PERGAMON

Computers in Biology and Medicine 31 (2001) 1–13

Computers in Biology and Medicine

www.elsevier.com/locate/compbiomed

Using patient-reportable clinical history factors to predict myocardial infarction

Samuel J. Wang[a,*], Lucila Ohno-Machado[b], Hamish S. F. Fraser[c], R. Lee Kennedy[d]

[a] *Clinical Information Systems Research & Development, Partners HealthCare System, Boston, MA, USA*
[b] *Decision Systems Group, Brigham and Women's Hospital, Boston, MA, USA*
[c] *Clinical Decision Making Group, Massachusetts Institute of Technology, Boston, MA, USA*
[d] *Department of Medicine, City Hospitals Sunderland & University of Sunderland, Sunderland, UK*

Received 3 February 2000; accepted 17 July 2000

Abstract

Using a derivation data set of 1253 patients, we built several logistic regression and neural network models to estimate the likelihood of myocardial infarction based upon patient-reportable clinical history factors only. The best performing logistic regression model and neural network model had *C*-indices of 0.8444 and 0.8503, respectively, when validated on an independent data set of 500 patients. We conclude that both logistic regression and neural network models can be built that successfully predict the probability of myocardial infarction based on patient-reportable history factors alone. These models could have important utility in applications outside of a hospital setting when objective diagnostic test information is not yet be available.

Keywords: Decision support techniques; Artificial intelligence; Regression analysis; Logistic models; Neural networks (computer); Myocardial infarction

1. Background

In the United States, approximately 1.5 million myocardial infarctions (MI) occur each year. The mortality rate associated with acute myocardial infarction in the US is approximately 30%, and more than half of these deaths occur without the patient even reaching the hospital [1]. A recent study of three UK regions suggest that the mortality rate may be even higher [2]. There is ample evidence that the earlier that appropriate treatment is initiated, the more likely that these patients will have a positive outcome [3–6]. Early identification of these patients is critical to successful treatment of

* Correspondence address: Partners Information Systems, 850 Boylston St., Suite 202, Chestnut Hill, MA 02467, USA. Tel.: +1-617-732-9050; fax: +1-978-334-7742.
E-mail address: sjwang@partners.org (S.J. Wang).

0010-4825/00/$ - see front matter
PII: S0010-4825(00)00022-6

the disease. It is therefore, essential that patients developing symptoms suggesting MI seek care and be admitted to a suitably equipped hospital as soon as possible.

Unfortunately, patients often delay substantially before seeking care [2,3]. Ruston et al. [7] suggests that this is due to several factors, including lack of understanding by patients of the symptoms of MI, emotional factors and denial by patients, and inadequate advice by health care workers when patients develop symptoms. Some studies [8] suggest the need for better characterization of symptoms that may be predictive of myocardial infarction. There is evidence that patients with better knowledge of the symptoms of MI will seek help earlier [7]. If patients could get rapid and accurate advice on whether their symptoms were likely to be serious, then it is possible that delays in seeking treatment could be further reduced [9].

Previous work has been published [10–17] that demonstrated that excellent prediction models for MI can be developed with a high degree of discrimination [15]. These models perform well primarily because of the use of EKG findings as predictors in the model [14]. Certain EKG findings, such as ST elevation, are well known to be highly indicative of myocardial injury, eventually leading to myocardial infarction. The discrimination performance of prediction models is typically measured in terms of the area under the receiver operating characteristic (ROC) curve,[1] and these models have areas under the ROC curve as high as 0.959 [15].

The purpose of this study was to determine how well a predictive model would perform based solely upon patient-reportable clinical history factors, without using diagnostic tests or physical exam findings. Even though we would not expect such a model to perform as well as one using these strong predictors, the model may have important practical applications. This type of prediction model might have application outside of the hospital setting to give accurate advice to patients to influence them to seek care in appropriate situations. For example, such a system might be used directly by patients in a patient-oriented software application, or might be used by healthcare workers as a decision support aid in telephone nurse triage. Given the very high mortality from MI in the US, even a small reduction in median time from onset of symptoms to treatment could translate into a substantial number of lives saved.

2. Materials and methods

2.1. Data

Two data sets totaling 1753 patients were used in this study. The derivation data set consisted of 1253 patients who presented to the emergency room in the Edinburgh Royal Infirmary in Scotland. The validation data set consisted of 500 patients from Sheffield, England. Both data sets contained the same 45 potential prediction attributes and a binary outcome variable indicating the presence or absence of MI. The prevalence of MI in the Edinburgh data set was 21.9%, while the prevalence

[1] The Receiver Operating Characteristic curve is a commonly used tool for evaluating the performance of diagnostic tests or prediction models [18–20]. The ROC curve is a plot of sensitivity vs. (1-specificity) over a range of threshold values. The area under the ROC curve is a standard measure of discrimination, which is one measure of the accuracy of a model. The greater the area under the ROC curve, the greater the ability of the test to discriminate between normal and abnormal values. The area under an ROC curve can range from 0.5 (no discriminating ability) to 1.0 (perfect discrimination of normal and abnormal cases).

Table 1
Patient-reportable clinical history factors that were candidates for inclusion as predictor covariates in the models[a]

Variable	Mean	
	Derivation set ($N = 1253$)	Validation set ($N = 500$)
1 Age (years)	57.600	59.994
2 Smokes	0.374	0.364
3 Ex-smoker	0.235	0.224
4 Family history of MI	0.228	0.254
5 Diabetes	0.070	0.098
6 Hypertension	0.160	0.194
7 Hyperlipidemia	0.030	0.036
8 Severe chest pain	0.951	0.926
9 Retrosternal pain	0.736	0.780
10 L chest pain	0.276	0.254
11 R chest pain	0.115	0.124
12 Back pain	0.105	0.148
13 L arm pain	0.465	0.526
14 R arm pain	0.168	0.164
15 Pleuritic pain	0.177	0.156
16 Postural	0.188	0.090
17 Sharp	0.279	0.200
18 Tightness	0.543	0.508
19 Sweating	0.410	0.530
20 Shortness of breath	0.417	0.438
21 Nausea	0.103	0.318
22 Vomiting	0.103	0.102
23 Syncope	0.036	0.066
24 Episodic	0.073	0.166
25 Worsening (min)	17.397	50.368
26 Duration (min)	8.840	12.344
27 Previous angina	0.442	0.438
28 Previous MI	0.333	0.246
29 Worse than angina	0.288	0.324
30 Sex (1 = male)	0.662	0.586
MI	0.219	0.308

[a]Mean values are shown for each variable for both the derivation (Edinburgh) and validation (Sheffield) data sets.

of MI in the Sheffield data set was 30.8%. The 45 potential prediction attributes contained clinical history, physical exam, and EKG findings. Kennedy et al. [14] previously published work using this data demonstrating the use of logistic regression (LR) models to successfully predict the likelihood of MI based on the full set of attributes. Since the purpose of this study was to build models based on patient-reportable history factors only, we removed all EKG findings and all physical exam findings that could not be reasonably expected to be reportable by the patient. Table 1 shows the resulting list of 30 patient-reportable history factors that were included as potential covariates in our

models. For comparison, the mean values for the history factors and the outcome variable (MI) are given. All variables were binary (0= absent or 1= present) except for "Age" (years), "Worsening" (minutes), and "Duration" (minutes). The variable "Postural" indicates whether the pain changes depending on whether the patient is sitting or recumbent. "Worsening" refers to how long the pain has been getting worse. "Worse than angina" refers to whether the patient feels that the current pain is worse than anginal pain. "Sex" refers to the gender of the patient, where 0= female and 1= male.

2.2. Randomization

For both the logistic regression and neural network model building, the derivation set of 1253 cases was randomly split into 2/3 training and 1/3 holdout sets. To compare the effects of different randomization splits on model building, we repeated the randomization process 10 times, and built models for each of the 10 splits.

2.3. Logistic regression

The SAS System for Windows, version 6.12 [21] was used for building the logistic regression (LR) models. We built several types of LR models for comparison. The first type used all 30 patient history factors as covariates. The remaining types used only a subset of history factors that were chosen using various variable selection methods. We built models using the automatic stepwise, forward, and backward variable selection algorithms in SAS using $p = 0.05$ as the entry and exit criteria. We then built a "most popular" model where we first tallied up the number of times a particular variable appeared in any of the automated variable selection algorithms (see "SAS tally" column in Table 2), and then we built a model using the 11 "most popular" variables (see the "SAS picks" column in Table 2). Finally, we asked a cardiologist to identify the history factors he considered to be very important predictors for MI, and we built a model using only these 17 variables (see "Expert picks", Table 2). Each algorithm was run on each of the 10 randomization splits.

2.4. Neural network

NevProp4 release 1 [22] was used for building the neural network models. NevProp4 requires a user-supplied random initialization seed to set the initial weights in the network, and also to randomly split the training set for the "AutoTrain" option. We chose 10 random initialization seeds to determine what effect initial conditions would have on the final model. We attempted to build a total of 100 different neural network models using the 10 initialization seeds on each of the same 10 training/holdout data splits that were used to build the logistic regression models above. For all neural network models, we built 2-layer networks using all 30 variables as inputs and 15 hidden units. For all models, we used a learning rate of 0.001 with momentum set at 0.0. We used NevProp's "AutoTrain" option with NSplits = 5, which randomly splits the allotted training data into training and holdout sets 5 times and takes the average of all of the runs to obtain a target error. The final model is then built by training on all train data using the target error as the stopping criterion.

Table 2
Variable selection for the logistic regression models[a]

Variable	SAS tally	SAS picks	Expert picks
1 Age	27	x	x
2 Smokes	25	x	x
3 Ex-smoker	0		
4 Family history of MI	4		
5 Diabetes	1		x
6 Hypertension	0		
7 Hyperlipidemia	0		x
8 Severe chest pain	0		x
9 Retrosternal pain	7		x
10 L chest pain	7		x
11 R chest pain	0		
12 Back pain	0		
13 L arm pain	27	x	x
14 R arm pain	7		
15 Pleuritic pain	25	x	x
16 Postural	7		x
17 Sharp	19	x	
18 Tightness	1		
19 Sweating	27	x	x
20 Shortness of breath	0		
21 Nausea	24	x	
22 Vomiting	0		
23 Syncope	0		
24 Episodic	25	x	x
25 Worsening	10		x
26 Duration	0		
27 Previous angina	27	x	x
28 Previous MI	25	x	x
29 Worse than angina	0		x
30 Sex	25	x	x

[a]The "SAS tally" column shows the number of times each covariate entered into one of the stepwise, forward, or backward variable selection logistic regression models. The "SAS picks" column indicates the variables that were chosen for the "Most popular" group of models. The "Expert picks" column shows a cardiologist's opinion of important prediction factors for MI; these variables were used in the "expert" models.

3. Results

Table 3 shows the logistic regression results from the derivation (Edinburgh) data set. The results shown are average C-index values (C-index [23] is equivalent to the area under the ROC curve [20]) from the 10 different randomization splits for each of the variable selection methods — all variables, stepwise, forward, and backward variable selection algorithms. Results are also shown for the "most popular" model built using the 11 most commonly selected variables and for the model built using the 17 variables selected by a cardiologist as being highly predictive of MI ("expert"). Three of the

Table 3
Logistic regression summary results showing the average C-indices from the 10 random splits of the derivation data set (Edinburgh, $N = 1253$)[a]

Logistic regression C-indices (Derivation holdout set)		
Model	Average	Range
All	0.8420	(0.8239,0.8681)
Stepwise	0.8352	(0.8084,0.8588)
Backward	0.8367	(0.8084,0.8608)
Forward	0.8350	(0.8084,0.8588)
Most popular	0.8450	(0.8228,0.8727)
Expert	0.8392	(0.8177,0.8631)

[a]The minimum and maximum C-indices found over all 10 splits is also shown. Standard SAS stepwise, forward, and backward variable selection models were run on each split. The 11 variables selected most often by the stepwise, forward, and backward algorithms were then used to build another LR model shown in the row labeled "Most popular". An additional LR model was built using the 17 variables selected by a cardiologist as being highly predictive of MI.

Table 4
Beta coefficients from the best logistic regression model. This model was obtained from split 8 using the 11 "most popular" selected variables

Logistic regression beta coefficients		
Variable	Beta	p-value
Intercept	−6.1005	0.0001
Age	0.0674	0.0001
Smokes	0.7002	0.0020
L arm pain	0.7165	0.0005
Pleuritic pain	−2.9265	0.0048
Sharp	−1.0132	0.0017
Sweating	1.1307	0.0001
Nausea	0.9580	0.0007
Episodic	−2.0136	0.0100
Previous ang	−0.9689	0.0001
Previous MI	−0.7715	0.0013
Sex	0.5236	0.0195

10 stepwise models were not included in the analysis because their build halted prematurely due to the Wald criterion.

Based on average C-index values obtained from the derivation holdout set, the best performing logistic regression models were the ones that used the 11 "most popular" variables, so we chose our final model from this group. We ranked all 10 of these models by C-index. The highest-ranking model had an aberrancy in that the β coefficient for "pleuritic pain" had an abnormally large p-value of 0.9631. This was presumably the result of an uneven random split of the pleuritic pain cases. Consequently, this model was thrown out and the second-highest ranking model was chosen as the final model, which turned out to be split 8. This model had a C-index of 0.8631 on the holdout set.

Table 5
Neural network C-indices from the derivation data set[a]

Neural network C-indices (Derivation holdout set)			
Trial	Seed	C-index Average	C-index Range
1	6754	0.8323	(0.8188,0.8480)
2	9689	0.8324	(0.8159,0.8626)
3	31532	0.8357	(0.8145,0.8611)
4	15202	0.8311	(0.8122,0.8517)
5	892	0.8311	(0.8076,0.8531)
6	6431	0.8356	(0.8155,0.8558)
7	12432	0.8295	(0.8059,0.8563)
8	22169	0.8309	(0.8139,0.8484)
9	5068	0.8348	(0.8098,0.8527)
10	3298	0.8355	(0.8168,0.8523)
Average:		0.8329	(0.8059,0.8626)

[a]For each trial we selected a random seed for NevProp to set initial conditions. Each seed was run on each of the 10 train/holdout splits of the derivation set. Each row gives the average and range of C-indices for that seed from the 10 train/holdout splits. The final number is the average C-index of all models.

This model was selected (Table 4) as the final "best" logistic regression model to be evaluated on the Sheffield validation data set below.

Table 5 shows derivation set (Edinburgh) results from the neural network models. The average C-indices across all 10 splits for each initialization seed are shown in the table. Three out of the 100 models failed to converge to a final model because of unstable oscillations during the training process. The average C-index of the 97 successful models was 0.8329. The network with the highest C-index (0.8611) was found in trial 3 of split 4, and this network was chosen as the "best" neural network to be evaluated on the Sheffield validation data set below.

After selecting the best logistic regression model and the best neural network from our analysis above, we then tested these models on our validation set of 500 patients from Sheffield, England. Fig. 1 shows the resulting ROC curves and C-indices. The final logistic regression model had a C-index of 0.8444 on the Sheffield validation set, and the final neural network had a C-index of 0.8503 on the Sheffield validation set. There was no statistical difference between the two models ($p = 0.38$). Calibration curves for the logistic regression and neural network models are shown in Figs. 2 and 3.

4. Discussion

Note that in the final logistic regression model, some predictive factors had negative β. For some factors, this is not surprising, since it is generally accepted that the presence of a pleuritic or postural component to the chest pain decreases suspicion of MI. However, it may not be as obvious why the other factors — episodic pain, history of previous angina, and history of previous MI — would

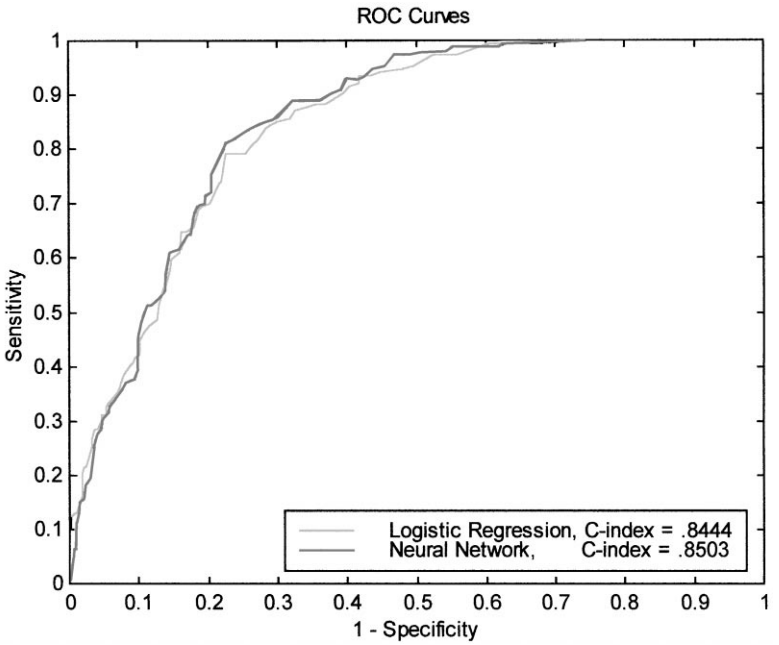

Fig. 1. Receiver operating characteristic curves for the best logistic regression model and the best neural network model when run on the independent Sheffield validation data set.

result in a decreased probability of MI. Nevertheless, somewhat plausible arguments may be devised for some of these attributes. For example, patients that have previously had confirmed angina may be more prone to seek care at an emergency room for any small hint of chest pain. This would tend to increase their rate of "false alarms" in the ER.

The variables selected by the SAS stepwise, forward, and backward selection algorithms were different from the ones named by the cardiologist as being important predictors for MI. There may be several explanations for this, but one of the factors could be the characteristics of the original data set used to derive the model. This is always an important factor to consider when building predictive models derived from retrospective data. By necessity, many simplifying assumptions must be made in order to distill a complex patient history into a series of binary variables for model building, and it likely that some important information is lost in this process. The exact definitions of the fields and the consistency with which the questions were asked are important factors in obtaining an accurate data set. For example, the question of "severe chest pain" is quite subjective. Note in Table 1 that our derivation data set had a greater than 95% positive response to this question. Given such homogeneity, it is obvious that this variable would not do well as a discriminating predictor variable. If a more objective and discriminating scale had been used, such as a 10-point pain rating scale, it may have increased the utility of this attribute in predicting MI. Also note from Table 1 that although the mean values for most of the attributes are comparable between the two data sets, there are more significant differences for some of the attributes. Some of this variation is to be expected when comparing data collected by different institutions.

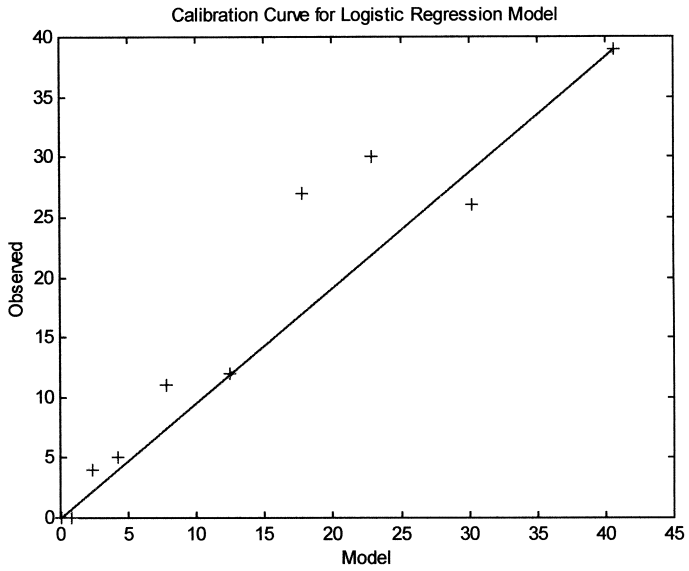

Fig. 2. Calibration curve for the best logistic regression model when run on the independent Sheffield validation data set.

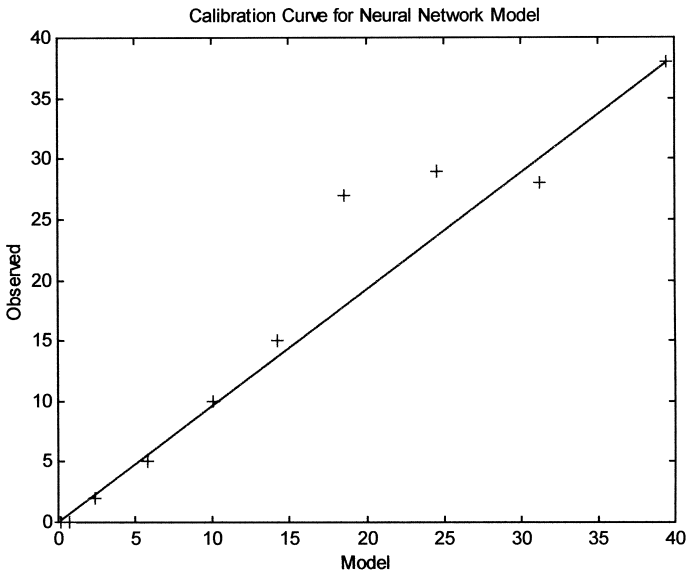

Fig. 3. Calibration curve for the best neural network model when run on the independent Sheffield validation data set.

These prediction models were built and tested using data sets obtained from patients presenting to emergency rooms. This introduces a form of "referral bias" since our derivation data set only contains patients who have already decided to seek care. The eventual goal of this work is to use these models for the population of patients who have not yet decided to seek care. It remains to be seen whether the models will perform as well in these situations where the prevalence rate of MI

will presumably be much lower. It is also possible that the presentation of MI patients in the US may differ from those in Scotland and England. In order to increase confidence in the robustness of our models, they should be further validated using other independently obtained data sets and with prospective testing.

We made a subjective determination as to which prediction factors could be easily reported by a patient. Further evaluation will be needed to determine if patients can accurately report these data items. The accuracy of these factors will obviously affect the performance of the prediction models.

The fact that both the logistic regression and neural network models had comparable performance appears to imply that there are no significant higher-order factors, interaction factors, or non-linearly separable factors in this data set. Furthermore, if we believe in the accuracy of this data set, then this generalizes to imply that for this clinical domain, the probability of MI can be modeled by a function of a simple linear combination of patient-reportable history factors. In future work, other neural network parameters could be adjusted to try to construct better networks. Other variations that could be tested include the number of hidden units, the number of layers, and the number and choice of input variables.

The original motivation for this work was to determine if a clinical software application could be built that could successfully predict the likelihood of a myocardial infarction based on clinical history factors alone. As expected, our models do not perform as well as those that used physical findings and EKG data, but they still performed remarkably well even without this objective information. The C-indices of our models were only about 0.11 lower than models [15] that also incorporated physical findings and EKG test data. Because of the shape of the ROC curve for these models (Fig. 1), a threshold could be chosen for these models to achieve a sensitivity of greater than 90% with a corresponding specificity of about 60%. This combination of sensitivity and specificity may be acceptable in situations such as a screening application where we are willing to tolerate a high rate of false positives in order to maximize the sensitivity of the application.

Our results lead us to believe that these models could be used in real software applications in a clinical setting. If the performance of these models holds up in further studies, software applications could be written using these models that could have important utility in settings outside of a hospital when a healthcare provider may not yet be available. For example, a software application could be designed to assist nurses doing telephone triage when they are assessing a patient's risk of MI over the phone. Alternatively, an application could be designed for direct use by patients to assist them in determining the seriousness of their chest pain symptoms. For example, a standalone software application that assesses chest pain symptoms could be designed to run on a home desktop computer or personal digital assistant (PDA) device. Patients could enter information about their chest pain symptoms and get back an estimate of the likelihood that they are experiencing symptoms of a heart attack and obtain advice on whether or not they should seek professional care. The purpose of these applications would be to enable patients to become better informed about their condition and to encourage them to seek professional care at an earlier stage in the appropriate situations.

5. Summary

The purpose of this study was to build predictive models for estimating the likelihood of myocardial infarction based upon patient-reportable clinical history factors only. These models could

be useful in settings outside of the hospital in encouraging patients seek appropriate care. Two data sets totaling 1753 patients were used to build and test these models. We built several models using logistic regression and neural networks and compared their performance. The first data set consisting of 1253 patients was used to build and train the models. Logistic regression models were constructed using all variables, and using standard stepwise, forward, and backward variable selection algorithms. The most commonly selected variables were then used to construct the "best" LR model. Neural networks were also constructed using the same data sets. The best performing logistic regression model and the best neural network were then chosen and evaluated using a previously unseen data set of 500 patients. The final logistic regression model had a C-index of 0.8444 on the validation data set. The final neural network had a C-index of 0.8503 on the same validation set. The difference in C-indices between logistic regression and neural networks was not statistically significant ($p = 0.38$). We conclude that logistic regression and neural network models can be built that successfully predict the probability of myocardial infarction based on patient-reportable history factors alone. Models that only require patient-reportable factors for prediction may have important applicability as screening tools in settings outside of the hospital when patients need advice on whether or not to seek professional care.

Acknowledgements

The authors would like to thank Fred Resnic, M.D. for providing the cardiologist's opinion. This work was funded in part by contract 467-MZ-802289 and grant R29 LM06538-01 from the National Library of Medicine. Dr. Wang was funded by NLM grant 2 T15 LM07092.

References

[1] E. Braunwald, E.M. Antman, Acute myocardial infarction, in: A.S. Fauci, E. Braunwald, K.J. Isselbacher, J.B. Martin (Eds.), Harrison's Principles of Internal Medicine, McGraw-Hill, New York, 1997 (Chapter 243).

[2] R.M. Norris, Fatality outside hospital from acute coronary events in three British health districts, 1994-5. United Kingdom Heart Attack Study Collaborative Group, Br. Med. J. 316 (1998) 1065–1070.

[3] Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group, Lancet 2 (1988) 349–360.

[4] An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO investigators, N. Engl. J. Med. 329 (1993) 673–682.

[5] Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group [published erratum appears in Lancet 343(8899) (1994) 742], Lancet 343 (1994) 311–322.

[6] J.M. Rawles, Quantification of the benefit of earlier thrombolytic therapy: five-year results of the Grampian Region Early Anistreplase Trial (GREAT), J. Am. Coll. Cardiol. 30 (1997) 1181–1186.

[7] A. Ruston, J. Clayton, M. Calnan, Patients' action during their cardiac event: qualitative study exploring differences and modifiable factors, Br. Med. J. 316 (1998) 1060–1064.

[8] A.P. Burke, A. Farb, G.T. Malcom, Y. Liang, J. Smialek, R. Virmani, Effect of risk factors on the mechanism of acute thrombosis and sudden coronary death in women, Circulation 97 (1998) 2110–2116.

[9] K. Dracup, A.A. Alonzo, J.M. Atkins et al., The physician's role in minimizing prehospital delay in patients at high risk for acute myocardial infarction: recommendations from the National Heart Attack Alert Program. Working Group

on Educational Strategies To Prevent Prehospital Delay in Patients at High Risk for Acute Myocardial Infarction, Ann. Intern. Med. 126 (1997) 645–651.
[10] B. Heden, H. Ohlin, R. Rittner, L. Edenbrandt, Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks, Circulation 96 (1997) 1798–1802.
[11] H.P. Selker, J.L. Griffith, S. Patil, W.J. Long, R.B. D'Agostino, A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients, J. Investig. Med. 43 (1995) 468–476.
[12] H.P. Selker, J.R. Beshansky, J.L. Griffith et al., Use of the acute cardiac ischemia time-insensitive predictive instrument (ACI-TIPI) to assist with triage of patients with chest pain or other symptoms suggestive of acute cardiac ischemia. A multicenter, controlled clinical trial, Ann. Intern. Med. 129 (1998) 845–855.
[13] D. Do, J.A. West, A. Morise, E. Atwood, V. Froelicher, A consensus approach to diagnosing coronary artery disease based on clinical and exercise test data, Chest 111 (1997) 1742–1749.
[14] R.L. Kennedy, A.M. Burton, H.S. Fraser, L.N. McStay, R.F. Harrison, Early diagnosis of acute myocardial infarction using clinical and electrocardiographic data at presentation: derivation and evaluation of logistic regression models, Eur. Heart J. 17 (1996) 1181–1191.
[15] S. Dreiseitl, L. Ohno-Machado, S. Vinterbo, Evaluating variable selection methods for diagnosis of myocardial infarction, Proceedings of AMIA Annual Fall Symposium, 1999, pp. 246–250.
[16] S. Vinterbo, L. Ohno-Machado, A genetic algorithm to select variables in logistic regression: example in the domain of myocardial infarction, Proceedings of AMIA Annual Fall Symposium, 1999, pp. 984–988.
[17] W.G. Baxt, Use of an artificial neural network for the diagnosis of myocardial infarction [published erratum appears in Ann. Intern. Med. 116(1) (1992) 94], Ann. Intern. Med. 115 (1991) 843–848.
[18] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine [published erratum appears in Clin. Chem. 39(8) (1993) 1589], Clin. Chem. 39 (1993) 561–577.
[19] C.E. Metz, Basic principles of ROC analysis, Semin. Nucl. Med. 8 (1978) 283–298.
[20] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1982) 29–36.
[21] The SAS System for Windows, version 6.12, The SAS Institute. Cary, NC. Available at: http://www.sas.com.
[22] P. Goodman, NevProp4: Artificial neural network software for statistical prediction, vers. 4, University of Nevada School of Medicine. Reno, Nevada. Available at: http://www.scs.unr.edu/nevprop.
[23] F.E. Harrell Jr., R.M. Califf, D.B. Pryor, K.L. Lee, R.A. Rosati, Evaluating the yield of medical tests, J. Am. Med. Assoc. 247 (1982) 2543–2546.

Samuel J. Wang is a senior informatics specialist in the Clinical Information Systems Research & Development group at Partners HealthCare System in Boston, MA. Dr. Wang is a graduate of Stanford University, where he received his M.D. with a Ph.D. in Medical Information Systems. He also completed a National Library of Medicine fellowship in the Harvard/MIT Medical Informatics Training Program. Dr. Wang's primary areas of interest include the design of electronic medical record systems, clinical decision support systems, and the evaluation of clinical information systems interventions.

Lucila Ohno-Machado is currently Assistant Professor of Radiology at Brigham and Women's Hospital, Harvard Medical School, and affiliated faculty at the Harvard/MIT division of Health Sciences and Technology. She studied medicine at the University of Sao Paulo and Medical Information Sciences and Computer Science at Stanford. Her primary interests are in building prognostic models in medicine, and developing new evaluation methods for assessing discrimination and calibration of those models. She does research at the Decision Systems Group with a team of physicians, computer scientists, and other professionals, and teaches at MIT.

Hamish S. F. Fraser is an Instructor at Boston Children's hospital and Harvard Medical School. He is also a Research Affiliate at the MIT Laboratory for Computer Science. His training is as a physician in general medicine and cardiology, predominantly from Edinburgh Royal Infirmary and related hospitals in Scotland. He holds a master's degree in Artificial Intelligence/Knowledge Based Systems from Edinburgh University. He recently completed a three-year fellowship in medical informatics split between the MIT Clinical Decision Making Group and the Divisions of Clinical Decision Making and Cardiology at the Tufts-New England Medical Center in Boston. Current research projects include collection of the data on patients with chest pain leading to development programs to assist with the

diagnosis of MI in the ER. He is currently involved in several projects to improve access to medical information in developing countries. Another key focus of his work is ensuring security and confidentiality of medical data.

R. Lee Kennedy is a medical graduate of Edinburgh University. He holds doctorates from the Universities of Edinburgh and Sheffield, and is a fellow of the Royal Colleges of Physicians of Edinburgh and London. His current post is as Professor of Medicine in Sunderland in the north east of England. He has previously held academic posts in the universities of Edinburgh, Cambridge and Sheffield. Research interests include complications of diabetes, and the application of information technology in medical domains.