

Research Paper ■

Using Computer-based Medical Records to Predict Mortality Risk for Inner-city Patients with Reactive Airways Disease

WILLIAM M. TIERNEY, MD, MICHAEL D. MURRAY, PHARM D, DENISE L. GASKINS, BS, XIAO-HUA ZHOU, PHD

Abstract **Objective:** To use routine data from a comprehensive electronic medical record system to predict death among patients with reactive airways disease.

Design: Retrospective cohort study conducted in an academic primary care internal medicine practice. Subjects were 1,536 adults with reactive airways disease: 542 with asthma and 994 with chronic obstructive pulmonary disease (COPD).

Measurements: The dependent variable was death from any cause within 3 years following patients' first primary care appointment in 1992. Multivariable logistic regression was used to identify independent predictors of 3-year mortality, with half of the patients used to derive the predictive model and the other half used to assess its predictability.

Results: Of the 1,536 study patients, 191 (12%) died in the 3-year follow-up period. From information available on or before patients' first primary care visit in 1992, multivariable predictors of 3-year mortality were coincidental heart failure, male sex, presence of COPD, lower weight, low serum albumin concentration level, and a prior arterial PO₂ of less than 60 mmHg; use of an inhaled corticosteroid was protective. The c-statistic (ROC curve area) in the validation cohort was 0.76, indicating good discrimination, and goodness of fit was excellent by Hosmer-Lemeshow chi-square ($P > 0.5$). Only 24% of the patients in the validation cohort were designated at high risk (estimated $\geq 15\%$ 3-year mortality), but this group contained more than half of the deaths within 3 years for the entire cohort.

Conclusions: Data generated during routine care and stored in a comprehensive electronic medical record can accurately predict mortality among patients with reactive airways disease. Such technology can be used by practices to control for severity of illness when assessing clinical practice and to identify high-risk patients for interventions to improve prognosis.

■ *J Am Med Inform Assoc.* 1997;4:313-321.

In recent years, there has been a disturbing trend toward increasing prevalence and severity of reactive airways disease¹⁻⁷ and associated mortality,^{1,4,5,8-9} especially among inner-city residents.⁹⁻¹¹ If patients with reactive airways diseases at higher mortality risk could be identified, interventions to reduce this risk

could be made more cost effective, targeting those most likely to benefit.

Predicting adverse outcomes depends on having a source of predictive information. One could prospectively collect such information by repeatedly exam-

Affiliations of the authors: Wishard Memorial Hospital (WMT, MDM); Regenstrief Institute for Health Care (WMT, MDM, DLG, XHZ); Indiana University School of Medicine (WMT, XHZ); Richard L. Roudebush Veterans Affairs Medical Center (WMT, MDM); Purdue School of Pharmacy (MDM), Indianapolis, IN.

Correspondence and reprints: William M. Tierney, MD, Regenstrief Institute for Health Care, Sixth Floor, RHC, 1001 West Tenth Street, Indianapolis, IN 46202.
E-mail: btierney@vax1.iupui.edu

Received for publication: 1/17/97; accepted for publication: 2/24/97.

ining and testing patients with reactive airways disease, but this would be expensive and time consuming. Routine clinical information, gathered during the process of delivering care, may predict utilization^{12,13} and adverse outcomes,¹⁴⁻¹⁷ yet such information is not usually accessible. Comprehensive electronic medical record systems could be one source for such predictive information,¹⁸ yet such systems are not currently widely available. If data in such systems could be shown to aid the delivery of cost-effective preventive care, this would be one more benefit to weigh against the cost of establishing and maintaining such systems.

Using a state-of-the-art comprehensive electronic medical information system,¹⁹ we tested the hypothesis that routine clinical data obtained from electronic medical records can accurately predict morbid outcomes in adults with reactive airways disease.

Methods

This study was performed in the General Medicine Practice (GMP), an academic primary care general internal medicine practice affiliated with an inner-city teaching hospital.²⁰ Approximately 13,000 patients visit this practice more than 50,000 times per year and are cared for by approximately 150 physicians, two thirds of whom are residents (predominantly in categorical internal medicine and medicine-pediatrics programs). This health care system is served by the Regenstrief Medical Record System (RMRS),²¹ a comprehensive information system that handles all laboratory, pharmacy, and appointment information for a network of inner-city facilities, including a 340-bed hospital, an adjacent outpatient facility with more than 65 offices and clinics, and more than 20 primary care neighborhood health centers and public health clinics.^{19,21}

The RMRS contains more than 150 million observations on more than 1.5 million separate patients cared for over the past 20 years. All data are stored permanently, with the exception of repeated inpatient clinical laboratory tests, for which the RMRS stores the first test value obtained upon admission, the last value before discharge, and the highest and lowest value for each week of the hospital stay. Death data are obtained from discharge case-abstracts, dictated death summaries, and the Indiana State Department of Health's computerized death certificate files. Names of patients in the death certificate files are matched with names in the RMRS using a validated matching algorithm.²²

Patients eligible for this study were those more than 14 years old who visited the GMP at least once in 1992

and had any one or more of the following indicators of reactive airways disease: any diagnosis of asthma, chronic obstructive pulmonary disease (COPD), emphysema, or chronic bronchitis recorded at any inpatient, emergent, urgent, or outpatient visit; any chest radiograph read as showing COPD; any prior prescription for inhaled beta-adrenergic agonists, cromolyn, parasympathetic blockers, or corticosteroids; or any prescription for oral beta-adrenergic agonists or theophylline. This definition was purposefully broad to be more sensitive (rather than specific) for patients with reactive airways disease in order to include a full spectrum of such patients.

For descriptive purposes, we arbitrarily assigned the diagnosis of "asthma" to patients who had ever had that specific diagnosis recorded from any site of care (and had never had the diagnosis of COPD recorded from any site) and to patients who had never had either specific diagnosis but had first taken one of the above-mentioned medications before 45 years of age. All other patients were deemed to have COPD. (We recognize that patients who never smoked should be considered asthmatic regardless of their age of onset of reactive airways disease. However, because our smoking information is not complete, and because of the blurring of the boundaries between asthma and chronic airways obstruction, we opted to arbitrarily separate these two groups of patients on the basis of their diagnoses and age at first treatment.)

We excluded patients who had never visited the GMP before their first visit in 1992 (who had too few data to analyze) as well as patients not known to have died within 3 years and who never returned to any inpatient, emergent, urgent, or outpatient site after their first GMP visit in 1992. All remaining eligible patients were included in the analysis.

We then extracted potential predictor variables from data stored in patients' electronic medical records on or before their first GMP visits in 1992. These variables fell into the following general categories: demographic data, vital signs, diagnoses from clinical and billing problem lists, diagnostic test results (clinical laboratory test results as well as coded results for imaging studies), drugs prescribed from any outpatient, emergency, or urgent care site, and clinical activity (visits to outpatient, emergent, urgent, or inpatient sites both before and after the first diagnosis of reactive airways disease). Because of the high prevalence of comorbid disease in this population, we assessed utilization (e.g., days hospitalized) both before and after the first evidence of reactive airways disease.

The dependent variable of interest was death from any cause after study patients' first GMP visit in 1992.

We chose to study all-cause mortality because of the known inaccuracy of causal information on death certificates^{23,24} and the possibility that deaths attributed to infectious or cardiac diseases may have had reactive airways disease as an inciting or contributing cause. Moreover, to patients, their families, and their physicians, the fact of death is more important than the immediate cause of death.

Using a random number generator in the Statistical Analysis System (SAS Institute, Cary, NC), we randomly selected half of our cohort for deriving the predictive model (the derivation cohort). The predictive power of this model was then assessed using the other half of the patients (the validation cohort). Because of the large number of potential predictor variables, we employed a two-stage data reduction process. The predictive power of individual variables among the derivation cohort was assessed using univariable logistic regression. All univariable significant predictors ($p < 0.10$) were then entered into multivariable logistic regression analysis (with backward removal of non-significant predictors). Selker²⁵ has recommended that the number of potential predictor variables should not exceed one tenth of the number of outcomes being predicted to avoid overfitting the model to the data. However, instead of arbitrarily excluding univariable significant predictors, we opted instead to only report the model's performance in the validation cohort, using the *c*-statistic (a measure of discrimination equal to the area under the Receiver Operating Characteristic, or ROC, curve)²⁶ and the Hosmer–Lemeshow chi-square statistic (a measure of model calibration or goodness-of-fit). Consistent with our prior modeling efforts,^{14,27} before we began the analyses we established a *c*-statistic (ROC curve area) of ≥ 0.70 as acceptable discrimination²⁸ and a Hosmer–Lemeshow chi-square *p*-value of ≥ 0.10 as acceptable calibration.

The resulting logistic regression model can be used to calculate a probability of dying in 3 years for each patient by using the following standard formula:

$$P = 1 / (1 + e^{-(B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n)})$$

where *P* is the calculated probability, B_0 is the intercept (constant), B_n is the nonstandardized parameter estimate for the *n*th variable, and X_n is the raw value of the *n*th predictor variable. All indicator variables (e.g., diagnoses, findings on imaging studies) were coded as 1 if they were found in a patient's electronic medical record on or before the patient's first GMP visit in 1992; otherwise, they were coded as 0. (This means that if a condition or finding was not mentioned in the patient's record or the imaging test had not been performed, the condition was assumed to be

absent.) Therefore, there were no missing data for these variables. Counting variables (e.g., number of prior emergency department visits) were likewise coded as 0 if none was noted in a patient's electronic record. Thus, the only variables that were potentially missing were clinical laboratory tests and vital signs. As in our previous modeling efforts,^{14,27} we arbitrarily excluded from analysis all such variables missing in more than 15% of subjects. We then dealt with missing data for the remaining variables by substituting the sample mean for the missing values. We repeated the analysis by excluding any patient whose record was missing any univariable significant laboratory test or vital sign. In a few instances, most notably arterial blood gases, we assumed that a patient who never had the test performed had normal results. We then created indicator variables for patients who had significant prior abnormalities on these tests (e.g., an arterial $pH < 7.30$ or an arterial $Po_2 < 60$ mmHg).

Results

In 1992, 1,557 patients visited the GMP who had prior evidence of reactive airways disease and had visited the GMP at least once before. Of these patients, 21 (1.3%) were excluded from analysis because they had no subsequent inpatient, emergent, urgent, or outpatient encounters and were not known to have died. Of the 1,536 remaining patients, 542 (35%) had asthma, while the other 994 (65%) had COPD by our definitions. The mean length of time that these 1,536 patients had received care in the GMP before their first visit in 1992 was 8.1 ± 5.2 (SD) years (range 0–16). The length of time followed in any facilities served by the RMRS prior to 1992 was 9.6 ± 5.2 years (range 0–16.5). The total time that these 1,536 patients were treated for reactive airways disease in this system was 6.9 ± 4.6 years (range 0–17).

By 3 years after their first GMP visits in 1992, 191 patients (12% of the entire study cohort) had died. Table 1 compares the baseline characteristics (i.e., data available on subjects' first GMP visits in 1992) of patients who died with those who survived. As expected, patients who died were significantly older, more often had COPD, less often had asthma, and had a greater burden of comorbid disease. There was more current treatment with ipratropium and home oxygen by patients who died but no difference in inhaled and oral corticosteroids, beta-adrenergic agonists, theophylline, or cromolyn. Patients who died had significantly lower weight and serum albumin concentrations than patients who survived. Patients who died had more prior hospitalizations and outpatient visits but no difference in emergency department visits than patients who survived. Hospitalizations specifically

Table 1 ■

Characteristics of Study Patients Who Died in 3 Years Compared With Those Who Survived

Patient Characteristic	Patients who Died (n = 191)	Patients who Survived (n = 1345)	P-value
Age (years)*	63.8 ± 12.8†	52.3 ± 16.1	<0.001
Sex (% women)	49.7‡	62.5	<0.001
Race (% white)	52.9	54.9	NS
Asthma	22.0	37.2	<0.001
Chronic obstructive pulmonary disease	88.0	60.2	<0.001
Hypertension	86.4	68.8	<0.001
Heart failure	50.8	19.3	<0.001
Ischemic heart disease	51.8	24.2	<0.001
Diabetes mellitus	47.6	32.1	<0.001
Obesity	24.6	25.4	NS
Bronchitis history	53.9	58.1	NS
Dyspnea history	16.7	13.2	NS
Esophageal reflux history	12.0	13.8	NS
Prior nonobstetric hospitalizations (no.)	4.0 ± 4.2	2.3 ± 3.1	<0.001
Prior hospitalizations for reactive airways disease (no.)	1.8 ± 2.4	1.1 ± 2.4	<0.001
Prior pneumonia hospitalizations (no.)	0.53 ± 1.2	0.17 ± 0.54	<0.001
Days hospitalized before onset of reactive airways disease (no.)	9.7 ± 19	6.1 ± 14	0.013
Days hospitalized after onset of reactive airways disease (no.)	27.7 ± 40	13.9 ± 28	<0.001
Prior GMP§ visits after onset of reactive airways disease (no.)	32 ± 28	22 ± 24	<0.001
All prior GMP visits (no.)	41 ± 33	30 ± 28	<0.001
Prior pulmonary clinic visits (no.)	0.79 ± 2.4	0.57 ± 2.5	NS
Prior emergency visits after onset of reactive airways disease (no.)	8.7 ± 11	8.3 ± 11	NS
Prior emergency department visits (no.)	11.8 ± 13	12.5 ± 14	NS
All prior system encounters (no.)	81 ± 63	65 ± 51	<0.001
Time followed in system (years)	10.4 ± 5.0	9.6 ± 5.2	0.027
Inhaled albuterol use¶	59.7	59.2	NS
Theophylline use	38.2	36.6	NS
Inhaled corticosteroid use	16.8	21.6	NS
Ipratropium use	22.0	15.5	0.002
Oral steroid use	20.9	16.8	NS
Cromolyn use	1.1	3.1	NS
Spacer device given ever	15.7	13.9	NS
Home oxygen therapy ever	19.9	6.6	<0.001
Theophylline level measured ever	53.9	50.2	NS
Most recent weight (lbs)	169 ± 59	185 ± 55	<0.001
Most recent systolic blood pressure (mmHg)	133 ± 22	131 ± 21	NS
Most recent diastolic blood pressure (mmHg)	79.8 ± 12	80 ± 11	NS
Most recent serum albumin (g/L)	3.7 ± .55	4.0 ± 45	<0.001
Prior blood gas measurements (no.)	8.3 ± 10	4.9 ± 9.0	<0.001
Prior blood counts (no.)	20 ± 178	15 ± 16	<0.001
Most recent hemoglobin (g/L)	138 ± 166	132 ± 205	>0.001
Most recent leukocyte count (no./mm ³)	9473 ± 9983	8913 ± 3227	NS
Any prior PO ₂ <60 mmHg	52.4	26.9	<0.001
Any prior PCO ₂ >50 mmHg	31.9	12.9	<0.001
Any prior PCO ₂ <30 mmHg	29.8	27.4	NS
Any prior pH <7.3	22.5	12.6	<0.001
Any prior FEV1 less than 50% of predicted	37.7	22.1	<0.001

*Age at the first primary care visit in 1992.

†±Standard deviation.

‡All figures are % of patients with that characteristic, unless otherwise specified, and all diagnoses come from problem lists and diagnostic tests.

§GMP = General Medicine Practice, site of this study.

¶All drugs are the percentage of patients who were currently taking them, or had them prescribed, at the time of their first GMP visit in 1992.

Table 2 ■

Multivariable Predictors of 3-Year Mortality in the Derivation Cohort

Predictor Variable*	Parameter Estimate	Chi-square	P-value	Multivariable Odds Ratio (95% CI)†
Diagnosis of heart failure	1.1213	18.20	0.0001	3.07 (1.83–5.14)
Sex = male	0.8205	10.68	0.0011	2.72 (1.39–3.72)
Diagnosis of chronic obstructive pulmonary disease‡	1.1921	8.95	0.0028	3.29 (1.51–7.19)
Most recent prior weight (in lbs)	–0.0060	6.10	0.0136	0.94§ (0.89–0.99)
Most recent prior serum albumin concentration (g/dL)	–0.6128	5.84	0.0157	0.54 (0.33–0.89)
Any prior PO ₂ less than 60 mmHg	0.6497	5.60	0.0197	1.92 (1.12–3.28)
Current treatment with inhaled corticosteroid	–0.8888	5.06	0.0244	0.41 (0.19–0.89)
Number of prior hospitalizations with pneumonia	0.2715	4.89	0.0270	1.31** (1.03–1.67)
Intercept	–0.4691	0.17	0.6827	—

*Data available on or before first General Medicine Practice (GMP) visit in 1992. Data are also coded as 1 if present, 0 if absent unless otherwise specified.

†95% CI = 95% confidence interval.

‡Diagnosis as recorded at any encounter or on any chest radiograph.

§Multivariable odds ratio (95% CI) for each 10 lbs of weight.

**Multivariable odds ratio (95% CI) for each prior hospitalization.

for reactive airways disease and pneumonia were also higher among patients who died. Finally, patients who died within 3 years more frequently had prior abnormal arterial blood gas values (PO₂ and pH) than patients alive 3 years after their first GMP visit in 1992.

There were 752 patients (49%) randomly assigned to the model derivation cohort, 93 (12%) of whom died within 3 years of their first GMP visit in 1992. Of the 90 potentially predictive variables assessed, 39 passed the univariable screen ($p < 0.10$) and were submitted to multivariable logistic regression analysis. Seventy-seven derivation cohort patients (10%) were missing one or more of the candidate predictor variables (predominantly laboratory test results). We performed the multivariable analysis with and without substituting the derivation sample mean for missing variables; the results were almost identical. We therefore present the results for the analysis using the entire dataset with mean substitution for missing values for candidate predictors.

Table 2 shows the eight variables that were retained in the final logistic regression model by virtue of their having a multivariable p -value less than 0.05. By far, the strongest predictor was the presence of heart failure (as evidenced by that diagnosis recorded during any encounter or by evidence of heart failure on car-

diac imaging studies). Patients with heart failure had three times the risk of dying in 3 years as did patients without heart failure. The next-strongest predictor was male sex (again, associated with an almost tripling of the risk of death in 3 years), followed by the diagnosis of COPD, lower weight, lower serum albumin concentration, having a prior PO₂ of less than 60 mmHg, and the number of prior hospitalizations for pneumonia. Current treatment with an inhaled corticosteroid (at the time of their first GMP visit in 1992) was associated with a 60% reduction in the risk of dying in 3 years.

The model's ability to predict 3-year mortality was assessed among the 784 patients (51%) who were randomly assigned to the model validation cohort, 98 (12%) of whom died within 3 years of their first GMP visit in 1992. The c -statistic (equivalent to area under the ROC curve) for the model was 0.76, which indicates good discrimination by our a priori criteria and those of others.²⁸ The Hosmer–Lemeshow goodness-of-fit chi-square statistic was 7.13, with 8 degrees of freedom ($p = .52$), which indicates excellent calibration. Figure 1 shows the calibration curve for the validation cohort, which also demonstrates excellent agreement between predicted and observed deaths and shows that the model does not routinely overestimate or underestimate 3-year mortality.

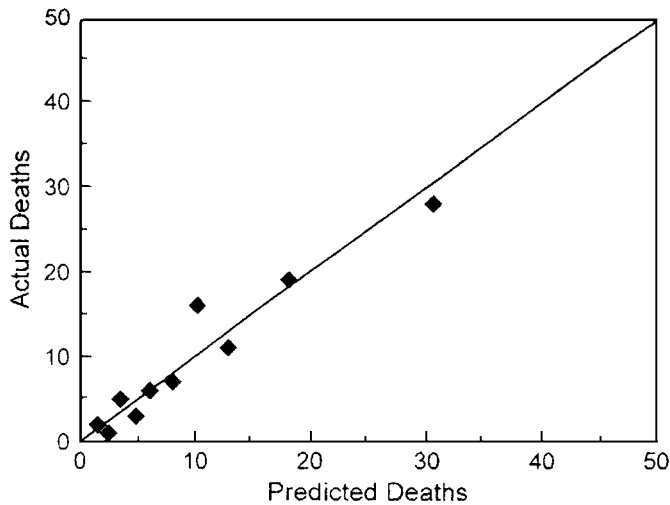


Figure 1 Calibration curve comparing predicted versus actual deaths in deciles of the model validation cohort, sorted by the logistic regression model's predicted probability of death.

As an example of how this model could be used to identify low- and high-risk subgroups of patients, we arbitrarily defined as high risk those patients in the validation cohort who had a calculated probability of dying in 3 years of more than 15% (Table 3). Of the 24% of patients in the validation cohort who would be defined as high risk using this criterion, 28.5% died within 3 years compared with only 7.5% mortality among the remaining 76% of validation set patients thus defined as low risk. The high-risk group contains more than half of all patients who died.

Discussion

This study demonstrates that information gathered during routine care by a state-of-the-art electronic medical record system can accurately predict 3-year mortality among patients with reactive airways diseases. If risk estimates based on such data were available for quality improvement activities, these activities could be made more cost effective by focusing on higher risk patients (Table 3). As health care resources become increasingly constrained, they and their associated costs can thus be allocated to those patients most likely to benefit from them. Using a predictive equation such as ours (Table 2), higher risk patients could, for example, be referred to pulmonary subspecialists, followed more closely by clinical nurse-specialists, or aggressively pursued for preventive care (such as influenza and pneumococcal vaccination).^{29,30} Primary care physicians and/or specialists might also treat comorbid conditions more aggressively or emphasize medication and visit compliance among higher risk patients.

Such predictive models can also be used to correct for severity of illness in evaluating the performance of physicians or other health care providers.³¹⁻³³ For example, assessments of our physicians' health care utilization for patients with reactive airways diseases could be adjusted for their patients' severity of illness, giving them "credit" for caring for sicker patients. With increasing frequency, health care is being scrutinized because of its high costs and the unexplained (and often unjustifiable) variability in the processes and outcomes of care observed when comparing physicians, practices, hospitals, and sometimes geographic areas.³⁴⁻³⁸ If providers are not given credit for caring for sicker patients, the higher risk of adverse events and greater costliness associated with these patients would be a disincentive for caring for them. Unfortunately, systems for adjusting for severity of illness often perform poorly in venues other than those in which they were developed.³⁹ Using one's own electronic medical records to adjust for illness severity avoids such problems with generalizability of others' predictive models.⁴⁰ However, adjusting for risk using such models is only valid within a health care system served by an electronic medical record. Comparisons between systems would be limited by the inherent differences in patients, data systems, and processes of providing care.

Predictive models can also provide clinicians with risk assessments for individual patients. For example, because all orders written in the GMP are entered by patients' primary care physicians directly into computer workstations,^{19,41} physicians writing outpatient orders can be shown the risk of each patient's dying in the next 3 years. As part of a larger trial of the effects of computer-based practice guidelines, we are currently using the model in Table 2 and another predicting mortality in patients with heart failure¹⁴ to provide mortality estimates during outpatient ordering for appropriate GMP patients.

Table 3 ■

Example of Using a Threshold of Predicted Probability of Death of >0.15 To Define Patients in the Validation Cohort as High Risk

Status	Predicted Probability of Death in 3 Years ≤0.15 (n = 598)	Predicted Probability of Death in 3 Years >0.15 (n = 186)
Actually alive in 3 years (n = 686)	553 (92%)	133 (72%)
Actually dead in 3 years (n = 98)	45 (8%)	53 (28%)

However, for such information to be useful, physicians must be able to understand and act upon quantitative probabilistic information. We believe they can. In a prior study of our emergency department patients with chest pain, treating physicians' numeric probability assessments were the strongest predictors of myocardial infarction.⁴² In another study performed using the order-writing workstations in the GMP, physicians shown low calculated probabilities that the tests they were ordering would be abnormal²⁷ ordered significantly fewer tests than did control physicians.⁴³

To have any effect on physician behavior, such predictive models must have face validity. Clearly, patients with heart failure alone are at increased risk of mortality.¹⁴ Their risk is even greater when combined with reactive airways diseases: 28% of validation cohort patients with heart failure died in 3 years compared with 8% of those without heart failure. Also, COPD (17% mortality in 3 years) was more morbid than asthma (8% mortality), which was also expected given the chronicity and irreversibility of physiologic dysfunction with COPD. It also makes sense that those with more severe pulmonary disease, as indicated by having a prior arterial PO₂ of less than 60 mmHg or prior hospitalizations with pneumonia, were at increased risk. Lower weight and serum albumin levels have been shown by us¹⁴ and others⁴⁴⁻⁴⁷ to be predictive of death. Although both factors can indicate general physical decline, chronic disease, or malnutrition, direct effects causing mortality have been postulated for hypoalbuminemia.⁴⁷⁻⁵¹ Men may have been at higher risk because, in this cohort, they were older and had a greater burden of comorbidity: they had 50% more coronary artery disease and COPD than women. They also had fewer prior visits for primary care, suggesting that men with reactive airways disease less often received care specifically aimed at forestalling morbidity. In a prior epidemiologic study using RMRS data, Murray et al.⁵² showed that, especially among adolescents and young adults, men were less likely to receive primary care, more likely to be cared for in the emergency department, and more likely to be hospitalized for reactive airways disease.

Finally, treatment with inhaled corticosteroids (in this case, predominantly beclomethasone) was independently associated with a marked reduction in 3-year mortality. However, further analyses of our data by type of reactive airways disease shows that the protective effect of inhaled corticosteroids is seen only among patients with asthma, not patients with COPD. This result is consistent with recent guidelines for treatment of asthma published by the National Insti-

tutes of Health, which reinforce the use of inhaled corticosteroids in what is now considered primarily an inflammatory condition.⁵³ This lends strong support for a computer reminder that encourages the use of inhaled corticosteroids among asthmatics.

Selker²⁵ has listed nine criteria upon which to assess predictive models. The model reported herein meets eight of them. The sole unmet criterion (not evaluating more than one independent variable for every ten patients suffering the outcome of interest—in this case, death in 3 years) is included by Selker to avoid overfitting the model to the data. However, this point is moot in our study because we have only reported our model's predictive power among patients in half of the patients who were excluded from the model-building analysis.

This study has limitations. First, the model shown in Table 2 may not be useful in other populations. These inner-city patients are not likely to be representative of most patients with reactive airways diseases. Moreover, the data come from the Regenstrief Medical Record System,¹⁹ which is unique, albeit a prototype of future electronic medical records.¹⁸ However, even though the model may not perform well in other populations, the methods we employed to derive our local model are certainly generalizable to practices with comprehensive electronic medical record systems. Our results should not only encourage those with sophisticated electronic medical record systems to profile their own patients' risk but also provide those lacking such systems with evidence that the value of these systems goes beyond generating bills and storing diagnostic test results.

This study also included some patients with very weak evidence of reactive airways disease. For example, a patient with one prior bout of acute bronchitis with bronchospasm treated with a single metered-dose inhaler would have been included in this study. However, those wishing to use electronic medical record systems to assess patient risk will also not have the luxury of verifying the prevalence of reactive airways disease in all patients with any evidence of its existence. If anything, including such patients is likely to bias the analysis *against* finding accurate predictors of mortality. We opted to include patients with *any* evidence of reactive airways disease rather than attempt to estimate its severity, which would be especially arbitrary among asthma patients, whose disease is by nature often intermittent and characterized by long intervening periods of normality.

We conclude that information stored in a state-of-the-art electronic medical record system can be used to accurately predict mortality among inner-city primary

care patients with reactive airways disease. Predictive information was found among demographic data (race), outpatient and inpatient diagnoses, arterial blood gas values and other diagnostic test results, and drug use data. As electronic medical record systems are installed into health care facilities or are created from existing electronic medical information resources,⁵⁴ those sufficiently comprehensive to include such data can (and should) be used to identify such patients at high mortality risk. Such prognostic models can be used to target treatment interventions and quantify severity of illness.

References ■

- Asthma—United States: 1980–1987. *MMWR*. 1990;39:493–7.
- Gergen PJ, Mullally DI, Evans R. National survey of prevalence of asthma among children in the United States, 1976 to 1980. *Pediatr*. 1988;81:1–6.
- Gerstman BB, Bosco LA, Tomita DK, Gross TP, Shaw MM. Prevalence and treatment of asthma in the Michigan Medicaid patient population younger than 45 years, 1980–1986. *J Allergy Clin Immunol*. 1992;83:1032–9.
- Mak H, Johnston P, Abbey H, Talamo RC. Prevalence of asthma and health service utilization of asthmatic children in an inner city. *J Allergy Clin Immunol*. 1982;70:367–72.
- Evans R, Mullally DI, Wilson RW, Gergen PJ, Tosenberg HM, Grauman JS. National trends in the morbidity and mortality of asthma in the US: prevalence, hospitalization, and death for asthma over two decades. *Chest*. 1987;91(Supplement 6):65S–74S.
- Halfon N, Newacheck PW. Trends in the hospitalization for acute childhood asthma, 1970–84. *Am J Public Health*. 1986;76:1308–11.
- Gerstman BB, Bosco LA, Tomita DK. Trends in the prevalence of asthma in the 5- to 14-year-old Michigan Medicaid population, 1980 to 1986. *J Allergy Clin Immunol*. 1993;91:838–43.
- Weiss KB, Wagener DK. Changing patterns of asthma mortality: Identifying target populations at high risk. *JAMA*. 1990;264:1683–7.
- Carr W, Zeitel L, Weiss K. Variations in asthma hospitalizations and deaths in New York City. *Am J Public Health*. 1992;82:59–67.
- Wissow LS, Gittelsohn AM, Szklo M, Starfield B, Mussman M. Poverty, race, and hospitalization for childhood asthma. *Am J Public Health*. 1988;78:777–82.
- Crain EF, Weiss KB, Bijur PE, Hersh M, Westbrook L, Stein REK. An estimate of the prevalence of asthma and wheezing among inner-city children. *Pediatrics*. 1994;94:356–62.
- Tierney WM, Fitzgerald JF, Miller ME, James MK, McDonald CJ. Predicting inpatient costs with admitting clinical data. *Med Care*. 1995;33:1–14.
- Herrmann FR, Safran C, Levkoff SE, Minaker KL. Serum albumin level on admission as a predictor of death, length of stay, and readmission. *Arch Intern Med*. 1992;152:125–30.
- Tierney WM, Takesue BY, Vargo DL, Zhou XH. Using electronic medical records to predict mortality in primary care patients with heart disease: prognostic power and pathophysiologic implications. *J Gen Intern Med*. 1996;11:83–91.
- Davis RB, Iezzoni LI, Phillips RS, Reiley P, Coffman GA, Safran C. Predicting in-hospital mortality: the importance of functional status information. *Med Care*. 1995;33:906–21.
- Evans RS, Burke JP, Classen DC, et al. Computerized identification of patients at high risk for hospital-acquired infection. *Am J Infect Control*. 1992;20:4–10.
- Evans RS, Classen DC, Stevens LE, et al. Using a hospital information system to assess the effects of adverse drug events. *Proc Annu Symp Comput Appl Med Care*. New York: McGraw-Hill, 1993;161–5.
- Institute of Medicine, Committee on Improving the Medical Record. *The Computer-based Patient Record: An Essential Technology for Health Care*. Washington, DC: National Academy Press, 1991.
- McDonald CJ, Tierney WM, Martin DK, Overhage JM. The Regenstrief Medical Record System: 20 years' experience in hospital outpatient clinics and neighborhood health centers. *MD Comput*. 1992;9:206–17.
- Tierney WM, Miller ME, Hui SL, McDonald CJ. Practice randomization and clinical research: the Indiana experience. *Med Care*. 1991;29:JS57–JS64.
- Overhage JM, McDonald CJ, Tierney WM. Design and implementation of the Indianapolis Network for Patient Care and Research. *MLA Bulletin*. 1995;83:48–56.
- Friedman C, Sideli R. Tolerating spelling errors during patient validation. *Comput Biomed Res*. 1992;25:486–509.
- Washko RM, Frieden TR. Tuberculosis surveillance using death certificate data, New York City, 1992. *Public Health Rep*. 1996;111:251–5.
- Slater DN. Certifying the cause of death: an audit of wording inaccuracies. *J Clin Pathol*. 1993;46:232–4.
- Selker HP. Systems for comparing actual and predicted mortality rates: characteristics to promote cooperation in improving hospital care. *Ann Intern Med*. 1993;118:821–3.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
- Tierney WM, Martin DK, Hui SL, McDonald CJ. Using clinical data to predict abnormal serum electrolytes and blood cell profiles. *J Gen Intern Med*. 1992;4:375–83.
- Weinstein MC, Feinberg HV. *Clinical Decision Analysis*. Philadelphia: WB Saunders, 1980;118.
- Canadian Thoracic Society Workshop Group. Guidelines for the assessment and management of chronic obstructive pulmonary disease. *Can Med Assoc J*. 1992;147:420–8.
- Rogers TK, Howard P. Pulmonary hemodynamics and physical training in patients with chronic obstructive pulmonary disease. *Chest*. 1992;101(suppl):289S–292S.
- Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. *Ann Intern Med*. 1995;123:763–70.
- Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. *Crit Care Med*. 1995;23:1327–35.
- Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory care groups: a categorization of diagnoses for research and management. *Health Serv Res*. 1991;26:53–74.
- Borowsky SJ, Kravitz RL, Laouri M, et al. Effect of physician specialty on use of necessary coronary angiography. *J Am Coll Cardiol*. 1995;26:1484–91.
- Carlisle DM, Valdez RB, Shapiro MF, Brook RH. Geographic variation in rates of selected surgical procedures within Los Angeles County. *Health Serv Res*. 1995;30:27–42.

36. Welch WP, Miller ME, Welch HG, Fisher ES, Wennberg JE. Geographic variation in expenditures for physicians' services in the United States. *N Engl J Med*. 1993;328:621-7.
37. Chassin MR. Explaining geographic variations: the enthusiasm hypothesis. *Med Care*. 1993(suppl);31:YS37-YS44.
38. Roos NP. Hospitalization style of physicians in Manitoba: the disturbing lack of logic in medical practice. *Health Serv Res*. 1992;27:361-84.
39. Iezzoni LI. "Black box" medical information systems: a technology needing assessment. *JAMA*. 1991;265:3006-7.
40. Tierney WM, McDonald CJ. Practice databases and their uses in clinical research. *Stat Med*. 1991;10:541-57.
41. Tierney WM, Miller ME, Overhage JM, McDonald CJ. Physician inpatient order-writing on microcomputer workstations: effects on resource utilization. *JAMA*. 1993;269:379-83.
42. Tierney WM, Fitzgerald J, McHenry R, et al. Physicians' estimates of the probability of myocardial infarction in emergency room patients with chest pain. *Med Decis Making*. 1986;6:12-17.
43. Tierney WM, McDonald CJ, Hui SL, Martin DK. Computer predictions of abnormal test results: effects on outpatient testing. *JAMA*. 1988;259:1194-8.
44. Phillips A, Shaper AG, Whincup PH. Association between serum albumin and mortality from cardiovascular disease, cancer, and other causes. *Lancet*. 1989;2:1434-6.
45. Kuller LH, Eichner JE, Orchard TJ, et al. The relation between serum albumin levels and risk of coronary heart disease in the Multiple Risk Factor Intervention Trial. *Am J Epidemiol*. 1991;134:1266-77.
46. Klonoff-Cohen H, Barrett-Connor EL, Edelstein SL. Albumin levels as a predictor of mortality in the healthy elderly. *J Clin Epidemiol*. 1992;45:207-12.
47. Corti M, Guralnik JM, Salive ME, Sorokin JD. Serum albumin level and physical disability as predictors of mortality in older persons. *JAMA*. 1994;272:1036-42.
48. Rossi EC. The effect of albumin upon the loss of enzymes from washed platelets. *J Clin Lab Med*. 1972;79:240-6.
49. Goodman DS. The interaction of human serum albumin with long-chain fatty acid anions. *J Am Chem Soc*. 1958;80:3892-8.
50. Hoak JC, Connor WE, Eckstein JW, Warner ED. Fatty acid-induced thrombosis and death: mechanisms and prevention. *J Lab Clin Med*. 1964;63:791-800.
51. Halliwell B. Albumin: an important extracellular antioxidant? *Biochem Pharmacol*. 1988;37:569-71.
52. Murray MD, Stang P, Tierney WM. Health care utilization by inner-city patients with asthma. *J Clin Epidemiol*. 1997;50:167-74.
53. National Asthma Education Program Expert Panel. Guidelines for the diagnosis and management of asthma. Bethesda, Maryland: National Institutes of Health, Publication No. 91-2042;1991.
54. McDonald CJ, Hammond WE. Standard formats for electronic transfer of clinical data. *Ann Intern Med*. 1992;110:333-5.