Dig Dis Sci (2007) 52:1871–1878
DOI 10.1007/s10620-006-9202-5

ORIGINAL PAPER

Applying Data Mining Techniques in the Development of a Diagnostics Questionnaire for GERD

Noya Horowitz · Menachem Moshkowitz · Zamir Halpern · Moshe Leshno

Received: 19 June 2005 / Accepted: 30 December 2005 / Published online: 10 April 2007

Abstract Gastroesophageal reflux disease (GERD) is a common condition, managed mostly in primary care practice. Heartburn and acid regurgitation are considered primary symptoms, and are usually highly specific. However, the symptom spectrum is much wider and in many cases it is difficult to determine whether the patient has GERD or dyspepsia from another origin. The aim of this study is to develop a symptom score and rule for the diagnosis of GERD, using data mining techniques, to provide a clinical diagnostic tool for primary care practitioners in the evaluation and management of upper gastrointestinal symptoms. A diagnostic symptom questionnaire consisting of 15 items and based on the current literature was designed to measure the presence and severity of reflux and dyspepsia symptoms using a 5-point Likert-type scale. A total of 132 subjects with uninvestigated upper abdominal symptoms were prospectively recruited for symptom evaluation. All patients were interviewed and examined, underwent upper gastrointestinal endoscopy, and completed the questionnaire. Based on endoscopic findings as well as the medical interview, the subjects were classified as having reflux disease (GERD) or non-reflux disease (non-GERD). Data mining models and algorithms (neural networks, decision trees, and logistic regression) were used to build a short and simple new discriminative questionnaire. The most relevant variables discriminating GERD from non-GERD patients were heartburn, regurgitation, clinical response to antacids, sour taste, and aggravation of symptoms after a heavy meal. The sensitivity and specificity of the new symptom score were 70%–75% and 63%–78%, respectively. The area under the ROC curve for logistic regression and neural networks were 0.783 and 0.787, respectively. We present a new validated discriminative GERD questionnaire using data mining techniques. The questionnaire is useful, friendly, and short, and therefore can be easily applied in clinical practice for choosing the appropriate diagnostic workup for patients with upper gastrointestinal complaints.

N. Horowitz · M. Moshkowitz · Z. Halpern · M. Leshno
Department of Gastroenterology and Liver Disease,
Tel-Aviv-Sourasky Medical Center,
Tel Aviv, Israel

N. Horowitz · M. Moshkowitz · Z. Halpern · M. Leshno
Sackler Faculty of Medicine, Tel Aviv University,
Tel Aviv, Israel

M. Leshno
Faculty of Management, Tel Aviv University,
Tel Aviv, Israel

M. Leshno (✉)
Faculty of Management, Tel-Aviv University,
Tel Aviv 69978, Israel
e-mail: leshnom@post.tau.ac.il

Introduction

Gastroesophageal reflux disease (GERD) is a common condition that places a considerable burden on health care resources [1]. Most patients are managed in general practice, and effective diagnosis and management of the disease remain a challenge. Heartburn and acid regurgitation are considered primary symptoms of GERD, and, when present as predominant symptoms, they are usually quite specific. In the rest of the cases, patients usually describe several different upper gastrointestinal (GI) symptoms, and in such cases it is difficult to assess whether they have GERD or other causes of dyspepsia.

GERD and dyspepsia are distinct problems and require different management. In the absence of alarm symptoms, the current recommended policy in young dyspeptic patients

is a "test and treat" strategy for *Helicobacter pylori*. On the other hand, in GERD patients, a therapeutic trial with proton pump inhibitors (PPI) is the treatment of choice. Several studies have been undertaken in an attempt to find an accurate symptom questionnaire for the diagnosis of GERD, most of which applied traditional statistical methods for data analysis. A significant drawback inherent in these methods is that questionnaire construction evaluation and validation are carried out on the same population sample ("in-sample analysis"), which may lead to overfitting of the model to the study population, and therefore not reflect the true validity of the questionnaire in daily clinical practice.

Data mining is a new discipline lying at the interface of statistics, database technology, pattern recognition, machine learning, and other areas [2–4] that may be described as the process of trawling through data to identify patterns in a given dataset. It involves many different algorithms to accomplish different tasks, all of which attempt to determine a model that best describes the characteristics of the data being examined. Several methods are being used in the data mining procedure, including traditional statistical methods (e.g., linear regression analysis), neural networks, and decision trees.

The aim of the present study was to identify a predictive cluster of symptoms that discriminates patients with GERD from patients with other types of dyspepsia, and to use methods of data mining—including in-sample and out-sample analysis—to develop a symptom score that might provide a clinical diagnostic tool for general practitioners.

Materials and methods

A diagnostic symptom questionnaire was developed to discriminate GERD from other types of dyspepsia. The items and the response structure of the questionnaire were based on a review of the published literature and feedback from patients, primary care physicians, and gastroenterologists. A preliminary GI symptom questionnaire was prepared by reviewing the instruments developed by other investigators for diagnostic or discriminatory purposes [5–30]. It included items that were found to be significant in previous studies. After review and revision by a group of 3 gastroenterologists, 2 primary care physicians, and 1 statistician, the questionnaire was administered to 15 patients who attended a hospital gastroenterology department with a chief complaint of the upper GI tract (such as epigastric pain, heartburn or other discomfort). The patients were asked to comment on the intelligibility of the questions, and after a minor change in terminology, no further revisions were needed.

The questionnaire was designed to measure both the presence and severity of reflux and dyspepsia using a 5-point Likert-type scale. The evaluated symptoms included epigastric pain/discomfort, retrosternal pain, heartburn, regurgitation, nausea, vomiting, belching, bloating, early satiety, postprandial fullness, halitosis, sour taste, and stress. The questionnaire also measured the influence of different factors on the patient's symptoms, including antacids, bending or lying down, and heavy meals. The presence of alarm symptoms (such as GI bleeding, weight loss, vomiting, dysphagia, anemia, GI polyps or tumors, and past abdominal surgery) was also addressed.

Subject selection

Adults presenting to the Gastroenterology Department at the Tel-Aviv Sourasky Medical Center with a chief upper GI tract complaint were eligible to participate in the study. Patients invited to enter the study provided written informed consent and the study was approved by the local ethics committee.

A total of 139 subjects with uninvestigated upper abdominal symptoms prospectively recruited for symptom evaluation from January 2002 through May 2003 were interviewed. The participants completed the detailed questionnaire regarding their GI symptoms, and later were interviewed by a gastroenterologist as well. All subjects were referred by the physician for gastroscopy; out of the 139 patients 132 consented to undergo upper GI endoscopy.

Subject evaluation

Upper GI endoscopy was performed using a Pentax video endoscope. The endoscopists were aware that the patients were being investigated for this study but were blind to their questionnaire answers. The macroscopic appearance of the esophageal mucosa was recorded according to the Los Angeles classification [31]. Subjects with any grade of esophagitis as seen in the endoscopy were diagnosed as having reflux disease (GERD) versus non-reflux disease (non-GERD). A finding of esophagitis was made by the physician who performed the gastroscopy and was approved by 2 other gastroenterologists reviewing the endoscopic findings. In addition, patients were diagnosed as having GERD—even when the gastroscopy was normal—when 3 senior gastroenterologists, based on the medical record of the patients, agreed that they had GERD. The gastroenterologists who evaluated the medical record were blind to the symptom questionnaire. Cases were classified as GERD if there was total agreement between the 3 gastroenterologists; other cases were diagnosed as non-reflux disease. A similar method was used previously to diagnose patients with GERD, presenting or not presenting with esophagitis [5, 14, 17).

Statistical methods

In this study, we used data mining techniques to find the best model based on the questionnaire to provide the best

classification tool to discriminate between reflux disease (GERD) and non-reflux disease (non-GERD). We used stepwise logistic regression to choose the best questions for classification in the questionnaire. Then we used neural network models (see below for more details), decision tree algorithms, and other classification algorithms to construct the best model to classify GERD, based on our data.

An important issue in data mining is the generalization of the classification (or prediction) capability of a given model or algorithm. The *generalization* of a classification model relates to its ability to classify a new dataset independent of the given dataset. Usually, to estimate the generalization capability one divides the data set into 2 separate sets: the first data set is called the *training set* and the second data set is called the *test set*. The classification-learning model (e.g., logistic regression) uses only the learning set; in the logistic regression model the parameter estimation is based only on the learning dataset. The generalization capability is estimated over the test set with the model obtained over the learning dataset. When the dataset is small, as is the case in our study, the "leave-one-out" method is usually used to estimate the generalization capability of a given model. We used the simple and common leave-one-out method of cross-validation, a technique for estimating the generalization error based on resampling. In *k*-fold cross-validation, the data are divided into *k* subsets of (approximately) equal size. The model (e.g., neural net) is then trained *k* times, each time leaving out 1 of the subsets, but using only the omitted subset to compute whatever error criterion is of interest. If *k* equals the sample size, this is called leave-one-out cross-validation (see Hastie *et al*. [2] pp 214–222). When the model's output is continuous (e.g., in logistic regression), one can also plot the receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC). In this study we estimated the generalization capability of a given model by the leave-one-out method.

We used Clementine 8.0 (SPSS, Chicago, IL) and Matlab 7.0 (Mathwork, Natick, MA) for the following data mining models and algorithms: logistic regression, neural networks, decision trees (C5.0 and Classification and Regression Trees—CART). We next elaborate on each of the algorithms we used to classify GERD.

Logistic regression

We used forward and backward logistic regression for feature selection, that is, to obtain those symptoms that significantly predict GERD. Variables with $P > .25$ were removed. GERD was diagnosed if the probability, using the logistic regression, was $> .5$. In addition, the ROC curve was calculated varying the probability threshold for GERD.

Table 1 Questionnaire items and the average symptom severity

Symptom assessed	Description in the questionnaire presented to the patients	Mean symptom severity non-GERD patients	Mean symptom severity GERD patients	*P*-Value
Abdominal pain	Upper abdominal pain or discomfort	3.12	2.80	.16
Chest pain	Chest pain or discomfort	2.15	1.85	.16
Bloating	Upper abdominal bloating, early satiety or post-prandial fullness	2.52	2.13	.10
Nausea	Nausea or vomiting	2.30	2.20	.68
Belching	Belching	2.00	2.69	.01
Heartburn	Retrosternal or upper abdominal burning sensation (heartburn)	2.42	3.44	.00
Regurgitation	Gastric content regurgitation	2.15	3.02	.00
Dysphagia	Difficulty or pain in swallowing	1.87	1.93	.91
Sour taste	Oral sour taste	1.92	2.44	.04
Halitosis	Bad breath	1.78	1.92	.77
Antacid	Symptom relieved by antacid	1.27	1.70	.13
Stress	Emotional stress	1.82	1.80	.96
Bend/lie aggravation	Aggravation of symptoms by bending or lying	0.95	1.64	.01
Heavy meal aggravation	Aggravation of symptoms after heavy meal	1.95	2.16	.51

Table 2 The relevant variables to discriminate GERD versus non-GERD patients

	Coefficient	SE	Sig	OR	95.0% CI for EXP(B)	
					Lower	Upper
Abdominal pain	−.466	.200	.019	.627	.424	.928
Chest pain	−.547	.205	.007	.579	.387	.864
Nausea	−.352	.176	.045	.703	.498	.993
Heartburn	.647	.187	.001	1.910	1.323	2.758
Regurgitation	.471	.220	.032	1.601	1.041	2.462
Antacid	.264	.147	.071	1.303	.977	1.737

Neural networks

Neural network theory is a collection of models of computation and learning that possess a number of properties similar to natural systems. Neural nets can learn from experience, generalize, infer from past examples to new ones, and extract essential attributes from inputs that include irrelevant information. These qualities, together with the ability to recognize what the input represents even when it is distorted or noisy, have made neural nets an attractive vehicle for attempts to develop medical diagnostic systems.

We used a neural network model consisting of one hidden layer (which suffices to approximate or classify any relationship between the input variables and the output variable, see Leshno *et al*. [32]) with various numbers of neurons in the hidden layer. Using the Backpropagation learning algorithm, we minimized the error by fitting the best weights of the neural networks on the data. Next we examined the sensitivity and specificity of the model. The criterion for best model was based on the area under the ROC curve.

Decision trees

A decision tree is a predictive modeling technique used in classification tasks that uses a "divide and conquer" technique to split the problem search into subsets. Whatever the technique used in their creation, decision trees are easy to use and efficient. Rules can be generated that are easy to interpret and understand. In this study we apply the CART algorithm.

Results

The combined endoscopic and clinical diagnosis of the 132 patients (M/F = 70/62, mean age 45.2 years, range 20–87) were GERD in 72 patients (55%), functional dyspepsia in 52 (39%), peptic ulcer in 7 patients (5%), and gastric cancer in 1 patient (1%).

Table 1 displays the questionnaire items and the average of the responses of GERD and non-GERD patients. The *P*-values of the difference between GERD and non-GERD patients were calculated using the Mann–Whitney test.

Using stepwise logistic regression, we identified the family of variables that were significant predictors of GERD. Table 2 displays the relevant variables with their coefficient and *P*-value in the logistic regression analysis. Note that when the questions for "Heartburn," "Regurgitation," and "Antacid" were rated higher the probability of GERD increased; however, when the questions for "Abdominal Pain," "Chest Pain," or "Nausea" were rated higher the probability of GERD decreased.

Using the leave-one-out method, we estimated the sensitivity, specificity, total error and the AUC of the ROC curve for the following classification algorithms: logistic regression, neural network model, and decision tree (as the output of a decision tree is dichotomous the AUC is irrelevant for this classification model). Table 3 summarizes the total error, sensitivity, and specificity, as well as the area under the ROC curve for each of the models. The sensitivity and specificity using logistic regression were 75.3% and 62.9%, respectively. The sensitivity and specificity using neural networks with 10 neurons in 1 hidden layer were 70% and 78%, respectively. The sensitivity and specificity using the decision trees algorithm were 55.0% and 76.0%, respectively.

Figures 1 and 2 present the ROC curve for logistic regression and neural networks respectively. The areas under the ROC curve were 0.783 and 0.957, respectively. Table 4 presents the AUC for the logistic regression and neural networks algorithms.

We next present the classification error over the in-sample data for logistic regression, decision trees and neural networks to compare our results to those of other published studies. Table 5 summarizes the total error, sensitivity, and specificity, as well as the area under the ROC curve for each of the models over the in-sample data. The sensitivity and

Table 3 Leave-one-out method to estimate sensitivity specificity AUC and total error

Model	Logistic Regression	Neural network	Decision tree
Sensitivity	75.3%	70.0%	55.0%
Specificity	62.9%	78.0%	76.0%
Total error	29.9%	26.8%	32.0%
AUC	0.783	0.787	—

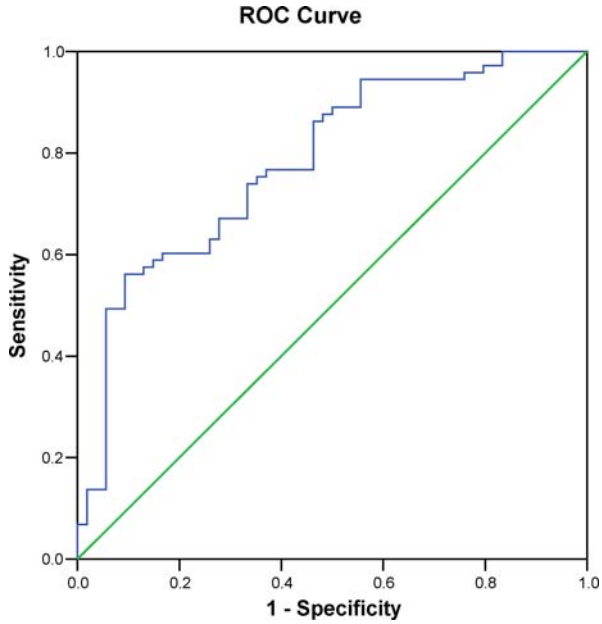

Fig. 1 ROC curve for logistic regression with the leave-one-out method

specificity using logistic regression were 80.8% and 66.7%, respectively. The sensitivity and specificity using neural networks with 10 neurons in 1 hidden layer were 87.7% and 94.4%, respectively. The sensitivity and specificity using the decision trees algorithm were 83.3% and 84.9%, respectively.

Figures 3 and 4 present the ROC curve for logistic regression and neural networks respectively in the in-sample analysis. The areas under the ROC curve were 0.824 and 0.957, respectively. Table 6 presents the AUC for the logistic regression and neural networks algorithms with the in-sample data analysis.

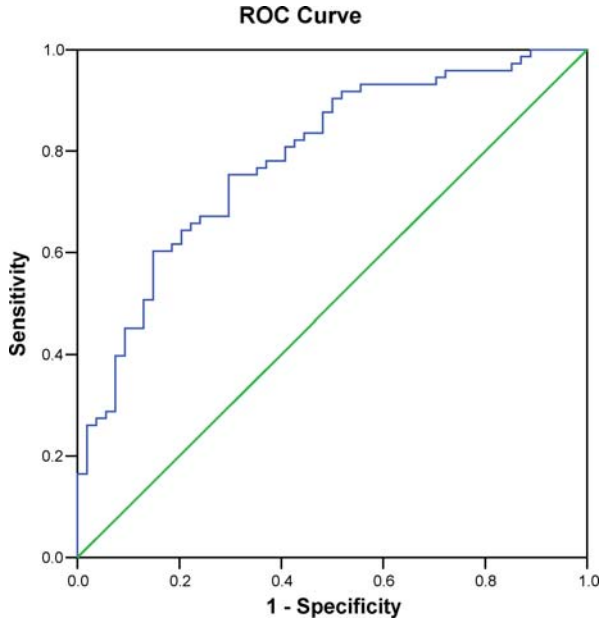

Fig. 2 ROC curve for neural networks with the leave-one-out method

Table 4 AUC for logistic regression and neural network with the leave-one-out method

	AUC	Standard Error	*P*-Value	95% CI Lower bound	95% CI Upper bound
Logistic regression	0.783	0.041	.0001	0.702	0.863
Neural network	0.787	0.040	.001	0.708	0.866

Discussion

Of the several GERD symptom questionnaires that have been published to date [5–19], some were developed mainly as an evaluative instrument for epidemiologic purposes or as a primary outcome in clinical trials [33]. The current questionnaire aims to be a discriminative instrument for the initial assessment and management of patients with upper GI complaints in the primary care practice. As such, it does not pretend to be the sole diagnostic method, but to guide the physician in choosing the appropriate diagnostic workup. Most important is to discriminate between patients with GERD and those with other causes of dyspepsia whose response to therapy is less predictable.

Using data mining techniques, primarily neural networks, we found the current questionnaire fairly sensitive and specific: 70% and 78%, respectively (out-sample analysis) and 87.7% and 94.4%, respectively (in-sample analysis). This does well in serving the questionnaire's aim, to identify most of the patients with GERD who will respond to a PPI empirical trial, while excluding patients suffering from other causes of dyspepsia, who will be referred for another diagnostic workup (*H. pylori* examination or endoscopy).

In developing the questionnaire, we determined the diagnosis of GERD according to the endoscopic results, and a comprehensive evaluation of the cases and the response to antisecretory drugs by 3 senior gastroenterologists. This ensures the inclusion of both erosive-GERD patients and

Table 5 In-sample data estimation for sensitivity specificity AUC and total error

Model	Logistic regression	Neural network	Decision tree
Sensitivity	80.8%	87.7%	83.3
Specificity	66.7%	94.4%	84.9
Total error	25.2%	9.4%	
AUC	0.824	0.957	—

Table 6 AUC for logistic regression and neural network with the in-sample data analysis

	AUC	Standard error	*P*-Value	95% CI Lower bound	95% CI Upper bound
Logistic regression	0.824	0.037	.0001	0.751	0.897
Neural network	0.957	0.019	.001	0.92	0.994

nonerosive GERD patients. The latter have been shown to be the majority of GERD patients, and no differences have been found between the 2 groups in relation to the character and severity of the symptoms [34–36].

Most of our patients did not undergo 24-hour pH monitoring. This test is not considered an established gold standard for GERD and recent studies suggest that the specificity of GERD diagnosis could be enhanced by using the therapeutic PPI trial and/or the physician's clinical diagnosis, based on a structured interview with the patient [17, 20, 37–39].

The main innovation of our study is the implementation of new methods in identifying the most predictive and relevant items for the GERD questionnaire. Previous studies used logistic regression methods for the development of the various GERD questionnaires and achieved fairly high sensitivity and specificity rates (Table 7). Most of them used in-sample sensitivity, specificity, and AUC of the ROC curve to evaluate the questionnaire validity. Significant decreases in sensitivity and specificity rates have been demonstrated using out-sample analysis. This is the result of the fact that using in-sample analysis means that the questionnaire construction and validation are evaluated on the same population

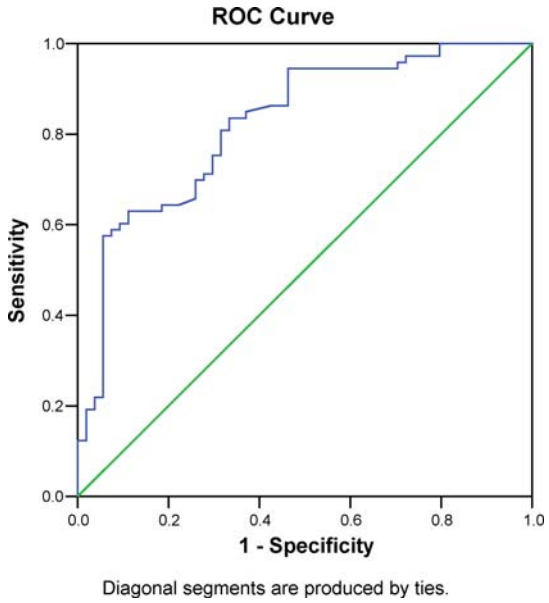

Fig. 3 ROC curve for logistic regression with the in-sample analysis

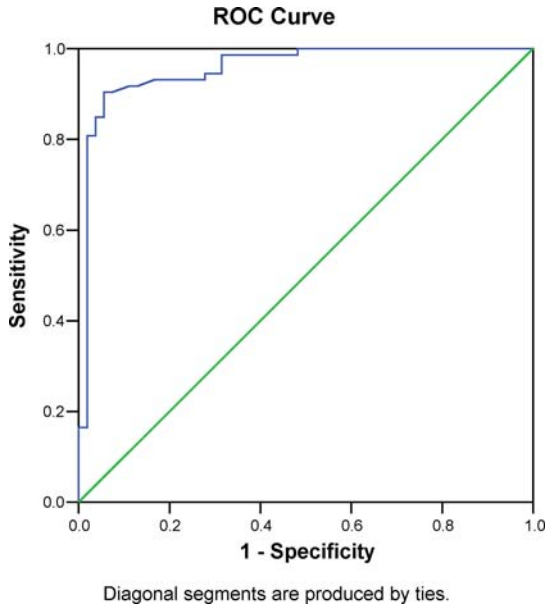

Fig. 4 ROC curve for neural networks with the in-sample analysis

sample, which may lead to overfitting of the model to the study population, and therefore not reflect the true validity of the questionnaire in daily clinical practice. For example, using our in-sample data, a complex neural network architecture with 40 neurons in the hidden layer resulted in sensitivity and specificity rates of 97% and 96%, respectively, and AUC of 0.994. However, using the leave-one-out method with this complex neural network architecture resulted in significantly lower sensitivity and specificity rates—64.4% and 50.0%, respectively—and an AUC of 0.663. Thus, we think that validation of a clinical questionnaire should be carried out on out-sample data and studies that validated their questionnaire using in-sample analysis cannot be reasonably compared to studies that used out-sample data. This is well demonstrated by the significant decrease in the specificity rate of the Carlsson-Dent questionnaire using out-sample versus in-sample analysis [10].

To be applicable for daily use in clinical practice the questionnaire has to be useful, short, and friendly to both patients and physicians. Our questionnaire fulfills these requirements. The discriminate models used in our study, namely logistic regression, neural nets, and decision trees, can be easily implemented in computerized medical systems. More specifically, patients will be asked to respond to 6 questions regarding the nature and severity of their symptoms. The data are then entered electronically into the patient's chart. Within this classification model, which will be transparent to the treating physician, a scoring module is generated from the chart, constituting a specific recommendation to the diagnostic workup of the patient.

Table 7 Data analysis of questionnaires for GERD and dyspepsia

Gold standard	AUC	Specificity	Sensitivity	In/out sample	Author
Gastroscopy	–	46%	70%	In-sample	Carlsson and Dent 1998 [10]
Gastroscopy & pH-monitoring	–	19%	92%	Out-sample	Carlsson & Dent 1998 [10]
Gastroscopy & response to PPI treatment	0.65	50–73%	48–73%	Out-sample	Numans 2003 [14]
pH-monitoring	0.93	94.9%	91.6%	In-sample	Manterola 2002 [12]
Clinical diagnosis (agreement) by 2 gastro-enterologists	0.89	83%	83%	In-sample	Ofman 2002 [17]
GP or Gastro-enterologist opinion	—	79% (GP population) 53% (GE population)	80% (GP population) 99% (GE population)	In-sample	Moayyedi 1998 [5]
Gastroscopy	0.86	80%	92%	In-sample	Brown 2003 [19]
Gastroscopy or pH-monitoring	0.92	84%	82%	In-sample	Wong 2003 [13]

Acknowledgements Noya Horowitz and Menachem Moshkowitz equally contributed to this work.

References

1. Ofman JJ (2003) The economic and quality-of-life impact of symptomatic gastroesophageal reflux disease. Am J Gastroenterol 98:S8–S14
2. Hastie T, Tibshirnai R, Friedman J (2001) The elements of statistical learning, data mining, inference and prediction. Springer, New York
3. Hand DJ, Mannila H, Smyth P (2001) Principles of data mining (adaptive computation and machine learning). MIT Press, Cambridge, MA
4. Klosgen W, Zytkow JM (eds) (2002) Handbook of data mining and knowledge discovery. Oxford University Press, New York
5. Moayyedi P, Duffett S, Braunholtz D, Mason S, Richards ID, Dowell AC, Axon AT (1998) The Leeds Dyspepsia Questionnaire: a valid tool for measuring the presence and severity of dyspepsia. Aliment Pharmacol Ther 12:1257–1262
6. Kline-Leidy N, Farup C, Rentz AM, Ganoczy D, Koch KL (2000) Patient-based assessment in dyspepsia. Development and validation of Dyspepsia Symptom Severity Index (DSSI). Dig Dis Sci 45:1172–1179
7. Talley NJ, Phillips SFP, Melton LJ 3rd, Wiltgen C, Zinsmeister AR (1989) A patient questionnaire to identify a bowel disease. Ann Intern Med 111:671–674
8. Shaw MJ, Talley NJ, Beebe TJ, Rockwood T, Carlsson R, Adlis S, Fendrick AM, Jones R, Dent J, Bytzer P (2001) Initial validation of a diagnostic questionnaire for gastroesophageal reflux disease. Am J Gastroenterol 96:52–57
9. El-Omar EM, Banerjee S, Wirz A, McColl KE (1996) The Glasgow Dyspepsia Severity Score—a tool for the global measurement of dyspepsia. Eur J Gastroenterol Hepatol 8:967–971
10. Carlesson R, Dent J, Bolling-Sternevald E, Johnsson F, Junghard O, Lauritsen K, Riley S, Lundell L (1998) The usefulness of a structured questionnaire in the assessment of symptomatic gastroesophageal reflux disease. Scand J Gastroenterol 33:1023–1029
11. Talley NJ, Boyce PM, Owen BK, Newman P, Paterson KJ (1995) Initial validation of a bowel symptom questionnaire and measurement of chronic gastrointestinal symptom in Australians. Aus N Z J Med 25:302–308
12. Manterola C, Munos S, Grande L, Bustos L (2002) Initial validation of a questionnaire for detecting gastroesophageal reflux disease in epidemiological settings. J Clin Epidemiol 55:1041–1045
13. Wong WM, Lam KF, Lai KC, Hui WM, Hu WH, Lam CL, Wong NY, Xia HH, Huang JQ, Chan AO, Lam SK, Wong BC (2003) A validated symptoms questionnaire (Chinese GERDQ) for the diagnosis of gastro-esophageal reflux disease in Chinese population. Aliment Pharmacol Ther 17:1407–1413
14. Numans ME, De Wit NJ (2003) Reflux symptoms in general practice: diagnostic evaluation of the Carlsson-Dent gastro-esophageal reflux disease questionnaire. Aliment Pharmacol Ther 17:1049–1055
15. Allen CJ, Parameswaran K, Belda J, Anvari M (2000) Reproducibility, validity and responsiveness of a disease-specific symptom questionnaire for gastroesophageal reflux disease. Dis Esophagus 13:265–270
16. Rothman M, Farup C, Stewart W, Helbers L, Zeldis J (2001) Symptoms associated with gastroesophageal reflux disease; development of a questionnaire for use in clinical trials. Dig Dis Sci 46:1540–1548
17. Ofman JJ, Shaw M, Sadik K, Grogg A, Emery K, Lee J, Reyes E, Fullerton S (2002) Identifying patients with gastroesophageal reflux disease; validation of a practical screening tool. Dig Dis Sci 47:1863–1869
18. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR (1994) A new questionnaire for gastroesophageal reflux disease. Mayo Clin Proc 69:539–547
19. Brown WH, Chey WD, Elta GH (2002) Number of responses on a review of systems questionnaire predicts the diagnosis of functional gastrointestinal disorders. J Clin Gastroenterol 36:222–227
20. Moayyedi P, Axon AT (1999) The usefulness of the likelihood ratio in the diagnosis of dyspepsia and gastroesophageal reflux disease. Am J Gastroenterol 94:3122–3125
21. Shaw M, Talley NJ, Adlis S, Beebe T, Tomshine P, Healey M (1998) Development of a digestive health status instrument: tests of scaling assumptions, structure and reliability in primary care population. Aliment Pharmacol Ther 12:1067–1078
22. Small PK, Loudon MA, Waldron B (1995) Importance of reflux symptoms in functional dyspepsia. Gut 36:189–192
23. Talley NJ, Weaver AL, Tesmer DL, Zinsmeister AR (1993) Lack of discriminant value of dyspepsia subgroups in patients referred for upper endoscopy. Gastroenterology 105:1378–1386
24. Junghard O, Lauritsen K, Talley NJ, Wiklund IK (1998) Validation of seven graded diary cards for severity of dyspepsia symptoms in patients with non ulcer dyspepsia. Eur J Surg Suppl 583:106–111
25. Wallace MB, Durkalski VL, Vaughan J, Palesch YY, Libby ED, Jowell PS, Nickl NJ, Schutz SM, Leung JW, Cotton PB (2001) Age and alarm symptoms do not predict endoscopic findings among patients with dyspepsia: a multicenter database study. Gut 49:29–34
26. Kolodny M (1998) Computerized disease management algorithms—the future is now. Eur J Surg Suppl 583:104–105

27. Stanghellini V, Tosetti C, Paternico A, De Giorgio R, Barbara G, Salvioli B, Corinaldesi R (1999) Predominant symptoms identify different subgroups in functional dyspepsia. Am J Gastroenterol 94:2080–2085
28. Mujica VR, Rao SSC (1999) Recognizing atypical manifestations of GERD. Asthma, chest pain and otolaryngologic disorders may be due to reflux. Postgrad Med 105:53–55
29. Weusten BL, Roelofs JM, Akkermans LM, Van Berge-Henegouwen GP, Smout AJ (1994) The symptom-association probability: an improved method for symptom analysis of 24-hour esophageal ph data. Gastroenterology 107:1741–1745
30. Grainger SL, Klass HJ, Rake MO, Williams JG (1994) Prevalence of dyspepsia: the epidemiology of overlapping symptoms. Postgrad Med J 70:154–161
31. Armstrong D, Bennett JR, Blum AL, Dent J, De Dombal FT, Galmiche JP, Luncell L, Margulies M, Richter JE, Spechler SJ, Tytgat GN, Wallin L (1996) The endoscopic assessment of oesophagitis: a progress report on observer agreement. Gastroenterology 111:85–92
32. Leshno M, Lin VY, Pinkus A, et al (1993) Multilayer feedforward networks with a non polynomial activation function can approximate any function. Neural Networks 6:861–867
33. Stanghellini V, Armstrong D, Monnikes H, Bardhan KD (2004) Systematic review: do we need a new gastro-oesophageal reflux disease questionnaire? Aliment Pharmacol Ther 19:463–479
34. Carlsson R, Holloway RH (2000) Endoscopy-negative reflux disease. Baillier Clin Gastroenterol 14:827–837
35. Stanghellini V, Cogliandro R, Cogliandro L, De Giorgio R, Barbara G, Corinaldesi R (2003) Unsolved problems in the management of patients with gastroesophageal reflux disease. Dig Liv Dis 35:843–848
36. Fass R, Fennerty MB, Vakil N (2001) Nonerosive reflux disease—current concepts and dilemmas. Am J Gastroenterol 96:303–314
37. Shaw M (2004) Diagnostic utility of reflux disease symptoms. Gut 53(Suppl 4):25–27
38. Schenk BE, Kuipers EJ, Klinkenberg-Knol EC, Festen HP, Jansen EH, Tuynman HA, Schrijver M, Dieleman LA, Meuwissen SG (1997) Omeprazole as a diagnostic tool in gastroesophageal reflux disease. Am J Gastroenterol 92:1997–2000
39. Fass R, Ofman JJ, Gralnek IM, Johnson C, Camargo E, Sampliner RE, Fennerty MB (1999) Clinical and economic assessment of the omeprazole test in patients with symptoms suggestive of gastroesophageal reflux disease. Arch Intern Med 159:2161–2168
40. Talley NJ, McNeil D, Piper DW (1987) Discriminant value of dyspeptic symptoms: a study of the clinical presentation of 221 patients with dyspepsia of unknown cause, peptic ulceration, or cholelithiasis. Gut 20:40–46