MEDICAL CARE
Volume 39, Number 9, pp 968–978

Screening for High Utilizing Somatizing Patients Using a Prediction Rule Derived From the Management Information System of an HMO

A Preliminary Study

Robert C. Smith, MD, ScM,* Joseph C. Gardiner, PhD,† Stacey Armatti, MS,† Monica Johnson, MD,‡ Judith S. Lyles, PhD,* Charles W. Given, PhD,§ Catherine Lein, RN,¶ Barbara Given, RN, PhD,¶ John Goddeeris, PhD,‖ Elie Korban, MD,* Robert Haddad, MD,* and Mohammed Kanj, MD*

Background. Somatization is a common, costly problem with great morbidity, but there has been no effective screening method to identify these patients and target them for treatment.

Objectives. We tested a hypothesis that we could identify high utilizing somatizing patients from a management information system (MIS) by total number of visits and what we termed "somatization potential," the percentage of visits for which ICD-9 primary diagnosis codes represented disorders in the musculoskeletal, nervous, or gastrointestinal systems or ill-defined complaints.

Methods. We identified 883 high users from the MIS of a large staff model HMO as those having six or more visits during the year studied (65th percentile). A physician rater, without knowledge of hypotheses and predictors, then reviewed the medical records of these patients and identified somatizing patients (n = 122) and nonsomatizing patients (n = 761). In two-thirds of the population (the derivation set), we used logistic regression to refine our hypothesis and identify predictors of somatization available from the MIS: demographic data, all medical encounters, and primary diagnoses made by usual care physicians (ICD-9 codes). We then tested our prediction model in the remaining one-third of the population (the validation set) to validate its usefulness.

Results. The derivation set contained the following significant correlates of somatization: gender, total number of visits, and percent of visits with somatization potential. The c-statistic, equivalent to the area under the ROC curve, was 0.90. In the validation set, the explanatory power was less with a still impressive c-statistic of 0.78. A predicted probability of 0.04 identified almost all somatizers, whereas a predicted probability of 0.40 identified about half of all somatizers but produced few false positives.

Conclusions. We have developed and validated a prediction model from the MIS that helps to distinguish chronic somatizing patients from other high utilizing patients. Our method requires corroboration but carries the promise of providing clinicians and health plan directors with an inexpensive, simple approach for identifying the common somatizing patient and, in turn, targeting them for treatment. The screener does not require clinicians' time.

Key words: Screening; somatization; utilization; prediction rule; HMO; management information system; hypothesis-testing. (Med Care 2001;39:968–978)

*From the Department of Medicine, Michigan State University, East Lansing, Michigan.

†From the Department of Epidemiology, Michigan State University, East Lansing, Michigan.

‡From the Department of Emergency Medicine, Michigan State University, East Lansing, Michigan.

§From the Department of Family Practice, Michigan State University, East Lansing, Michigan.

¶From the College of Nursing, Michigan State University, East Lansing, Michigan.

‖From the Department of Economics, Michigan State University, East Lansing, Michigan.

Address correspondence and reprint requests to: Robert C. Smith, B312 Clinical Center, 138 Service Road, East Lansing, MI 48824. E- mail: smithrr@pilot.msu.edu

Received September 21, 2000; initial decision November 20, 2000; accepted April 3, 2001.

Somatization is defined here as the presence of physical symptoms of at least 6 months duration with no organic disease explanation or, occasionally, where an organic disease is present but does not fully explain the frequency or intensity of a patient's symptoms.[1–3] Clinicians seldom recognize somatization[4–7] and may treat it inappropriately when they do,[7–9] leading to frustrated patients, repeated visits, and many requests for tests in a desperate attempt to get well.[7–9] Unnecessary tests, ill-advised hospitalizations and surgery, "trial" treatments (eg, antibiotics, corticosteroids) for unidentified organic diseases, and use of addicting and other medications lead to high iatrogenic complication rates.[7,8,10–12]

Chronic somatizers have high psychiatric comorbidity, persistent high use of medical and mental health services, high levels of disability, and increased unemployment.[13] Compared with chronically ill medical patients as well as to normals, they have decreased physical functioning, lower perceptions of their state of health, and worse mental health.[14] Severe chronic somatizing patients spent an average of 7 days in bed in the month before one study, compared with one-half day for the general medical population; 83% had stopped work because of their symptoms.[14] Many studies show high costs and excessive utilization in somatizing patients.[11,14–19] Somatizers thus are a disabled group of patients who are not improving and, if anything, getting worse in spite of high utilization and costly care.

We addressed one dimension of the somatization problem by evaluating a screening procedure to identify somatizers from computerized data systems in a staff model HMO. We wanted to efficiently identify high utilizing patients with no organic disease explanation for physical symptoms of at least 6 months' duration.[1,3] A successful screening procedure will not burden busy clinicians and will identify "at risk" patients who can be targeted for more effective treatment than usually is provided.[20–22]

During an unpublished pilot study, we observed that many chronic somatizing patients could be identified by screening computerized data tapes (Management Information System – MIS) for number of visits and types of diagnoses (ICD-9 codes) during a 1 year period. Review of patients' medical charts then confirmed the presence (or absence) of somatization. In this paper we test the hypothesis that we can identify high utilizing somatizing patients from the MIS by total number of visits and "somatization potential," the percentage of visits for which the ICD-9 primary diagnosis code represents any disorder in the musculoskeletal, nervous, or gastrointestinal systems or ill-defined complaints. We developed a prediction rule in one group of patients and validated it in another.

Materials and Methods

From the three Lansing branches of the largest HMO in Michigan, we first generated a list from the MIS of all high users of care, arbitrarily defined as having six or more visits in the year of study (65th percentile). Then a physician rater, unaware of the hypothesis or predictors, reviewed patients' charts and identified chronic somatizing patients (as the gold standard) so that we could test our prediction.

Only computerized descriptive information and data in patients' clinical charts were involved in this study, which did not include informed consent and was in compliance with requirements for human subjects. The MIS included information on age, gender, all medical encounters, primary diagnoses (ICD-9 codes) from patients' usual care and other physicians, revenue codes, and charges for services. Subjects with any visits for pregnancy, substance abuse, or psychiatric attention were excluded.

Screening to Identify Somatizing Patients

From the MIS, we first identified all 21 to 55 year old patients who had at least one visit during 1995 to a physician, physician assistant, nurse practitioner, specialist, or emergency room; each hospitalization was counted as one visit. We did not use older patients because our goal, for another project, was to identify chronic somatizing patients with minimal comorbid organic disease. Of the 15,505 members at that time, 5,423 had six or more visits (65th percentile), and 1,000 of these patients were randomly selected for further evaluation. Of the 1,000, 94 were excluded because of pregnancy, substance abuse, visits for psychiatric care, or because they were employees of the HMO; 23 were excluded because of incomplete data. The remaining 883 patients constituted the study population. Excluded patients differed from the study group patients in age, gender, and employer

group, but not on the amount of copay ($P = 0.58$) and relationship to subscriber ($P = 0.23$). Excluded patients were on average younger (35.7 years vs. 40.3 years, $P < 0.0001$) and 88% were female versus 67% for patients included in the overall study ($P < 0.0001$).

From our hypothesis, patients screened positive with increasing numbers of visits per year and increasing percentages of ICD-9 diagnoses in the following ranges: 320 to 389, 520 to 579, 710 to 739, or 780 to 799. These codes represent, respectively, all diagnoses in the nervous system, gastrointestinal system, musculoskeletal system, and ill-defined complaints, the diagnoses we referred to as having somatization potential; the value of clustering ICD-9 codes for other purposes has been demonstrated.[23] The following are examples from each of the four categories and the diagnoses that often had no organic disease explanation (on our chart review) are italicized: nervous system: meningitis, encephalitis, *migraine, vertiginous syndromes*; digestive system: cholelithiasis, regional enteritis, *gastritis/duodenitis, irritable colon, constipation*; musculoskeletal system: systemic sclerosis, polymyositis, *unspecified disorder of joint, rheumatism and fibrositis, lumbago/sciatica/backache*; ill-defined condtions: coma, epistaxis, *malaise/fatigue, dizziness/giddiness, headache*. Table 1 shows the percent of all visits that were in the four main diagnostic categories. For example, in our sample of 883 patients, 231 had one or more visits that were coded as "disease of the digestive system" (gastrointestinal [GI]). For these patients, the GI visits constituted on average 17.7% of all visits (median 14.3%; range 1.4% to 83.3%).

Gold standard diagnoses were established by a physician rating the 883 charts according to specific criteria. To offset expectation and diagnostic suspicion biases,[24] the rater was blinded to the hypothesis and predictors and did not use gender, number of visits, or specific symptom complexes in her diagnostic evaluations. Rather, a patient was evaluated based upon the presence or absence of data indicating organic diseases. Patients were classified categorically by their primary physical problem for the entire year as organic disease, somatization, or minor acute illness; psychiatric disease was also rated but rarely found. The primary problem was defined as the one generating the largest number of visits during the year. *Organic disease* was diagnosed by standard medical criteria and based upon clear physical signs of disease (eg, laceration, enlarged liver) and, almost always, definitive laboratory and consultative investigation; the rater relied upon expert judgment and referred to text material as needed.[25] Somatization was rated when, following objectively-based diagnostic evaluation (definitive testing and consultative evaluation), patients were free of organic disease that contributed significantly to at least one physical symptom which must have been of at least 6 months' duration. *Minor acute illness* was rated when all physical symptoms were less than 6 months' duration and there was no documentation of an organic disease explanation for any symptoms or their degree of severity. Because minor acute problems typically were not severe or disabling, in contrast to somatization, definitive testing and consultation only occasionally had been performed, ie, patients with no investigation were not rated as somatizers. Symptom syndromes, such as irritable bowel syndrome or fibromyalgia, were classified by the above criteria as somatization or minor acute illness, and criteria unique for each symptom syndrome were not used, so that we have avoided some of the symptom overlap problem that can occur.[26] Psychiatric disease was defined by nonphysical symptoms such as stress or depression.

We used a single rater because we lacked funding to support multiple raters and because rating had been straightforward. After 10 h initial training, including practice rating on nonstudy charts, the rater rated 20 charts of nonstudy high utilizing patients. A priori, we set an agreement rate with the trainer, one of the authors (RCS), for primary diagnosis, of 90% (18 of 20 charts) before the rater began rating study patients. During the study, the trainer rated sets of 20 study charts already rated by the rater on three evenly spaced occasions during 1995, 1996, and 1997. The rater had high levels of agreement with the trainer throughout, varying from 90% to 95%. This level of agreement is not surprising because the trainer trained the rater, reflected also in the kappa of 0.93. Detailed instructions for the rater are available from the authors.

Statistical Methods

We used patient clinical chart review as the gold standard for somatization. Comparisons between somatizers and nonsomatizers were made by χ^2

TABLE 1. Distribution of All Visits by Diagnostic Category

Category/ICD codes	N†	Percent of Total Visits*		
		Mean (SD)	Median	Range
Digestive system, 520–579	231	17.7 (12.6)	14.3	1.4–83.3
Musculoskeletal and connective tissue, 710–739	450	27.9 (18.9)	22.2	1.4–100
Nervous system and sense organs, 320–389	311	19.3 (14.0)	14.3	2.4–87.5
Symptoms, signs, and ill-defined conditions, 780–799	446	19.8 (13.4)	16.7	2.4–85.7

*Computed for patients with visits of specified type.

†N = number patients with at least one visit in specified category. Because patients overlap categories, total number exceeds sample size of 883.

analyses for categorical variables and by t-tests for continuous measures. When continuous measures did not exhibit a normal distribution we used the Wilcoxon Rank test instead.

We randomly selected two-thirds of our 883 patients as a derivation set for estimating our model. We employed logistic regression to assess the relative influence of the number of patient visits, ICD-9 diagnostic categories, and patient demographic factors on the likelihood of somatization—all the variables available from the MIS. Demographic variables included age, gender, copay, relationship to subscriber, and employment group (Table 2).

A logistic model for somatization status was built on a derivation set of a randomly selected sample of 588 patients. We defined *somatization potential* as the percent of visits for which ICD-9 primary diagnosis groups fell into one of the following four: nervous system (codes 320–389), gastrointestinal system (codes 520–579), musculoskeletal system (codes 710–739), and ill-defined complaints (codes 780–799). Thus, all patients within these coding ranges, including those with organic diseases, were classed as having somatization potential. Somatization potential and total number of visits were treated as continuous variables as well as in multilevel categories. However, cross-classification of these factors with somatization status and gender revealed cell counts that were too small that resulted in the logistic model being unstable. Therefore, we used both somatization potential and number of visits as continuous variables in our model. The final logistic model was derived through a series of univariable models. To allow interactions of variables that otherwise might not exhibit significant association with the outcome, all variables that demonstrated a $P < 0.15$ relationship with somatization in a univariate analysis were included in the initial multivariable model. We then used a step-down approach to retain only variables that revealed a strong independent association ($P < 0.05$) with somatization. Selected interaction terms between predictor variables and variables excluded at the preliminary model building stage were then assessed for inclusion, but none were found to materially affect the predictive power of the final model. The final model was submitted to a series of influential diagnostics. We also assessed the model's goodness-of-fit by deviance statistics and the Hosmer-Lemeshow test.[27] A pseudo R^2-statistic[28] was calculated for the final model, which in the context of logistic regression models represents the proportion of explained variation of the model relative to the maximum achievable value of R^2.

The accuracy of the model based on MIS screening to predict somatization status was determined by assessing its agreement with the chart review diagnosis using sensitivity (proportion of true positives) and specificity (proportion of true negatives) computed from the likelihood of somatization as predicted by the logistic model.[24,29] To reduce bias in assessing these probabilities a modified jackknife procedure was used. We also computed positive and negative predictive values for a screening test based on the 14% prevalence of somatization among our high utilizing population.[24,29,30] The positive predictive value was the likelihood of somatization when the screener was positive, and the negative predictive value was the likelihood of not being a somatizer when the screener was negative. To depict the relationship between the proportion of true positives (sensitivity) and the proportion of false positives (one

minus specificity) we plotted the receiver operating characteristic curve (ROC curve). The greater the area under the ROC curve, measured quantitatively by the c-statistic, the greater the power of our model in discriminating between somatizers and nonsomatizers.[24,29,30]

We validated the derivation set model using the remaining one-third sample of the 883 patients. Sensitivity and specificity were calculated based on the estimated likelihood of somatization from the derived model and a ROC curve was plotted for the validation data set.

Results

The demographic characteristics of all 883 patients in our study (122 somatizers, 761 nonsomatizers) are shown in Table 2 (column 3). The mean age was 40.3 years (range 20–55 years) and 67% were women. The three main employers in the area, Michigan State University, General Motors, and the State of Michigan, employed almost 60% of our participants. Approximately 63% were direct subscribers to the health insurance, whereas 34% received their benefits as covered spouse. Because we sampled users who were 21 years or older, we had few who were covered as dependents of the subscriber.

Overall, the mean number of visits was 10.7 (median 9, interquartile range 7–13). The mean somatization potential was 35.6% (SD = 25.1%), the median 33.3% and interquartile range 14.3% to 53.8%. Comparison of the somatizers and nonsomatizers (Table 2, columns 4 and 5) revealed significant differences by gender ($P < 0.001$), in the number of visits ($P < 0.0001$), and in somatization potential ($P < 0.0001$). Somatizers were more likely to have higher somatization potential, more visits, and be of female gender. There were no significant differences between these groups in age, employer group, copay amount or relationship of the participant to the main subscriber (Table 2, columns 4–6).

Derivation Set

A randomly selected sample of approximately two-thirds of the patients (n = 588) formed the derivation set for constructing a logistic model for assessing the correlates of somatization status. As expected, the participant profile in the derivation set was similar to that of the whole sample.

Candidate variables for the final model were gender and, as continuous variables, the number of visits and somatization potential (percent of visits with the predicted ICD-9 codes). The final model contained as significant correlates of somatization, gender, total number of visits, and somatization potential (Table 3). A search for potential influencing observations and outlying values did not reveal any observations whose exclusion would substantively improve the predictive power. Overall, scaled pseudo $R^2 = 45\%$, and the c-statistic, equivalent to the area under the ROC curve, was 0.90.

On the basis of our model we computed the probability of somatization in a given patient and derived the sensitivity and specificity afforded by the model at different probability cutoff points. These are shown in Table 4 and the corresponding receiver operating characteristic (ROC) curve is the upper curve in Fig. 1. As highlighted, using a cutoff point of 0.04, the logistic model provided a sensitivity of 97.5% and a specificity of 54%. Based on the three predictor variables in Table 3, this means that a subject is predicted to be a somatizer if the model yields a probability of somatization $P \geq 0.04$. This probability P is determined from

$$\log\{P/(1-P)\} = -7.420 + 1.146 \times Gender + 0.166 \times NVisits + 0.057 \times SPotential$$

where *Gender* = 1 if female, and 0 otherwise, *NVisits* = Total number of visits, and *Spotential* = Somatic Potential.

Varying the cutoffs produces a range of sensitivities and specificities shown in Table 4. With a 14% prevalence of somatization, the corresponding predictive values positive and negative were 25.7% and 99.2% respectively. Using a cutoff of 0.40, the sensitivity was 49.4%, the specificity 95.9%, the predictive value positive 66.1%, and the predictive value negative 92.1%.

Validation Set

The model was prospectively validated in the remaining group of 295 patients. The validation set did not differ from the derivation set on somatization status ($P = 0.64$), gender ($P = 0.19$), age ($P = 0.65$), total number of visits ($P = 0.33$), and

TABLE 2. Characteristics of Patients (N = 883)

Characteristic	Subgroup	Total N (%)	N (%) by Somatizing Status: Somatizers (N = 122)	Nonsomatizers (N = 761)	*P* Value*
Age, years	20–29	118 (13.4)	16 (13.1)	102 (13.4)	0.339
	30–39	277 (31.4)	30 (25.0)	247 (32.5)	
	40–49	376 (42.6)	59 (48.4)	317 (41.7)	
	50+	112 (12.7)	17 (13.9)	95 (12.5)	
Gender	Male	283 (32.1)	21 (17.2)	262 (34.4)	<0.001
	Female	600 (67.0)	101 (82.8)	499 (65.6)	
Employer†	MSU	72 (8.2)	8 (6.6)	64 (8.4)	0.339
	State	208 (23.6)	32 (25.4)	177 (23.3)	
	GM	246 (27.9)	27 (22.1)	219 (28.8)	
	Other†	357 (40.4)	56 (45.9)	301 (39.6)	
Copay, $	0	487 (55.2)	67 (54.9)	420 (55.2)	0.598
	5	275 (31.1)	35 (28.7)	240 (32.5)	
	7	101 (11.4)	18 (14.8)	83 (10.9)	
	10	20 (2.3)	2 (1.6)	18 (2.4)	
Relationship to subscriber	Self	554 (62.7)	71 (58.2)	483 (63.5)	0.284
	Spouse	296 (33.5)	48 (39.3)	248 (32.6)	
	Dependent	33 (3.7)	3 (2.5)	30 (3.9)	
Number of visits	<8	301 (34.1)	11 (9.0)	290 (38.1)	<0.0001
	8+	583 (65.9)	111 (91.0)	471 (61.9)	
Somatic potential	No	658 (74.5)	48 (39.3)	610 (80.2)	<0.0001
	Yes	225 (25.5)	74 (60.7)	151 (19.8)	

*P value for comparison between somatizers and non somatizers.
†MSU = Michigan State University, State = State of Michigan, GM = General Motors.

somatization potential ($P = 0.09$). Univariate comparisons between somatizers and nonsomatizers in the validation set showed significant differences in gender ($P < 0.05$), number of visits ($P < 0.02$), and in somatization potential ($P < 0.01$).

Using the same probability cutoff of 0.04 to classify patients, the algorithm correctly identified 39 of the 43 somatizers (sensitivity = 0.91), and 129 of the 252 nonsomatizers (specificity = 0.51). With a prevalence of 14% this led to a positive predictive value of 23% and negative predictive value of 97%. The ROC curve for the validation set is the lower curve in Fig. 1; the c-statistic was 0.78 (Table 3).

Discussion

We developed a model that distinguished chronic somatizing patients from other high utilizing patients. We tested this prediction model on a second data set and validated its predictive ability.

This study exhibits several shortcomings, especially from our need to rely upon chart review to make gold standard diagnoses. 1) We depended upon how aggressively a doctor attempted to diagnose organic illness and upon how completely busy clinicians recorded their findings; with our study design, we were unable to obtain physicians' opinions but plan to do so in the future. 2) Without access to patients, we were unable to identify the patient's unique perspective or to diagnose specific

TABLE 3. Logistic Model for Somatization Status: Derivation Set (N = 588)

	β	χ^2	*P*
Intercept	−7.420	117.16	<0.0001
Female gender	1.146	9.66	0.0019
Number of visits	0.166	42.39	<0.0001
Somatization potential	0.057	66.03	<0.0001
c-statistic			
Derivation	0.90		
Validation	0.78		
Maximum R^2	0.45		
Hosmer-Lemeshow Test-*P* value	0.247		

TABLE 4. Sensitivity, Specificity, Predictive Value Positive, and Predictive Value Negative:* Derivation Data Set

Probability Cutoff	Sensitivity	Specificity	Predictive Value Positive	Predictive Value Negative
0.020	98.7	37.1	20.4	99.4
0.040	**97.5**	**54.0**	**25.7**	**99.2**
0.060	89.9	63.1	28.4	97.5
0.080	88.6	69.9	32.4	97.4
0.100	88.6	75.0	36.6	97.6
0.120	86.1	77.8	38.7	97.2
0.140	82.3	82.5	43.4	96.6
0.160	81.0	83.7	44.7	96.4
0.180	77.2	86.1	47.4	95.9
0.200	72.2	88.0	49.5	95.1
0.220	72.2	89.0	51.6	95.2
0.240	68.4	89.8	52.1	94.6
0.260	59.5	91.2	52.3	93.3
0.280	55.7	92.9	56.2	92.8
0.300	55.7	93.1	56.9	92.8
0.320	54.4	93.5	57.7	92.7
0.340	51.9	94.1	58.9	92.3
0.360	51.9	94.7	61.4	92.4
0.380	50.6	95.5	64.6	92.2
0.400	**49.4**	**95.9**	**66.1**	**92.1**
0.420	45.6	96.3	66.5	91.6
0.440	44.3	96.7	68.3	91.4
0.460	43.0	97.4	73.3	91.3
0.480	39.2	97.4	71.4	90.8
0.500	38.0	97.6	72.4	90.6

*Predictive values computed at a 14% prevalence of somatization.

somatoform disorders[1] using, for example, the Diagnostic Interview Schedule,[31] the World Health Organization Composite Diagnostic Interview,[32] or direct patient interviews. 3) Our designation of somatization depended on identifying symptoms of 6 or more months duration, the minimal criteria specified for DSM-IV somatoform diagnoses.[1] If busy clinicians did not record duration, some somatization may have been missed in chart review. 4) We cannot be certain that either somatizing or minor acute patients did not in fact have organic diseases because we were unable to investigate each patient ourselves from a biomedical perspective. We guarded against this problem by requiring that definitive diagnostic evaluation have been performed before making a diagnosis of somatization. During follow-up chart review (for other purposes), we have observed no instances of an important organic disease having been missed when a diagnosis of somatization was made in its place initially. 5) We used our chart data in a categorical, mutually exclusive way to reflect an overall 1-year diagnosis so that we were unable to account for the significant overlap among categories that we observed. This, of course, reduced the information we gained, and we presently are developing a more refined rating procedure to accommodate this important problem, eg, patients with combined organic disease and somatization status. 6) We required a whole year of data before we could identify a patient as somatization, and we know only how well the model predicts somatization during the year rated. We do not know how well it predicts for the subsequent year.

In spite of problems expected when chart review is the only available source of diagnostic data,

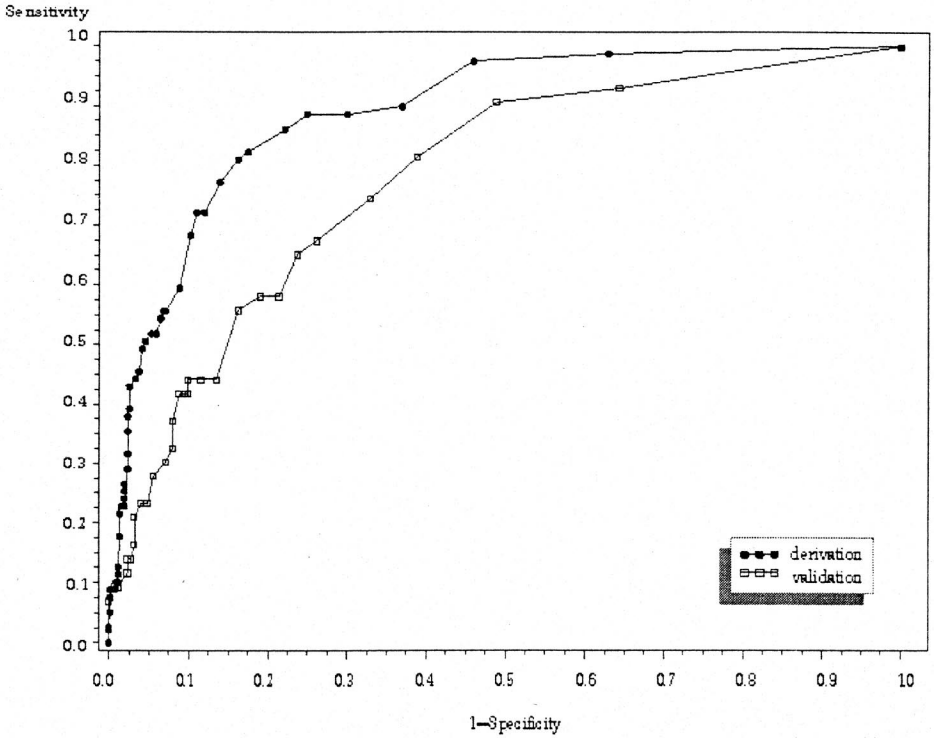

FIG. 1. Receiver operating characteristic curves.

this research provides a starting point on a screening procedure not requiring clinicians' time for a common and important problem. Our observations in somatizing patients did in fact indicate that we had studied a typical chronic somatizing population. Table 5 shows the breakdown of symptoms and diagnoses identified by the rater for the 122 somatizers. The commonest problems were low back pain, muscle/joint pain, headaches, and irritable bowel syndrome—with approximately half of all symptoms and diagnoses in the musculoskeletal system.

In addition, others have found that chart review itself may be superior to the Diagnostic and Statistical Manual of Mental Disorders, 4th edition, (DSM-IV)[1] in identifying somatizing patients[33,34] and that, indeed, DSM-IV has important shortcomings.[35,36] Nonetheless, DSM-IV diagnoses should be sought to better understand these patients. We suspect that many patients we rated as somatizers will be in the less severe abridged somatization disorder category,[13] rather than full-fledged somatization disorder.[1] Although DSM-IV diagnoses and other psychiatric measures are needed to better define this high utilizing population, their absence in no way negates the importance of our findings from medical patients' charts—which presumably reflect what patients actually reported to providers as their major reasons for seeking care. Rather, complementing our chart-based data with DSM-IV diagnoses and other psychiatric measures is the necessary next-step.

There are additional caveats about this study. 1) The study focused on an HMO population, which may be different from other populations, and this study is based upon data from 1995, which may be different from 2001. We also do not know if other HMOs have similar capabilities on their MIS. 2) We were somewhat concerned about the decrease in the explanatory power of the validation model, the c-statistic falling to 0.78 from 0.90 in the derivation model. In discussing this with experts, we believe that the disparity reflects an inordinately high c-statistic from the derivation model and that the 0.78 value on the validation model is acceptable. Also concerning the high values for the c-statistic, the rater was blinded and did not use

TABLE 5. Symptoms and Diagnoses of 122 Somatizing Patients*

	N	Diagnoses Recorded
Pain Symptoms		
Low back pain	29	Back pain, Sciatica, Lumbago
Joint/muscular pain	20	Fibromyalgia, Myofascial Syndrome
Headache	17	Migraine, Tension Headache, Cephalgia
Neck pain	9	Cervicalgia, Neck Pain
Abdominal pain	8	Irritable Bowel Syndrome†
Chest pain	8	Costochrondritis
Pelvic pain	4	Pelvic Pain
Miscellaneous pain	7	Hip, wrist, shoulder, arm, mouth, foot pain
Total	**102**	
Nonpain Symptoms		
Diarrhea/constipation	9	Irritable Bowel Syndrome†
Vertigo	4	Dizziness
Fatigue	2	Chronic Fatigue Syndrome
Nausea	2	Chronic Nausea
Visual	1	Visual changes
Impotence	1	Impotence
Weak knee	1	Exaggerated organic disease
Total	**20**	

*Symptom counts by body system follow: Musculoskeletal = 65, Nervous System = 22, Gastrointestinal System = 19, Cardiovascular System = 8, Genitourinary System = 8.
†Combined number of cases of Irritable Bowel Syndrome is 17.

criteria contained in the model (gender, number of visits, specific diagnoses) to identify patients for study and, rather, relied entirely upon criteria to exclude organic diseases. 3) Although it is likely that we excluded significant numbers of somatizers from study, our exclusion criteria were determined by our needs to identify a population of chronic somatizers for a management intervention. We thus identified a group of patients who were unlikely to be under specific psychological management but who were in need of treatment.[9,37] 4) We did not consider patients with fewer than 6 visits per year. Although this population may also contain somatizing patients, we believed high utilizing patients would be of greater concern to physicians and those concerned with costs of health care. 5) As a preliminary study, these data need to be confirmed in other HMOs. Similarly, a broader range of variables will require evaluation as possible predictors; eg, specific diagnosis codes within each broad ICD-9 category and other combinations of diagnosis codes.

This study was designed to assist clinicians and health plan directors. If corroborated, these data suggest that if HMOs wish to identify most high utilizing somatizers, we would recommend a cutoff of 0.04. This would require subsequent chart review to exclude large numbers of false positives. Alternatively, physicians could identify false positive patients during clinical assessments. If chart review was not possible and the HMO did not wish to use physicians' time to exclude false positives, a cutoff point of 0.40 would identify about half of chronic somatizers and the low false positive rate would preclude the need for chart review; physicians could identify the few false positives during clinical assessment. From the raw data alone, we also know that the group of patients with at least one visit in any of the four somatization potential categories contained all 122 somatizers, 86 of whom had a somatization potential of at least 50%, and 650 nonsomatizers, of whom 201 had a somatization potential of at least 50%. Thus, at present, more research is needed to refine what

could be a useful and simple procedure for identifying the often troublesome chronic somatizing patient. At this point, however, these data are mainly applicable to researchers and more work is needed. Future work will be most useful if it provides simple prediction rules and informs how they affect the probabilities of somatization, perhaps weighting predictors and even generating a simple score. Operationally, a system likely would build a program into the computer that entered gender, number of visits, and percent visits with one or more critical ICD-9 codes and, then, provided probabilities of somatization for providers and the health care plan.

Previous attempts to develop screeners for somatization have been made using self-report questionnaires[21] and interviews,[22] but the PrimeMD has gained most prominence.[38–40] With a different methodology from ours for defining somatization, using a structured interview, and by relying on physicians to decide if a symptom was unexplained, we must be cautious in comparing the PrimeMD to our screener. The PrimeMD takes the physician more than 10 minutes per patient to make a psychiatric diagnosis, and it has used cutoffs of three symptoms[38] and seven symptoms[38,40] to effectively achieve its goals. If our preliminary data can be corroborated, the MIS screener reported here has the potential attributes of being less expensive and placing no demands upon clinicians' time. Although its potential has been recognized,[20] we are aware of no previous studies screening for somatizing patients from computerized databases.

Others agree that an effective screening procedure for somatizing patients is needed.[7,21,40] Effective screening encourages a focus on management[20,40] which, at present, is not well established or research-based in primary care.[41,42] Successful RCTs for treating depression in primary care, however, provide hope and guidance.[43] Nonetheless, useful guidelines for treating somatization exist,[8,9,37] which can be used in developing and testing effective management programs.

Acknowledgments

This work was supported by a grant from The Institute for Managed Care, Michigan State University, East Lansing, MI. We also thank Sharon Roble, an honors medical student, for her hard, creative work.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association, 1994.

2. **Kirmayer LJ, Robbins JM.** Three forms of somatization in primary care: prevalence, co-occurrence, and sociodemographic characteristics. J Nerv Ment Dis 1991;179:647–655.

3. **Lipowski ZJ.** Somatization: the concept and its clinical application. Am J Psychiatry 1988;145:1358–1368.

4. **deGruy F, Columbia L, Dickinson P.** Somatization disorder in a family practice. J Fam Pract 1987;25:45–51.

5. **deGruy F, Crider J, Hashimi DK, et al.** Somatization disorder in a university hospital. J Fam Pract 1987;25:579–584.

6. **Bridges KW, Goldberg DP.** Somatic presentation of DSM III psychiatric disorders in primary care. J Psychosom Res 1985;29:563–569.

7. **Barsky AJ, Borus JF.** Somatization and medicalization in the era of managed care. JAMA 1995;274:1931–1934.

8. **Ford CV.** The somatizing disorders: illness as a way of life. New York, NY: Elsevier Biomedical; 1983.

9. **Smith RC** A clinical approach to the somatizing patient. J Fam Pract 1985;21:294–301.

10. **Wickramasekera I.** Somatizers the health care system, and collapsing the psychological distance that the somatizer has to travel for help. Professional Psychology: Research and Practice 1989;20:105–111.

11. **Zoccolillo MS, Cloninger CR.** Excess medical care of women with somatization disorder. South Med J 1986;79:532–535.

12. **Lightfoot JRW, Luft BJ, Rahn DW, et al.** Empiric parenteral antibiotic treatment of patients with fibromyalgia and fatigue and a positive serologic result for Lyme disease. Ann Intern Med 1993;119:503–509.

13. **Escobar JI, Swartz M, Rubio-Stipec M, et al.** Medically unexplained symptoms: distribution, risk factors, and comorbidity. In: Kirmayer LJ, Robbins JM, eds. Current Concepts of Somatization: Research and Clinical Perspectives. Washington, DC: American Psychiatric Press; 1991:63–78.

14. **Smith GR Jr, Monson RA, Ray DC.** Patients with multiple unexplained symptoms. Arch Intern Med 1986;146:69–72.

15. **Shaw J, Creed F.** The cost of somatization. J Psychosom Res 1991;35:307–312.

16. **Simon GE.** Psychiatric disorder and functional somatic symptoms as predictors of health care use. Psychiatr Med 1992;10:49–59.

17. **Lloyd AR, Pender H.** The economic impact of chronic fatigue syndrome. Med J Aust 1992;157:599–601.

18. **Katon W, von Korff M, Lin E, et al.** Distressed high utilizers of medical care: DSM-III-R diagnoses and treatment needs. Gen Hosp Psychiatry 1990;12:355–362.

19. **Smith GR Jr.** The course of somatization and its effects on utilization of health care resources. Psychosomatics 1994;35:263–267.

20. **Bass C, Bond A, Gill D, et al.** Frequent attenders without organic disease in a gastroenterology clinic–patient characteristics and health care use. Gen Hosp Psych 1999;21:30–38.

21. **Peveler R, Kilkenny L, Kinmonth A-L.** Medically unexplained physical symptoms in primary care: a comparison of self-report screening questionnaires and clinical opinion. J Psychosom Res 1997;42:245–252.

22. **Bucholz KK, Dinwiddie SH, Reich T, et al.** Comparison of screening proposals for somatization disorder empirical analysis. Compr Psychiatry 1993; 34:59–64.

23. **Weiner JP, Starfield BH, Steinwachs DM, et al.** Development and application of a population-oriented measure of ambulatory care case-mix. Med Care 1991;29:452–472.

24. **Sackett DL, Haynes RB, Tugwell P.** Clinical epidemiology: a basic science for clinical medicine. Boston, MA: Little, Brown and Co.; 1985.

25. **Humes HD, DuPont HL, Gardner LB, et al, eds.** Kelley's Textbook of Internal Medicine. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2000.

26. **Aaron LA, Burke MM, Buchwald D.** Overlapping conditions among patients with chronic fatigue syndrome, fibromyalgia, and temporomandibular disorder. Arch Intern Med 2000;160:221–227.

27. **Hosmer DW, Lemeshow S.** Applied Logistic Regression. New York, NY: John Wiley & Sons; 1989.

28. **Nagelkerke NJD.** A note on a general definition of the coefficient of determination. Biometrika 1991;78:691–692.

29. **Sackett DL, Richardson WS, Rosenberg W, et al.** Evidence-based Medicine—How to practice and teach EBM. New York, NY: Churchill Livingstone; 1997.

30. **Fletcher RH, Fletcher SW, Wagner EH.** Clinical Epidemiology–The Essentials. 3rd ed. Philadelphia: Williams & Wilkins; 1996.

31. **Robins LN, Helzer JE, Croughan J, et al.** National Insitute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. Arch Gen Psychiatry 1981;38:381–389.

32. **Sartorius N.** Composite International Diagnostic Interview (CIDI)—Core Version 1.1. Copyright World Health Organization.

33. **Fink P.** Physical complaints and symptoms of somatizing patients. J Psychosom Res 1992;36:125–136.

34. **Simon GE, Gureje O.** Stability of somatization disorder and somatization symptoms among primary care patients. Arch Gen Psychiatry 1999;56:90–95.

35. **Norquist G, Hyman SE.** Advances in understanding and treating mental illness: implications for policy. Health Aff 1999;18:32–47.

36. **Krueger RF.** The structure of common mental disorders. Arch Gen Psychiatry 1999;56:921–926.

37. **Smith RC.** Somatization disorder: defining its role in clinical medicine. J Gen Intern Med 1991;6:168–175.

38. **Spitzer RL, Williams JBW, Kroenke K, et al.** Utility of a new procedure for diagnosing mental disorders in primary care—the PRIME-MD Study. JAMA 1994;272:1749–1756.

39. **Spitzer RL, Kroenke K, Linzer M, et al.** Health-related quality of life in primary care patients with mental disorders—results from the PRIME-MD Study. JAMA 1995;274:1511–1517.

40. **Kroenke K, Spitzer RL, deGruy FV, et al.** A symptom checklist to screen for somatoform disorders in primary care. Psychosomatics 1998;39:263–272.

41. **Andrews G, Hadzi-Pavlovic D, Christensen H, et al.** Views of practicing psychiatrists on the treatment of anxiety and somatoform disorders. Am J Psychiatry 1987;144:1331–1334.

42. **Smith GR Jr.** Effectiveness of treatment for somatoform disorder patients. Psychiatric Med 1991;9:545–558.

43. **Katzelnick DJ, Simon GE, Pearson SD, et al.** Randomized trial of a depression management program in high utilizers of medical care. Arch Fam Med 2000;9:345–351.