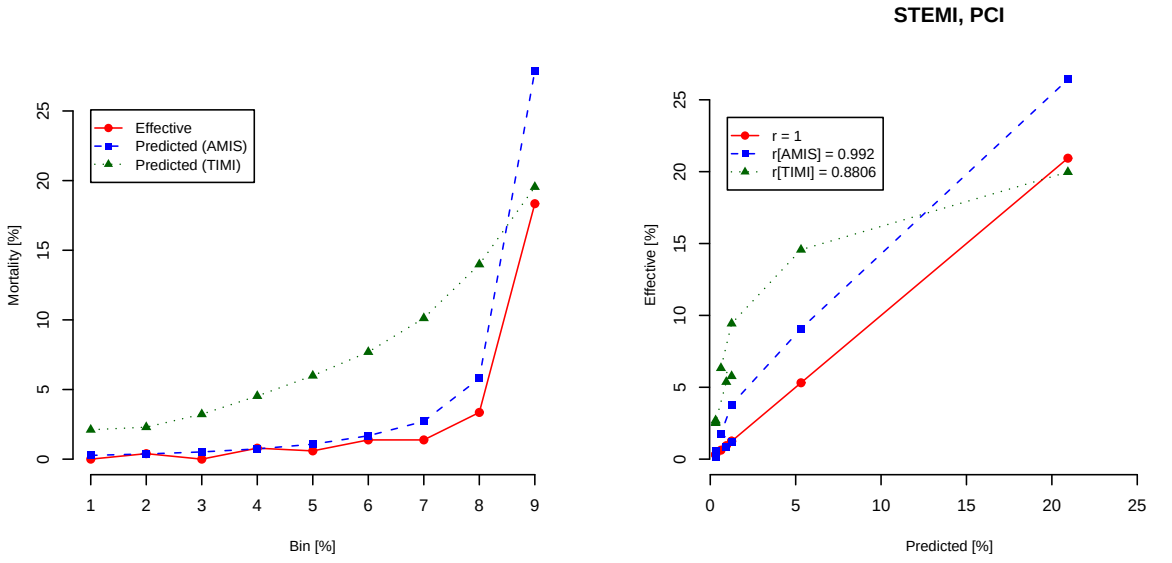


Figure: Predicted-Effective Plots for a percentile-based 9 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	517	0	0.0 %	2.55 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	517	1	0.31 %	2.46 %	0.19%	26.65 %	$\approx$ 0.0 %	$\approx$ 0.0 %
3	99.41 %	98.82 %	517	4	0.66 %	4.15 %	0.77%	38.68 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	98.82 %	98.82 %	517	8	1.18 %	5.15 %	1.55%	25.05 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	98.82 %	97.06 %	517	11	2.08 %	7.48 %	2.13%	46.88 %	$\approx$ 0.0 %	0.02 %
6	97.06 %	95.86 %	517	19	3.32 %	9.03 %	3.68%	33.4 %	$\approx$ 0.0 %	$\approx$ 0.0 %
7	95.86 %	89.94 %	517	36	6.74 %	14.97 %	6.96%	42.12 %	$\approx$ 0.0 %	0.01 %
8	89.94 %	80.0 %	517	79	13.55 %	19.3 %	15.28%	13.82 %	0.62 %	9.95 %
9	80.0 %	46.75 %	517	185	35.4 %	23.62 %	35.78%	42.64 %	$\approx$ 0.0 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =166.3317 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 9 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	320	0	0.0 %	2.65 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	320	0	0.13 %	2.58 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
3	99.41 %	99.41 %	320	1	0.59 %	3.99 %	0.31%	18.76 %	$\approx$ 0.0 %	0.0 %
4	99.41 %	98.82 %	320	3	0.95 %	3.97 %	0.94%	49.42 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	98.82 %	98.22 %	320	2	1.3 %	4.88 %	0.63%	6.39 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	98.22 %	97.06 %	320	4	2.38 %	8.04 %	1.25%	3.44 %	$\approx$ 0.0 %	$\approx$ 0.0 %
7	97.06 %	95.29 %	320	10	3.55 %	8.42 %	3.13%	33.16 %	$\approx$ 0.0 %	0.02 %
8	95.29 %	87.06 %	320	12	8.62 %	14.67 %	3.75%	$\approx$ 0.0 %	$\approx$ 0.0 %	$\approx$ 0.0 %
9	87.06 %	46.75 %	320	67	27.18 %	19.99 %	20.94%	0.27 %	34.08 %	13.27 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =110.9293 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 9 bin discretisation based on the AMIS score.

Error Plots

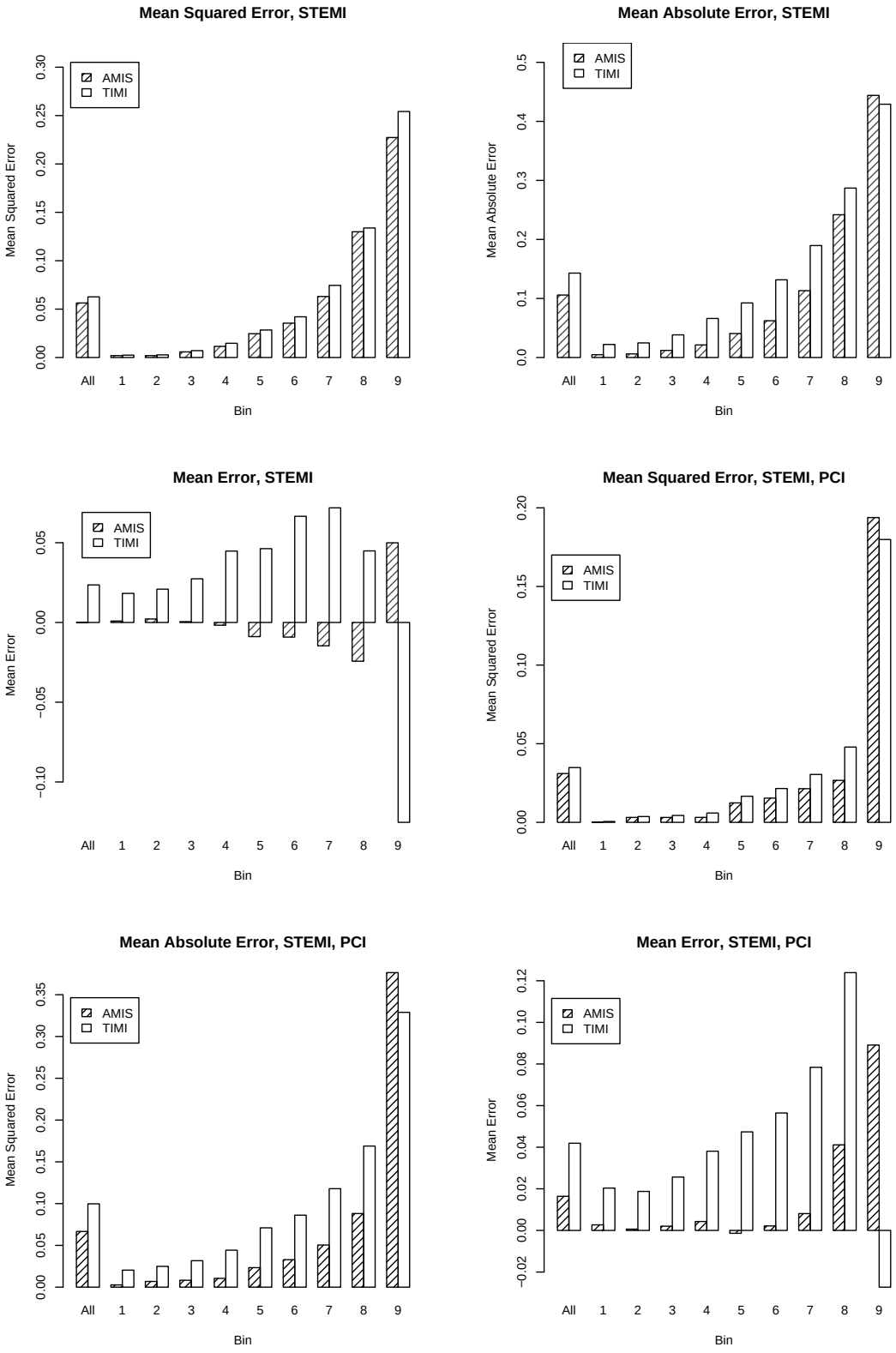


Figure: ErrorPlots for a percentile-based 9 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	517	0	0.0	0.0011	-0.0011	-10.357	≈ 0.0 %	AMIS
2	517	1	0.0019	0.0028	$-9.0 \cdot 10^{-4}$	-8.6465	≈ 0.0 %	AMIS
3	517	4	0.0077	0.01	-0.0023	-3.152	0.08 %	AMIS
4	517	8	0.0152	0.0178	-0.0025	-2.8106	0.25 %	AMIS
5	517	11	0.0208	0.0243	-0.0035	-2.3288	1.0 %	AMIS
6	517	19	0.0354	0.0418	-0.0064	-5.2111	≈ 0.0 %	AMIS
7	517	36	0.0645	0.0737	-0.0093	-3.1807	0.07 %	AMIS
8	517	79	0.1302	0.1343	-0.0041	-1.2336	10.87 %	?
9	517	185	0.2209	0.2506	-0.0297	-3.7032	0.01 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 9 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	517	0	0.0	0.0255	-0.0255	-27.1618	≈ 0.0 %	AMIS
2	517	1	0.0051	0.0264	-0.0214	-25.6173	≈ 0.0 %	AMIS
3	517	4	0.0143	0.048	-0.0337	-18.0253	≈ 0.0 %	AMIS
4	517	8	0.0269	0.0648	-0.0379	-19.5657	≈ 0.0 %	AMIS
5	517	11	0.0411	0.0906	-0.0495	-19.789	≈ 0.0 %	AMIS
6	517	19	0.0675	0.1208	-0.0532	-20.2152	≈ 0.0 %	AMIS
7	517	36	0.127	0.1956	-0.0686	-18.2992	≈ 0.0 %	AMIS
8	517	79	0.247	0.2829	-0.036	-8.7143	≈ 0.0 %	AMIS
9	517	185	0.4353	0.4229	0.0125	1.573	5.79 %	?

Table: Mean Absolute Error Statistics for a percentile-based 9 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	517	0	0.0	0.0255	-0.0255	27.1618	0.0 %	AMIS
2	517	1	0.0012	0.0226	-0.0214	25.7559	0.0 %	AMIS
3	517	4	-0.0011	0.0337	-0.0348	19.0181	0.0 %	AMIS
4	517	8	-0.0037	0.036	-0.0397	21.2422	0.0 %	AMIS
5	517	11	$-5.0 \cdot 10^{-4}$	0.0536	-0.0541	23.3853	0.0 %	AMIS
6	517	19	-0.0036	0.0535	-0.0571	23.0597	0.0 %	AMIS
7	517	36	-0.0022	0.0801	-0.0823	25.9805	0.0 %	AMIS
8	517	79	-0.0173	0.0402	-0.0575	15.8724	0.0 %	AMIS
9	517	185	-0.0038	-0.1216	0.1178	-19.5176	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 9 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	320	0	0.0	0.0012	-0.0012	-8.053	≈ 0.0 %	AMIS
2	320	0	≈ 0.0	0.0011	-0.0011	-8.6565	≈ 0.0 %	AMIS
3	320	1	0.0031	0.0063	-0.0031	-7.3508	≈ 0.0 %	AMIS
4	320	3	0.0093	0.0118	-0.0026	-7.3659	≈ 0.0 %	AMIS
5	320	2	0.0063	0.0099	-0.0037	-9.4243	≈ 0.0 %	AMIS
6	320	4	0.0124	0.0191	-0.0067	-4.8134	≈ 0.0 %	AMIS
7	320	10	0.0303	0.0354	-0.0051	-3.1496	0.08 %	AMIS
8	320	12	0.0391	0.0536	-0.0145	-5.6285	≈ 0.0 %	AMIS
9	320	67	0.1628	0.1712	-0.0084	-0.9529	17.04 %	?

Table: Mean Squared Error Statistics for a percentile-based 9 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	320	0	0.0	0.0265	-0.0265	-20.7589	≈ 0.0 %	AMIS
2	320	0	0.0013	0.0258	-0.0245	-20.8619	≈ 0.0 %	AMIS
3	320	1	0.0090	0.0429	-0.0339	-14.7843	≈ 0.0 %	AMIS
4	320	3	0.0186	0.0487	-0.0301	-15.0679	≈ 0.0 %	AMIS
5	320	2	0.0191	0.0547	-0.0356	-15.6056	≈ 0.0 %	AMIS
6	320	4	0.0357	0.0902	-0.0545	-17.6016	≈ 0.0 %	AMIS
7	320	10	0.0645	0.1096	-0.0451	-14.4878	≈ 0.0 %	AMIS
8	320	12	0.1172	0.1728	-0.0556	-11.9833	≈ 0.0 %	AMIS
9	320	67	0.3427	0.3221	0.0206	2.1916	1.42 %	TIMI

Table: Mean Absolute Error Statistics for a percentile-based 9 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	320	0	0.0	0.0265	-0.0265	20.7589	0.0 %	AMIS
2	320	0	0.0013	0.0258	-0.0245	20.8619	0.0 %	AMIS
3	320	1	0.0028	0.0367	-0.034	14.8447	0.0 %	AMIS
4	320	3	$1.0 \cdot 10^{-4}$	0.0304	-0.0303	15.261	0.0 %	AMIS
5	320	2	0.0067	0.0425	-0.0358	15.7727	0.0 %	AMIS
6	320	4	0.0113	0.0679	-0.0566	18.9872	0.0 %	AMIS
7	320	10	0.0042	0.053	-0.0487	16.5775	0.0 %	AMIS
8	320	12	0.0487	0.1092	-0.0604	13.5882	0.0 %	AMIS
9	320	67	0.0624	-0.0095	0.0719	-8.3892	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 9 bin discretisation based on the AMIS score. STEMI, PCI

D.2.7 10 Bins Percentile-based Discretisation on AMIS

Plots: All Patients (STEMI)

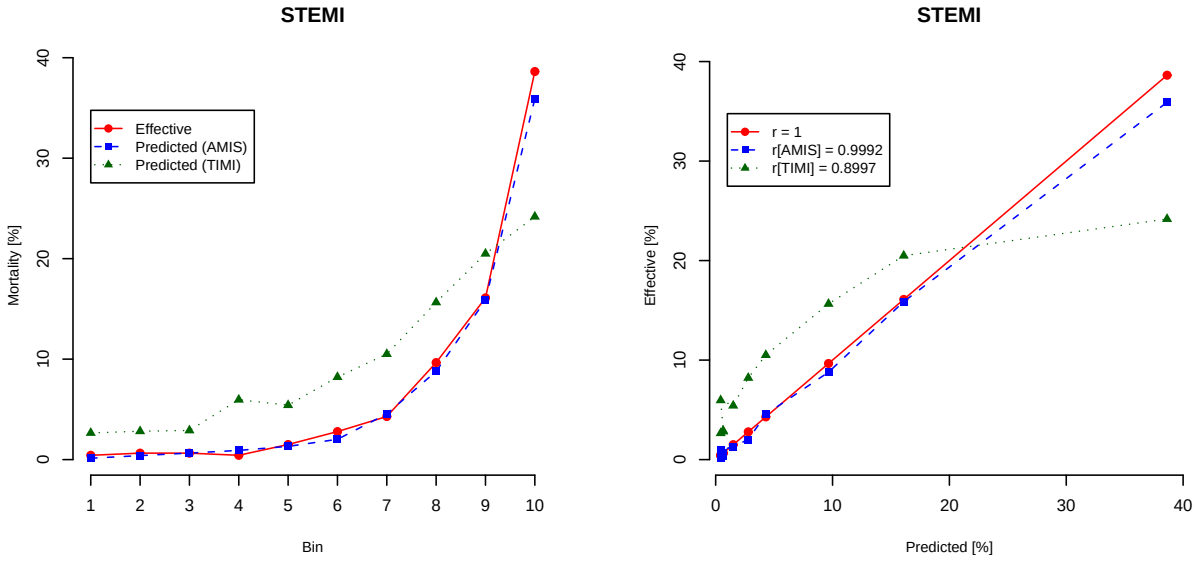


Figure: Predicted-Effective Plots for a percentile-based 10 bin discretisation based on the AMIS score.

Plots: PCI Patients (STEMI)

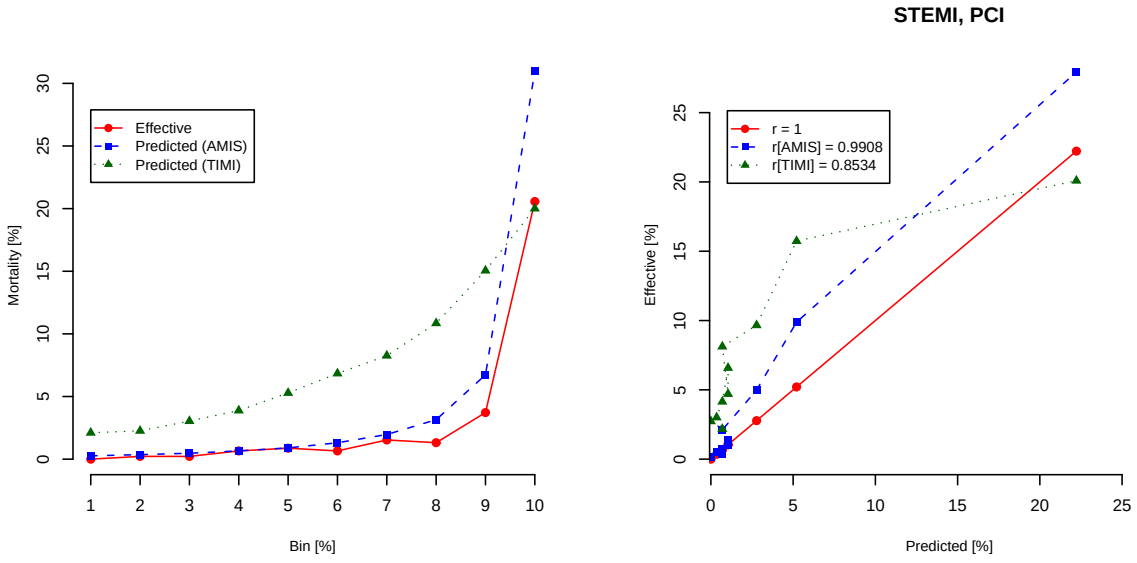


Figure: Predicted-Effective Plots for a percentile-based 10 bin discretisation based on the AMIS score. Only PCI patients.

Statistics: All Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	466	0	0.0 %	2.56 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	466	0	0.22 %	2.45 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
3	99.41 %	99.41 %	466	4	0.59 %	3.91 %	0.86%	26.59 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	99.41 %	98.82 %	466	6	1.07 %	4.85 %	1.29%	33.59 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	98.82 %	97.63 %	466	7	1.46 %	5.44 %	1.5%	47.31 %	$\approx$ 0.0 %	0.04 %
6	97.63 %	97.04 %	466	17	2.69 %	9.07 %	3.65%	13.64 %	$\approx$ 0.0 %	0.02 %
7	97.04 %	93.53 %	466	15	3.95 %	10.44 %	3.22%	18.64 %	$\approx$ 0.0 %	$\approx$ 0.0 %
8	93.53 %	88.17 %	466	41	8.54 %	15.6 %	8.8%	42.1 %	$\approx$ 0.0 %	0.38 %
9	88.17 %	80.0 %	466	84	14.87 %	20.22 %	18.03%	3.84 %	11.38 %	8.59 %
10	80.0 %	46.75 %	466	176	37.58 %	24.26 %	37.77%	46.65 %	$\approx$ 0.0 %	$\approx$ 0.0 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =168.9846 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 10 bin discretisation based on the AMIS score.

Statistics: PCI Patients (STEMI)

Bin	Interval [%]		n	$n_\dagger$	$\bar{s}$	$\bar{\tau}$	$\bar{m}$	P_A	P_T	P_{WK}
1	100.0 %	100.0 %	288	0	0.0 %	2.66 %	0.0%	$\approx$ 0.0 %	$\approx$ 0.0 %	0.0 %
2	100.0 %	99.41 %	288	0	0.02 %	2.58 %	0.0%	0.22 %	$\approx$ 0.0 %	0.0 %
3	99.41 %	99.41 %	288	1	0.59 %	3.26 %	0.35%	24.34 %	$\approx$ 0.0 %	$\approx$ 0.0 %
4	99.41 %	98.82 %	288	1	0.72 %	3.85 %	0.35%	13.87 %	$\approx$ 0.0 %	$\approx$ 0.0 %
5	98.82 %	98.82 %	288	3	1.18 %	4.82 %	1.04%	40.79 %	$\approx$ 0.0 %	$\approx$ 0.0 %
6	98.82 %	97.63 %	288	2	1.67 %	6.33 %	0.69%	2.4 %	$\approx$ 0.0 %	$\approx$ 0.0 %
7	97.63 %	96.47 %	288	6	2.76 %	8.08 %	2.08%	21.28 %	$\approx$ 0.0 %	$\approx$ 0.0 %
8	96.47 %	93.53 %	288	8	4.18 %	10.07 %	2.78%	7.55 %	$\approx$ 0.0 %	$\approx$ 0.0 %
9	93.53 %	86.98 %	288	16	9.81 %	15.34 %	5.56%	0.09 %	$\approx$ 0.0 %	$\approx$ 0.0 %
10	86.98 %	46.75 %	288	62	28.75 %	19.88 %	21.53%	0.12 %	25.18 %	15.63 %

HOSMER-LEMESHOW-statistics:

H_{TIMI} =109.7708 (P-value 0.0 %)

H_{AMIS} =$\approx$ 0.0 (P-value $\approx$ 0.0 %)

Table: Statistics:Predicted/Effective mortalities for a percentile-based 10 bin discretisation based on the AMIS score.

Error Plots

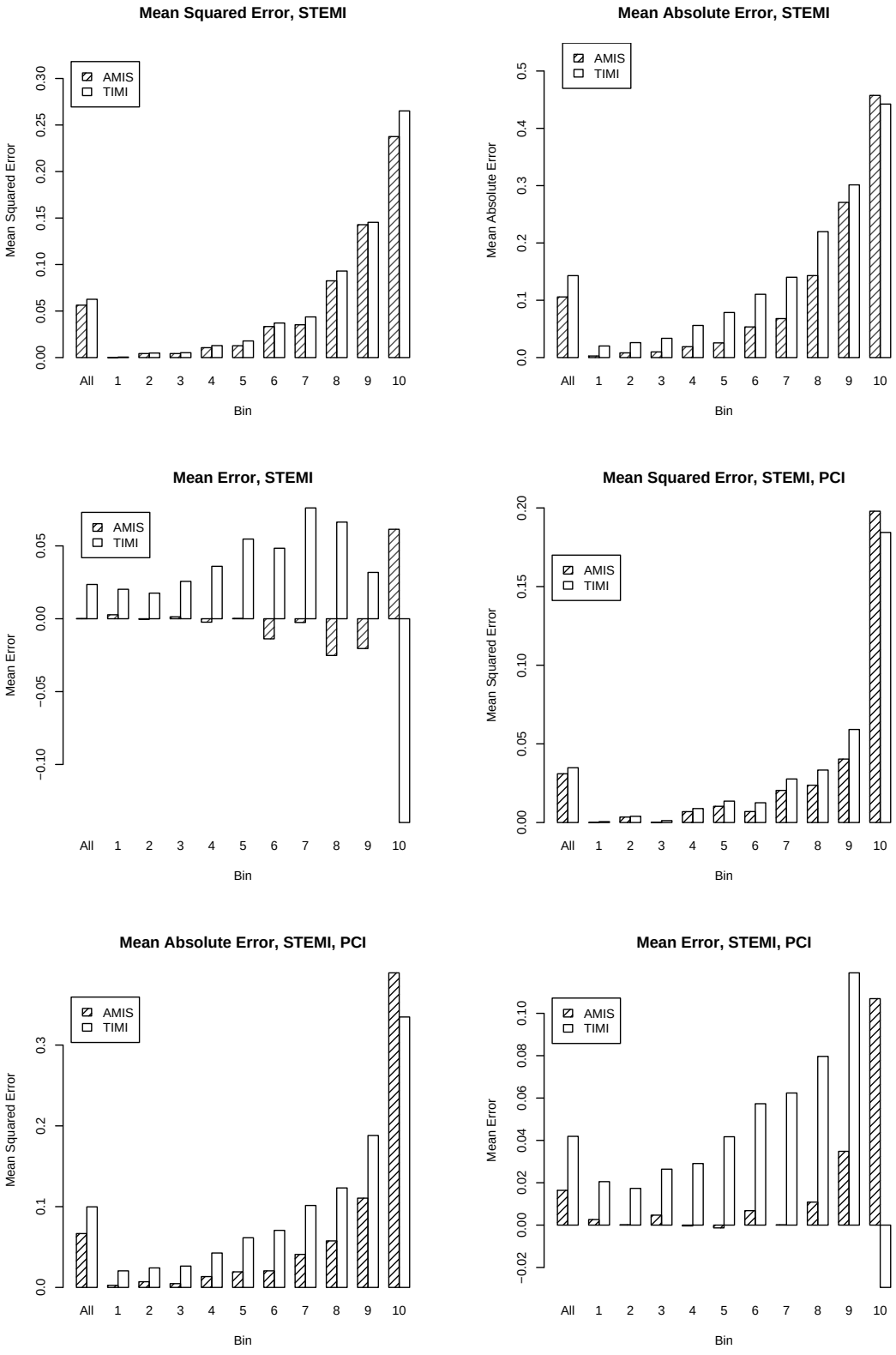


Figure: ErrorPlots for a percentile-based 10 bin discretisation based on the AMIS score.

Error Statistics

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.0555	0.0626	-0.0072	-6.7122	≈ 0.0 %	AMIS
1	466	0	0.0	0.0011	-0.0011	-9.7896	≈ 0.0 %	AMIS
2	466	0	≈ 0.0	0.0010	$-9.0 \cdot 10^{-4}$	-9.8499	≈ 0.0 %	AMIS
3	466	4	0.0085	0.0104	-0.0018	-2.3391	0.97 %	AMIS
4	466	6	0.0127	0.0154	-0.0027	-3.0161	0.13 %	AMIS
5	466	7	0.0148	0.017	-0.0022	-2.1477	1.59 %	AMIS
6	466	17	0.0352	0.0406	-0.0054	-3.2494	0.06 %	AMIS
7	466	15	0.0312	0.0393	-0.0081	-4.7037	≈ 0.0 %	AMIS
8	466	41	0.0805	0.0879	-0.0074	-2.3744	0.88 %	AMIS
9	466	84	0.1486	0.1536	-0.0050	-1.322	9.31 %	?
10	466	176	0.2228	0.2593	-0.0365	-4.0893	≈ 0.0 %	AMIS

Table: Mean Squared Error Statistics for a percentile-based 10 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	0.1078	0.143	-0.0352	-28.3109	≈ 0.0 %	AMIS
1	466	0	0.0	0.0256	-0.0256	-25.6369	≈ 0.0 %	AMIS
2	466	0	0.0022	0.0245	-0.0223	-24.8491	≈ 0.0 %	AMIS
3	466	4	0.0144	0.0464	-0.0321	-17.3487	≈ 0.0 %	AMIS
4	466	6	0.0232	0.0597	-0.0365	-18.2062	≈ 0.0 %	AMIS
5	466	7	0.0292	0.0665	-0.0373	-17.6692	≈ 0.0 %	AMIS
6	466	17	0.0614	0.1199	-0.0585	-20.2257	≈ 0.0 %	AMIS
7	466	15	0.0691	0.129	-0.0599	-19.8579	≈ 0.0 %	AMIS
8	466	41	0.1582	0.2137	-0.0555	-13.3289	≈ 0.0 %	AMIS
9	466	84	0.2745	0.3077	-0.0332	-7.3182	≈ 0.0 %	AMIS
10	466	176	0.4445	0.4355	0.0090	1.0224	15.33 %	?

Table: Mean Absolute Error Statistics for a percentile-based 10 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	4661	351	-0.0042	0.0235	-0.0278	21.643	0.0 %	AMIS
1	466	0	0.0	0.0256	-0.0256	25.6369	0.0 %	AMIS
2	466	0	0.0022	0.0245	-0.0223	24.8491	0.0 %	AMIS
3	466	4	-0.0027	0.0305	-0.0332	18.4258	0.0 %	AMIS
4	466	6	-0.0022	0.0356	-0.0378	19.4281	0.0 %	AMIS
5	466	7	$-4.0 \cdot 10^{-4}$	0.0394	-0.0398	19.7389	0.0 %	AMIS
6	466	17	-0.0095	0.0542	-0.0638	24.1842	0.0 %	AMIS
7	466	15	0.0073	0.0722	-0.0649	23.3298	0.0 %	AMIS
8	466	41	-0.0026	0.068	-0.0706	19.4139	0.0 %	AMIS
9	466	84	-0.0316	0.0219	-0.0535	13.0617	0.0 %	AMIS
10	466	176	-0.0018	-0.1351	0.1332	-21.1407	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 10 bin discretisation based on the AMIS score. STEMI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0294	0.0348	-0.0055	-5.0712	≈ 0.0 %	AMIS
1	288	0	0.0	0.0012	-0.0012	-7.616	≈ 0.0 %	AMIS
2	288	0	≈ 0.0	0.0011	-0.0011	-8.0915	≈ 0.0 %	AMIS
3	288	1	0.0035	0.0055	-0.0020	-5.8202	≈ 0.0 %	AMIS
4	288	1	0.0034	0.0063	-0.0028	-6.1991	≈ 0.0 %	AMIS
5	288	3	0.0103	0.0138	-0.0035	-9.0597	≈ 0.0 %	AMIS
6	288	2	0.0070	0.0118	-0.0048	-3.6315	0.01 %	AMIS
7	288	6	0.0204	0.0272	-0.0068	-6.0306	≈ 0.0 %	AMIS
8	288	8	0.0274	0.0364	-0.0090	-4.9569	≈ 0.0 %	AMIS
9	288	16	0.0543	0.0662	-0.0119	-3.6054	0.02 %	AMIS
10	288	62	0.1661	0.1745	-0.0084	-0.8679	19.28 %	?

Table: Mean Squared Error Statistics for a percentile-based 10 bin discretisationbased on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.0678	0.0997	-0.0319	-22.6403	≈ 0.0 %	AMIS
1	288	0	0.0	0.0266	-0.0266	-19.6812	≈ 0.0 %	AMIS
2	288	0	$2.0 \cdot 10^{-4}$	0.0258	-0.0257	-20.6629	≈ 0.0 %	AMIS
3	288	1	0.0093	0.036	-0.0267	-13.8488	≈ 0.0 %	AMIS
4	288	1	0.0106	0.0417	-0.0311	-13.3046	≈ 0.0 %	AMIS
5	288	3	0.022	0.0581	-0.0361	-15.2875	≈ 0.0 %	AMIS
6	288	2	0.0234	0.0687	-0.0453	-15.3565	≈ 0.0 %	AMIS
7	288	6	0.0472	0.0987	-0.0515	-16.0956	≈ 0.0 %	AMIS
8	288	8	0.0674	0.1229	-0.0555	-14.9057	≈ 0.0 %	AMIS
9	288	16	0.1422	0.1896	-0.0474	-9.5136	≈ 0.0 %	AMIS
10	288	62	0.3535	0.3244	0.0291	2.8612	0.21 %	TIMI

Table: Mean Absolute Error Statistics for a percentile-based 10 bin discretisation based on the AMIS score. STEMI, PCI

Bin	n	$n_\dagger$	e^{AMIS}	e^{TIMI}	Δe	t_p	P	Winner
All	2882	101	0.015	0.0419	-0.0269	18.6617	0.0 %	AMIS
1	288	0	0.0	0.0266	-0.0266	19.6812	0.0 %	AMIS
2	288	0	$2.0 \cdot 10^{-4}$	0.0258	-0.0257	20.6629	0.0 %	AMIS
3	288	1	0.0024	0.0292	-0.0267	13.928	0.0 %	AMIS
4	288	1	0.0038	0.0351	-0.0313	13.4361	0.0 %	AMIS
5	288	3	0.0014	0.0378	-0.0364	15.479	0.0 %	AMIS
6	288	2	0.0097	0.0564	-0.0467	16.2419	0.0 %	AMIS
7	288	6	0.0067	0.0599	-0.0532	17.1735	0.0 %	AMIS
8	288	8	0.014	0.0729	-0.0589	16.6295	0.0 %	AMIS
9	288	16	0.0425	0.0978	-0.0553	11.7909	0.0 %	AMIS
10	288	62	0.0723	-0.0165	0.0887	-10.0083	≈ 0.0 %	TIMI

Table: Mean Error Statistics for a percentile-based 10 bin discretisation based on the AMIS score. STEMI, PCI

References

[Hunt, 2006] Hunt, Katrin: Evaluation of novel algorithms to optimize Risk Stratification Scores in Myocardial Infarction; Master thesis at the Institute for Informatics, University of Zurich, 2006.

[Morrow et al, 2000] 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.

[Witten/Frank, 2005] Witten, Ian H.; Eibe, Frank: Data Mining. Practical Machine Learning Tools and Techniques. 2nd Edition. Morgan Kaufmann, 2005.

[Aktinson, 1969] Atkinson, Anthony C.: A Test for Discriminating between Models. Biometrika, Vol. 56, No. 2. 1969.

[Prince, 1982] Prince, E.: Comparison of the Fits of Two Models to the Same Data Set. National Measurement Laboratory, National Bureau of Standards, Washington, DC. In: Acta Cystallographica, B38, 1982.

[Himmelblau, 1970] Himmelblau, D. M.: Process Analysis by Statistical Methods, pp. 216-221. New York: John Wiley, 1970.

[Kurz et al, 2006] Kurz, David J.; Hunt, Katrin; Bernstein, Abraham; Radocanovic, Dragana; Erne, Paul E.; Stauffer, Jean-Christophe; Bertel, Osmund: Inadequate performance on the TIMI risk prediction score for patients with ST-elevation myocardial infarction in the modern era. Working Paper.

[Domingos/Pazzani, 1996] Domingos, Pedro; Pazzani, Michael: Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier. Proceedings of the 13th International Conference on Machine Learning, Bari/Italy, 1996.

[Zhang et al, 2005] Zhang, Harry; Liangxiao, Jiang; Jiang, Su: Augmenting Naive Bayes for Ranking. Proceedings of the 22nd International Conference on Machine Learning, Bonn/Germany, 2005.

[Zadrozny/Elkan, 2001] Zadrozny, Bianca; Elkan, Charles: Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. Department of Computer Science and Engineering, University of California, San Diego, La Jolla, California. 92093-0114.

[Vinterbo/Ohne, 1999] Vinterbo, Staal; Ohne-Machade, Lucila: A Recalibration Method for Predictive Models with Dichotomous Outcomes. In Predictive Models in Medicine: Some Methods for Construction and Adaptation. Norwegian University of Science and Technology, 1999.

[Fawcett, 2004] Fawcett, Tom: ROC Graphs: Notes and Practical Considerations for Researchers. HP Laboratories, 2004. Published by Kluver Academic Publishers, 2004.

[Provost/Domingos, 2003] Provost, F.; Domingos, P.: Tree induction for probability-based ranking. Machine Learning, 52 (3), 199-203, September 2003.

[Cestnik, 1990] Cestnik, B.: Estimating Probabilities: A crucial task in machine learning. Proceedings of the Ninth European Conference on Artificial Intelligence (pp. 147-149). Pitman.

[Murphy, 1973] Murphy, A. H.: A new vector partition of the probability score. Journal of Applied Meteorology, 12:595-600, 1973.

[Hosmer/Lemeshow, 1989] Hosmer, David W.; Lemeshow, Stanley: Applied Logistic Regression. Wiley Series in Probability and Mathematical Statistics, 1989.