

RESEARCH ARTICLE

# Accuracy and Calibration of Computational Approaches for Inpatient Mortality Predictive Modeling

Christos T. Nakas<sup>1,2</sup>, Narayan Schütz<sup>1</sup>, Marcus Werners<sup>3</sup>, Alexander B. Leichtle<sup>1\*</sup>

**1** University Institute of Clinical Chemistry, Centre of Laboratory Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, **2** Laboratory of Biometry, University of Thessaly, Volos, Greece, **3** Central Controlling Unit, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

\* [alexander.leichtle@insel.ch](mailto:alexander.leichtle@insel.ch)



OPEN ACCESS

**Citation:** Nakas CT, Schütz N, Werners M, Leichtle AB (2016) Accuracy and Calibration of Computational Approaches for Inpatient Mortality Predictive Modeling. PLoS ONE 11(7): e0159046. doi:10.1371/journal.pone.0159046

**Editor:** Takeru Abe, Yokohama City University, JAPAN

**Received:** January 8, 2016

**Accepted:** June 7, 2016

**Published:** July 14, 2016

**Copyright:** © 2016 Nakas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are restricted from public sharing by Cantonal Ethics Committee of Bern and the Inselspital Bern. Data may be retrieved from the authors upon ethical approval. Please contact Cantonal ethics committee Bern: <https://www.kek-bern.ch/kontakt.html>, or Scientific secretary: Dr. sc. nat. Dorothy Pfiffner, e-mail: [pfiffner@kek.unibe.ch](mailto:pfiffner@kek.unibe.ch). For the Inselspital, the interested researchers might refer to the corresponding author, who can coordinate the request and support the retrieval of the respective waivers: [alexander.leichtle@insel.ch](mailto:alexander.leichtle@insel.ch).

**Funding:** The authors have no support or funding to report.

## Abstract

Electronic Health Record (EHR) data can be a key resource for decision-making support in clinical practice in the “big data” era. The complete database from early 2012 to late 2015 involving hospital admissions to Inselspital Bern, the largest Swiss University Hospital, was used in this study, involving over 100,000 admissions. Age, sex, and initial laboratory test results were the features/variables of interest for each admission, the outcome being inpatient mortality. Computational decision support systems were utilized for the calculation of the risk of inpatient mortality. We assessed the recently proposed Acute Laboratory Risk of Mortality Score (ALaRMS) model, and further built generalized linear models, generalized estimating equations, artificial neural networks, and decision tree systems for the predictive modeling of the risk of inpatient mortality. The Area Under the ROC Curve (AUC) for ALaRMS marginally corresponded to the anticipated accuracy (AUC = 0.858). Penalized logistic regression methodology provided a better result (AUC = 0.872). Decision tree and neural network-based methodology provided even higher predictive performance (up to AUC = 0.912 and 0.906, respectively). Additionally, decision tree-based methods can efficiently handle Electronic Health Record (EHR) data that have a significant amount of missing records (in up to >50% of the studied features) eliminating the need for imputation in order to have complete data. In conclusion, we show that statistical learning methodology can provide superior predictive performance in comparison to existing methods and can also be production ready. Statistical modeling procedures provided unbiased, well-calibrated models that can be efficient decision support tools for predicting inpatient mortality and assigning preventive measures.

## Introduction

The use of Electronic Health Records (EHR) for building mortality predictive models is a modern practice that is expected to enhance patient care by pointing physicians to patients at risk, that would potentially be missed in clinical routine [1]. The vast amount of EHR and

**Competing Interests:** The authors have declared that no competing interests exist.

laboratory databases allow for the construction of scoring models that quantify patients' risks for event or mortality prediction and strongly suggest the use of emerging "big data" strategies for analyzing these frequently incomplete, unordered, and usually for documentation purposes only, collected data [2–4].

Models such as the Acute Laboratory Risk of Mortality (ALaRMS) score [5,6] have been proposed in the literature to this goal, as the rapid assessment of clinical severity using EHR data available at the time of admission may aid decision support and improve healthcare quality. For an elderly cohort in Israel, Smolin et al. [7] applied multivariate logistic regression analysis to predict the risk of 6-month mortality from laboratory and clinical anamnesis data, whereas Lee et al. [8] proposed a method for personalized mortality prediction based on electronic health record (EHR) data and patient similarity metrics. Scoring systems such as the Acute Physiology and Chronic Health Evaluation (APACHE IV), the Simplified Acute Physiology Score (SAPS), the Laboratory-based Acute Physiology Score (LAPS), the Comorbidity Point Score (COPS), and others are typically used in clinical practice and may aid to the decision making process offering a simplified, swift patient assessment [2, 3]. However, such scoring systems almost always offer simplistic quantification of patients' risks.

In this work, we reproduced the method proposed by Tabak et al. [5], namely the Acute Laboratory Risk of Mortality Score (ALaRMS) as a modern representative of scoring systems, built our models using statistical learning methodology, and compared all of them. We show that models that are constructed based on statistical learning methodology are well-calibrated and offer superior diagnostic accuracy to previous or traditional regression approaches.

As a result, statistical learning methodologies may provide more accurate predictions and enable the use of incomplete data, not only for retrospective studies but potentially also for real-time application.

The following section describes the database on which we based our assessment, develops on the methods we used, and covers the implementation procedures. Results are presented next along with intuition about these. We end with a discussion and conclusions about proper strategies for predicting inpatient mortality based on relevant retrospective outcome studies.

## Methods

In this section, we present the experiment design which simply involves the extraction of the database and methodology used for modeling purposes.

### Database

The complete database of hospital admissions to the Inselspital from early 2012 to late 2015 was used, involving over 100,000 admissions. The Inselspital provides as the University Hospital of Bern quaternary medical care mainly for the canton of Bern (about one million inhabitants) with 78 departments, about 900 beds, and more than 6000 employees (17% physicians, 38% care-givers). Age, sex, and initial laboratory test results were included in the database, the outcome being discharge disposition, which identifies inpatient mortality status. A total of 23 numeric laboratory test results were included for reasons of consistency with the construction of the ALaRMS model. Those were serum chemistry (specifically: albumin, aspartate transaminase, alkaline phosphatase, blood urea nitrogen (BUN), calcium, creatinine, glucose, potassium (K), sodium (Na), and total bilirubin); hematology and coagulation parameters (bands, hemoglobin, partial thromboplastin time, prothrombin time international normalized ratio (PT INR), platelets, and white blood cell count (WBC)); arterial blood gas (partial pressure of carbon dioxide (pCO<sub>2</sub>), partial pressure of oxygen (pO<sub>2</sub>), and pH value); cardiac markers (brain natriuretic peptide (BNP) or NT-proBNP, creatine phosphokinase MB (CK MB), and

troponin T (to replace troponin I in the score). An ethics dispensation from the cantonal ethics committee Bern (NoZ023/2014) was issued for the anonymized use of these data.

Data handling involved filtering and pre-processing. Entries with missing values for discharge disposition and patient/admission ID were eliminated. A single duplicate entry was found and verified which was also eliminated from the database. The final database included 106,688 admissions. Crude inpatient mortality risk was calculated to be 2.41% (corresponding to 2,568 admissions). Multiple imputation was used as a second step in order to handle missing data since not all patients had results for the 23 laboratory tests that were considered. We have then employed modeling approaches both for the database with the missing and for the one with the imputed data.

## Statistical analysis methods and implementation

Initial data analysis involved the calculation of basic descriptive statistics based on the available data as a first step. Next, the ALaRMS score was calculated for each admission along the lines described in Tabak et al. [5]. The ALaRMS system was chosen as representative of scoring systems being a most recent addition to the relevant bibliography [5]. The following step was to build statistical models that can be used as a prognostic tool for inpatient mortality combining age, sex, and lab measurements per se, without having to resort to a scoring system. A random 80% of the admissions were used as training sample while the rest were used as the testing sample.

Linear classification methodology included generalized estimating equation (GEE) techniques and Generalized Linear Modeling using LASSO (GLM). Regarding the GEE model, admission ID defined the subject-level and age, sex, and the different laboratory test measurements were introduced as explanatory variables. GEEs can handle missing values effectively and were thus used for the non-imputed database. GEE parameter estimates are consistent even when the covariance structure is misspecified [9]. We have used an independent structure for the working correlation matrix, which reduced significantly computational burden.

GLM was considered for an imputed dataset. Specifically, we have used penalized multiple logistic regression with LASSO penalization, a choice that can be equivalent to Bayesian Model Averaging [10]. The tuning parameter lambda was chosen via cross-validation [11].

Neural artificial network and decision tree based algorithms were adjusted to be directly usable on raw data. For the neural network we employed a Multi-Layer Perceptron (MLP) with 3 hidden layers, rectified linear units (as activation function for the hidden units) and dropout (for better generalization).

For the decision tree approach we used a rule-based model via the C5.0 algorithm [12]. This algorithm was chosen as it is able to handle missing data relatively well, and besides computational performance gains, also allows for boosting out of the box, which generally results in better predictive performance. In the final decision rule model, an ensemble of 11 models was used for prediction. Soft-classification was used for both the MLP and C5 model, in order to derive probabilistic predictions for class assignment. For calibration a simple logistic regression was put on top of both models to calculate the respective class posterior probabilities. ROC curve analysis was used for the assessment of the accuracy of markers and scores that were calculated.

The R software [13] was used for data pre-processing (packages used: 'dplyr' [14], 'plyr' [15], 'tidyr' [16], 'doBy' [17]) and data analysis (packages used: 'lme4' [18] for the GEE, 'pROC' [19] for ROC analysis, 'glmnet' [20] for binary multiple logistic regression with LASSO penalization, 'C50' [21] for the application of the C5.0 algorithm for the rule-based model, 'RSNNS' [22] for the MLP). Multiple imputation was applied using the 'mi' package [23].

Additionally, Stata 13.1 (Stata Corp., College Station, TX) was used for the calculation of descriptive statistics, linear regression modeling, generalized estimating equation modeling verification, and figure art.

### Results

For the admissions to Inselspital Bern recorded between 2012 and 2015, a complete database was built including information such as age, sex, and initial laboratory test results along with the outcome, discharged ‘alive’ or ‘dead’. After data cleaning 106,688 records remained for analysis. There were 48,497 women (45.5%) and 58,191 men (54.5%) with 1,030 deaths among women (2.1%) and 1,538 deaths among men (2.6%). The average age was 52.03 ( $\pm 24.66$ ), being 51.69 ( $\pm 24.68$ ) for those discharged alive and 66.16 ( $\pm 19.42$ ) for those discharged dead.

The average number of initial laboratory tests administered was 10.58 ( $\pm 4.48$ ). Among those who died the average number of administered tests were 15.11 ( $\pm 4.03$ ), while they were equal to 10.46 ( $\pm 4.43$ ) among those discharged alive. The average ALaRMS score was 16.98 ( $\pm 18.50$ ), being 16.21 ( $\pm 17.59$ ) for those discharged alive and 48.50 ( $\pm 25.38$ ) for those discharged dead.

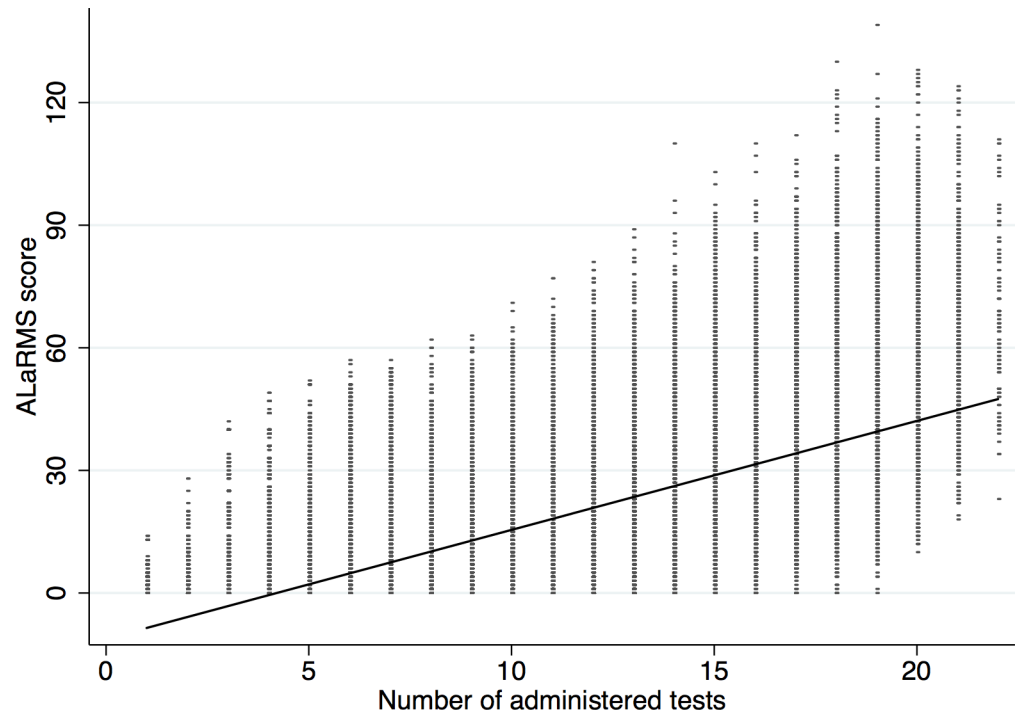
[Table 1](#) shows descriptive statistics for laboratory test results for all 106,688 admissions.

The ALaRMS score has been proposed as a tool for the prediction of death based on age, sex, and initial laboratory test results [5]. Our results regarding the accuracy of ALaRMS resemble those reported in [5,24,25]. Specifically, the area under the ROC curve was estimated to be equal to 0.858 (95% CI: 0.851, 0.865) comparable to 0.87 in the publication introducing this [5]. However, our findings demonstrate a clear linear trend between the ALaRMS score and the number of tests ordered by the physicians [26] (cf. [Fig 1](#)). We eliminated this trend

**Table 1. Laboratory test results descriptive statistics for all admissions.**

Test	Mean	SD	Median	min	max
ASAT (U/L)	61.67	345.9	26	4	26442
Albumin (g/L)	31.16	7.03	32	4	56
BNP (pg/mL)	631.35	898.57	260	10	5000
(NT-pro)-BNP (pg/mL)	5196.21	13385.65	1484	5	299792
CK-MB ( $\mu$ g/L)	8.98	30.63	2	0.3	396.7
Calcium (mmol/L)	2.23	0.2	2	0.5	15.47
Creatinin ( $\mu$ mol/L)	89.8	82.21	73	3	3334
Glucose (mmol/L)	6.67	3.12	5.9	0.2	101
Urea (mmol/L)	7.82	6.23	6	0.8	98.4
Hemoglobin (g/L)	127.65	20.46	130	1	238
INR	1.18	0.52	1.02	1	11
Potassium (mmol/L)	4.08	0.6	4	1.5	30.9
Leukocytes (G/L)	9.02	7.03	8	0	575
Sodium (mmol/L)	138.7	3.8	139	65	182
Band neutr. (%)	16.9	14.21	13.5	0	91.5
Thrombocytes (G/L)	238.35	97.31	225	1	2848
Troponin T ( $\mu$ g/L)	0.28	1.43	0.021	0.003	84.4
aPTT (s)	40.53	36.8	31.5	0	300
alk. Phosphatase (U/L)	106.56	130.52	74	10	8156
pCO <sub>2</sub> (mmHg)	40.7	9.99	39	7	172
pH	7.38	0.089	7.39	5.85	7.79
pO <sub>2</sub> (mmHg)	98.62	91.95	66	0	3678

doi:10.1371/journal.pone.0159046.t001



**Fig 1. Linear trend of ALaRMS score vs. number of administered tests ordered by the physician in charge.**

doi:10.1371/journal.pone.0159046.g001

using the ratio of the ALaRMS score over the number of administered tests as a prognostic index for death. As expected, the resulting AUC for the ratio had lower accuracy, being equal to 0.819 (95%CI: 0.813,0.826) and comparable to the accuracy of a model that simply combines the number of administered tests plus age and sex (model shown in Table 2). Simply using the number of administered tests to predict death results in AUC equal to 0.786 (95% CI: 0.777, 0.795), while adding age and sex as predictors to a binary logistic regression model yields AUC equal to 0.801 (95% CI: 0.792, 0.809).

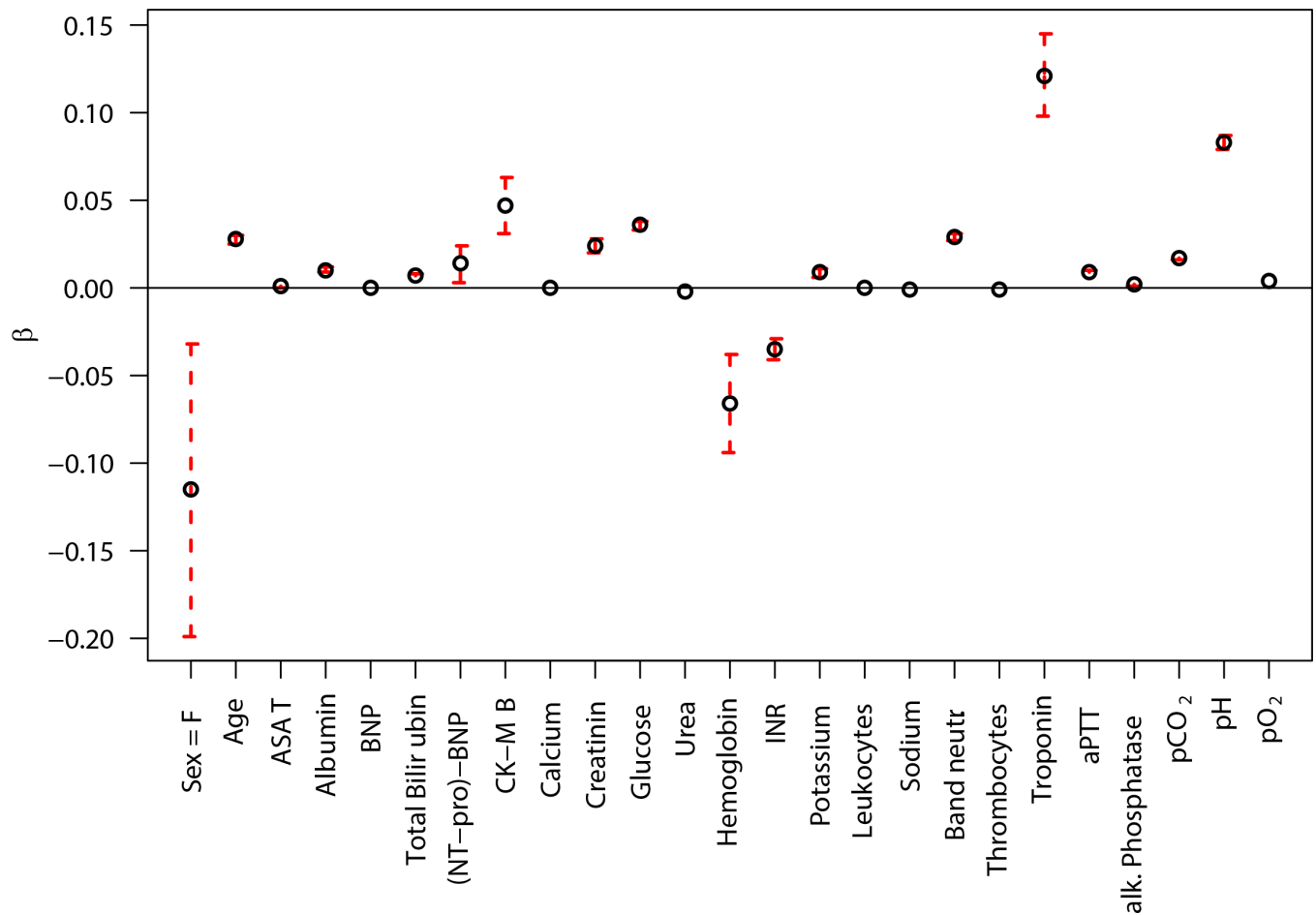
Furthermore, considering the fact that the prior probability of death is quite low (2.41%), the positive predictive value (PPV) of the ALaRMS model is significantly weakened. For example, the pair of sensitivity and specificity corresponding to the Youden index-based cut-off point for the ALaRMS model was estimated to be 81.1% and 74.4% respectively. These result in a PPV = 7.26% and a high negative predictive value (NPV) equal to 99.38%. In order to achieve sensitivity in the vicinity of 80% one needs to select a cut-off point of 25 for the ALaRMS score. A cut-off point equal to 73 would result in a much higher PPV (i.e. PPV = 30.53% with corresponding NPV = 98%), but in this case the sensitivity of ALaRMS drops to 18%.

The resulting AUC for the GEE for the testing sample was 0.809 (95% CI: 0.801, 0.817). A graphical representation of the coefficients of the GEE model is shown in Fig 2. Although the

**Table 2. A simplistic alternative to the ALaRMS score with comparable accuracy to the adjusted-for-number of administered tests ALaRMS score.**

Parameter	B	Std. Error	Sig.	OR	95% CI for OR	
(Intercept)	-7.941	0.1093	<0.01	-		
Sex = F	-0.034	0.0416	0.415	0.967	0.891	1.049
Number of administered tests	0.232	0.0054	<0.01	1.262	1.248	1.275
Age	0.021	0.0012	<0.01	1.021	1.018	1.023

doi:10.1371/journal.pone.0159046.t002



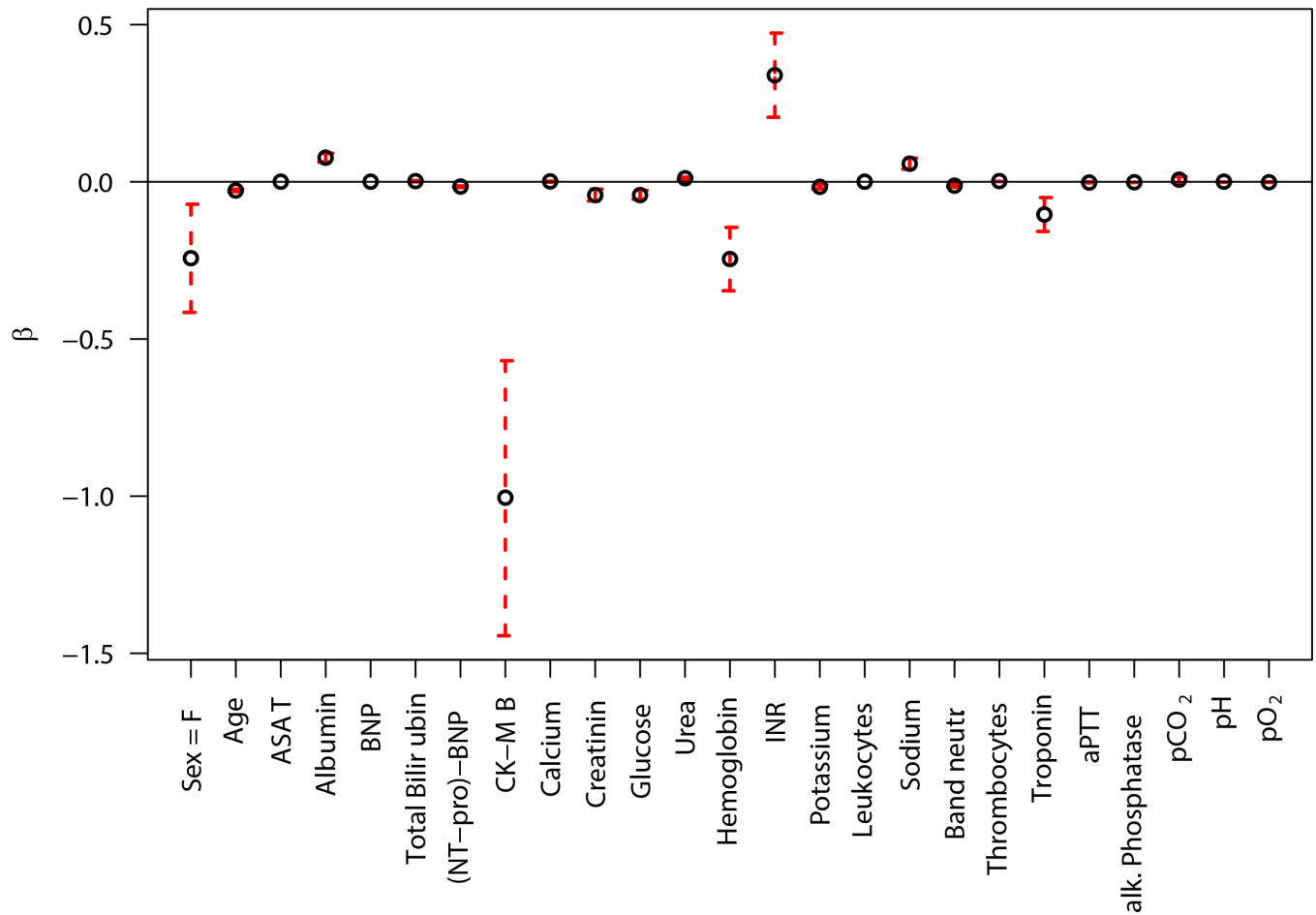
**Fig 2. Absolute values of model parameters, along with 95% confidence intervals, for the GEE model.**

doi:10.1371/journal.pone.0159046.g002

importance of specific laboratory measurements on the outcome is obvious and may provide clinical insight, this model does not provide an adequate accuracy in terms of ROC AUC. This model was applied to the non-imputed dataset, as it can easily handle missing data. However, application of the GEE to the imputed dataset resulted in a very slight increase of the ROC AUC (0.814).

Model parameters after imputation and restricted multiple logistic regression with LASSO penalization, which yielded an AUC equal to 0.872 (95% CI: 0.859, 0.885) for the testing sample, are presented in Fig 3.

Similar results, in terms of accuracy, were obtained for the rule-based random forest model (C5.0) and MLP. Specifically, C5.0 using soft classification, without feature selection on the imputed data resulted in AUC equal to 0.870 (95% CI: 0.862, 0.878) for the testing sample. The model using feature selection resulted in ROC AUC of 0.847 (95% CI: 0.834, 0.857). Attribute importance of the C5.0 model is shown in Table 3, the more important a variable in the model is, the more important it is for making decisions regarding class assignment, thus the higher its discriminative power is. However, given that these classification models are not naturally probabilistic, they may produce distorted class probability distributions [27]. Calibration of these models was achieved using standard logistic regression after initial soft-classification, with this methodology we can directly calculate the respective class posterior probabilities which results in well calibrated models. The fit and model calibration of the ALARMS, GEE, and GLM relative to patients' age are shown in Figs 4 and 5.



**Fig 3. Absolute values of model parameters, along with 95% confidence intervals, for the GLM.**

doi:10.1371/journal.pone.0159046.g003

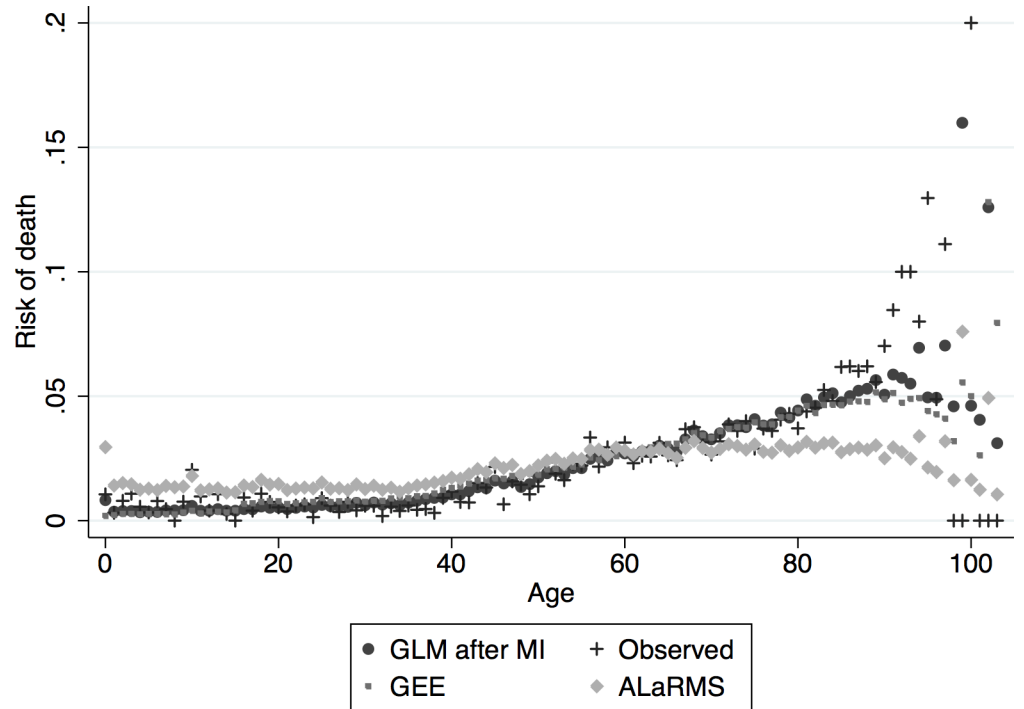
### Discussion

We demonstrated that statistical learning methodology could provide superior predictive performance in comparison to existing methods given that estimated ROC AUCs of the models

**Table 3. Attribute importance (average of 11 decision trees; C5.0 algorithms used splits) for the imputed dataset.**

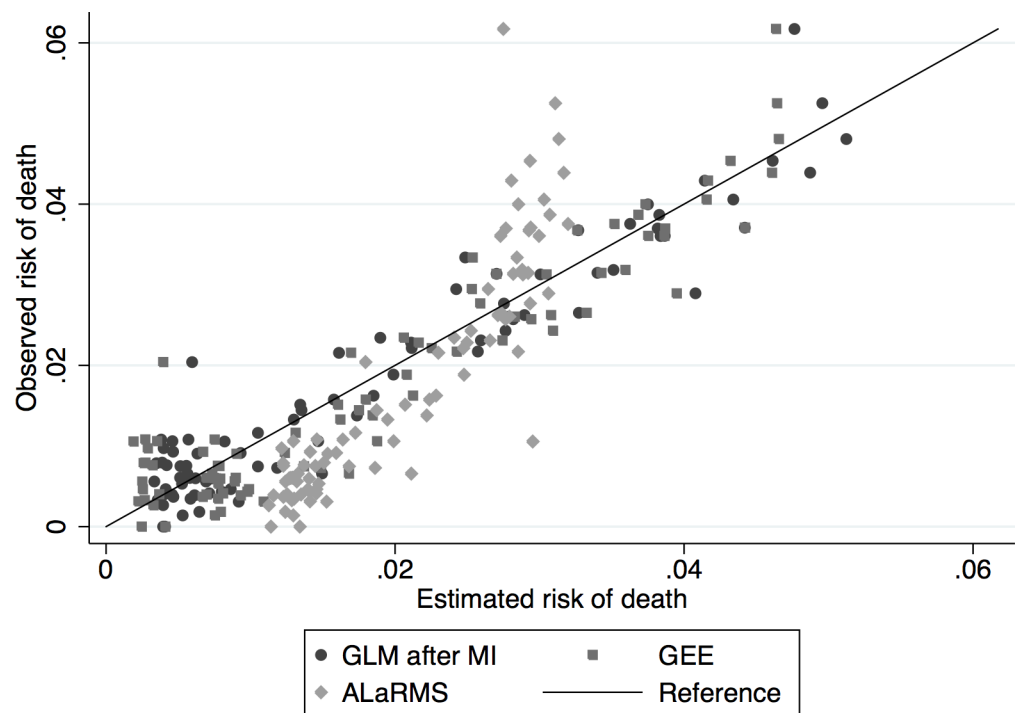
100.00%	Band neutr.		
100.00%	pH	98.12%	CK-MB
100.00%	Calcium	97.96%	BNP
99.94%	ASAT	97.67%	(NT-pro) BNP
99.92%	Hemoglobin	96.69%	Thrombocytes
99.90%	pCO <sub>2</sub>	96.54%	Albumin
99.59%	Age	95.20%	Creatinin
99.28%	pO <sub>2</sub>	95.18%	Glucose
99.15%	Urea	86.10%	Sodium
99.02%	Total Bilirubin	67.07%	INR
98.49%	Leukocytes	44.73%	Sex
98.35%	alk. Phosphatase	32.17%	Troponin T
98.24%	aPTT	0.42%	Potassium

doi:10.1371/journal.pone.0159046.t003



**Fig 4. Model calibration according to 'Age'.** Observed vs. Expected risk of death for GEE, GLM, and ALaRMS.

doi:10.1371/journal.pone.0159046.g004



**Fig 5. Model calibration according to 'Age'.** Expected vs. Observed risk of death for GEE, GLM, ALaRMS.

doi:10.1371/journal.pone.0159046.g005



that were built provide evidence of better prognostic accuracy. The ALaRMS score that has been recently proposed as an indicator signaling inpatient mortality for a hospital admission would in our case constitute a biased assessment largely influenced by the subjective opinion of the physician ordering a batch of lab tests for the admitted patient [28] as, by its methodological construction, it takes on larger values as the number of administered tests increases. The ALaRMS score also suffers from poor calibration based on our findings. Counting the number of tests requested for a certain patient and combining it with age and sex yields an AUC of 0.801 (compared to 0.858 of the ALaRMS score). Since ALaRMS is an additive score of points assigned to a limited number of lab result ranges, it only roughly covers continuous effects and does not account for e.g. mutual information. However, innovative predictive modeling strategies provide sufficient accuracy in incomplete, biased data sets like EHR and laboratory data [2–4].

Imputation allows building a model as if no biased assessment existed. Our results show that this works pretty well in practice. A possible drawback is the large amount of missing values and that the GLM results in high accuracy when we actually use imputation. However, using modern algorithms for boosting the performance of rule-based methods, one can produce models with high accuracy without any imputation, since we have achieved AUCs of over 0.90 (specifically, up to AUC = 0.912 and 0.906, for the C5 and MLP respectively) by properly adjusting the rule-based algorithms parameters (feature weighing) and sampling schemes (subject selection).

Although lab requests usually should follow uniform diagnostic paths, their patterns are frequently highly variable—between disciplines, between clinics, and between physicians. Open “menu” request systems without digital expert systems or administrative restrictions facilitate the selection of favorite sets of “biomarkers”, that are neither mirrored by the actual guidelines nor by computational evidence [29], and even if published or hospital-based recommendations exist, they are scarcely followed [30]. The high degree of collinearity present in many routinely measured lab results not only points at tests potentially not additionally informative, but also blurs the contribution of a single parameter to a certain prediction or differentiation. Statistical learning approaches such as in our case the MLP and C5.0 algorithm can incorporate not only collinearities but also mutual information between individual variables.

Albumin e.g. as a negative acute-phase protein is inversely correlated to markers of inflammation (e.g. leukocyte count), and CK-MB and Troponin as markers of myocardial damage are frequently jointly elevated. The actual “diagnostic value” of a given lab test therefore depends not only from its actual level, but also from its shared variance with other markers and also the pre-test probability in the respective patient population. The assessment of a lab test’s predictivity for a certain end point or differential diagnosis therefore remains a medical as well as computational challenge with great implications on patient care and healthcare costs.

Comparison of the GEE model coefficients in Fig 2 with the GLM coefficients in Fig 3 suggests that low hemoglobin is significant for the GEE model (typically associated with e.g. kidney problems, GI bleeding, and trauma), whereas high hemoglobin values are important in the GLM model (associated especially with bone marrow disorders). Although smaller in magnitude, there appears to be also a reversal in estimated influence of albumin. This suggests that separate modeling of elevated versus depressed levels for some markers may improve model performance. We can also assume that within the whole “model space” there are many opposing and inverted, but likewise predictive marker combinations. A second observation in looking at the model fits is the large coefficient for Troponin in the GEE model with smaller importance of CK-MB, with the GLM model having a very large coefficient for CK-MB and small for Troponin. Given that data is missing on occasion, a model approach that puts similar weights on highly correlated factors could improve model predictiveness for individuals where

one or the other measurement is missing. Feature/marker weighing and subject selection are automatically accommodated with the C5 and MLP approaches but not with the current GLM and GEE ones. Future work includes further refinement of all of our models and production of a ready-to-use system for everyday clinical practice.

## Conclusion

Employing computation intensive methods for decision support could help save patients' lives as they may help physicians to better assess a patient's risk. Efficient strategies that link machine learning methodology to clinical decision making are described in this article and could enhance diagnostic accuracy and patient assessment.

Future research also includes expanding such models to more types of data, for example, medication, clinical history, proteomics/metabolomics data, microarray data or even full genome scans.

## Acknowledgments

The authors would like to thank Heinz von Allmen for his outstanding support querying the laboratory data. The authors would also thank the reviewers and likewise the editor for their valuable comments.

## Author Contributions

Conceived and designed the experiments: CTN ABL. Performed the experiments: NS CTN. Analyzed the data: CTN NS. Contributed reagents/materials/analysis tools: MW. Wrote the paper: CTN NS MW ABL. Extracted, transformed, and provided the data: MW.

## References

1. Toga AW, Foster I, Kesselman C, Madduri R, Chard K, Deutsch EW, et al. Big biomedical data as the key resource for discovery science. *J Am Med Inform Assoc. The Oxford University Press*; 2015; 22: 1126–1131. doi: [10.1093/jamia/ocv077](https://doi.org/10.1093/jamia/ocv077) PMID: [26198305](https://pubmed.ncbi.nlm.nih.gov/26198305/)
2. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood). Project HOPE—The People-to-People Health Foundation, Inc*; 2014; 33: 1123–1131. doi: [10.1377/hlthaff.2014.0041](https://doi.org/10.1377/hlthaff.2014.0041)
3. Roski J, Bo-Linn GW, Andrews TA. Creating Value In Health Care Through Big Data: Opportunities And Policy Implications. *Health Affairs. Project HOPE—The People-to-People Health Foundation, Inc*; 2014; 33: 1115–1122. doi: [10.1377/hlthaff.2014.0147](https://doi.org/10.1377/hlthaff.2014.0147) PMID: [25006136](https://pubmed.ncbi.nlm.nih.gov/25006136/)
4. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst. BioMed Central Ltd*; 2014; 2: 3. doi: [10.1186/2047-2501-2-3](https://doi.org/10.1186/2047-2501-2-3) PMID: [25825667](https://pubmed.ncbi.nlm.nih.gov/25825667/)
5. Tabak YP, Sun X, Nunez CM, Johannes RS. Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS). *J Am Med Inform Assoc. BMJ Publishing Group Ltd*; 2014; 21: 455–463. doi: [10.1136/amiajnl-2013-001790](https://doi.org/10.1136/amiajnl-2013-001790) PMID: [24097807](https://pubmed.ncbi.nlm.nih.gov/24097807/)
6. Froom P, Shimoni Z. Prediction of hospital mortality rates by admission laboratory tests. *Clinical Chemistry. American Association for Clinical Chemistry*; 2006; 52: 325–328. doi: [10.1373/clinchem.2005.059030](https://doi.org/10.1373/clinchem.2005.059030) PMID: [16449218](https://pubmed.ncbi.nlm.nih.gov/16449218/)
7. Smolin B, Levy Y, Sabbach-Cohen E, Levi L, Mashiach T. Predicting mortality of elderly patients acutely admitted to the Department of Internal Medicine. *Int J Clin Pract.* 2015; 69: 501–508. doi: [10.1111/ijcp.12564](https://doi.org/10.1111/ijcp.12564) PMID: [25311361](https://pubmed.ncbi.nlm.nih.gov/25311361/)
8. Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. Emmert-Streib F, editor. *PLoS ONE. Public Library of Science*; 2015; 10: e0127428. doi: [10.1371/journal.pone.0127428](https://doi.org/10.1371/journal.pone.0127428) PMID: [25978419](https://pubmed.ncbi.nlm.nih.gov/25978419/)
9. Hardin JW, Hilbe JM. *Generalized Estimating Equations, Second Edition.* CRC Press; 2012.
10. Yuan M, Lin Y. Efficient Empirical Bayes Variable Selection and Estimation in Linear Models. *Journal of the American Statistical Association.* 2005; 100: 1215–1225. doi: [10.1198/016214505000000367](https://doi.org/10.1198/016214505000000367)

11. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer Science & Business Media; 2013. doi: [10.1007/978-1-4614-7138-7](https://doi.org/10.1007/978-1-4614-7138-7)
12. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York, NY: Springer Science & Business Media; 2013. doi: [10.1007/978-1-4614-6849-3](https://doi.org/10.1007/978-1-4614-6849-3)
13. The R Core Team. R [Internet]. 3rd ed. Vienna. Available: <https://www.R-project.org/>
14. Wickham H, Francois R. *A Grammar of Data Manipulation*. 0 ed. 2015. Available: <http://CRAN.R-project.org/package=dplyr>
15. Wickham H. plyr. *Journal of Statistical Software*. 2011; 40: 1–29.
16. Wickham H. Easily Tidy Data with 'spread()' and 'gather()' Functions. 0 ed. 2015. Available: <http://CRAN.R-project.org/package=tidyr>
17. Højsgaard S, Halekoh U. doBy [Internet]. 4 ed. Available: <http://CRAN.R-project.org/package=doBy>
18. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Soft*. 2015; 67: 1–48. doi: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)
19. Robin X, Centre GS, Turck N, Hainard A, Tiberti N, Lisacek F, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. BioMed Central Ltd; 2011. pp. 77–8. doi: [10.1186/1471-2105-12-77](https://doi.org/10.1186/1471-2105-12-77) PMID: [21414208](https://pubmed.ncbi.nlm.nih.gov/21414208/)
20. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Soft*. NIH Public Access; 2010; 33: 1–968. doi: [10.1109/TPAMI.2005.127](https://doi.org/10.1109/TPAMI.2005.127)
21. Kuhn M, Weston S, Coulter N, Culp M. C50: C5.0 Decision Trees and Rule-Based Models [Internet]. 0 ed. Available: <http://CRAN.R-project.org/package=C50>
22. Bergmeir C, Benítez JM. CORE: Connecting Repositories. *J Stat Soft*. 2012.
23. Su Y-S, Yajima M, Gelman A, Hill J. Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *Journal of Statistical Software*. 2011; 45: 1–31.
24. Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. *Med Care*. 2007; 45: 789–805. doi: [10.1097/MLR.0b013e31803d3b41](https://doi.org/10.1097/MLR.0b013e31803d3b41) PMID: [17667314](https://pubmed.ncbi.nlm.nih.gov/17667314/)
25. Novack V, Pencina M, Zahger D, Fuchs L, Nevzorov R, Jotkowitz A, et al. Routine laboratory results and thirty day and one-year mortality risk following hospitalization with acute decompensated heart failure. Hernandez AV, editor. *PLoS ONE*. Public Library of Science; 2010; 5: e12184. doi: [10.1371/journal.pone.0012184](https://doi.org/10.1371/journal.pone.0012184) PMID: [20808904](https://pubmed.ncbi.nlm.nih.gov/20808904/)
26. Boef AGC, Dekkers OM, le Cessie S, Vandenbroucke JP. Reporting Instrumental Variable Analyses. *Epidemiology*. 2013; 24: 937–938. doi: [10.1097/01.ede.0000434433.14388.a1](https://doi.org/10.1097/01.ede.0000434433.14388.a1) PMID: [24076999](https://pubmed.ncbi.nlm.nih.gov/24076999/)
27. Zadrozny B, Elkan C. Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. *ICML*. 2001.
28. Hardison JE. To Be Complete. *N Engl J Med*. Massachusetts Medical Society; 1979; 300: 193–194. doi: [10.1056/NEJM197901253000410](https://doi.org/10.1056/NEJM197901253000410) PMID: [759846](https://pubmed.ncbi.nlm.nih.gov/759846/)
29. Leichtle A. Biomarker—vom Sein und Wesen. *J Lab Med*. 2015; 39: 97–101. doi: [10.1515/labmed-2015-0033](https://doi.org/10.1515/labmed-2015-0033)
30. Teich N, Mössner J, Keim V. How effective is published medical education? *Lancet*. Elsevier; 2004; 363: 1326. doi: [10.1016/S0140-6736\(04\)16015-7](https://doi.org/10.1016/S0140-6736(04)16015-7) PMID: [15094287](https://pubmed.ncbi.nlm.nih.gov/15094287/)