Health Outcomes Research in Medicine

ORIGINAL RESEARCH ARTICLE

HEALTH ■ CLINICAL ■ POLICY

# Accuracy of Claims-Based Measures of Severity of Childhood Illnesses

Farrokh Alemi, PhD, Maria Uriyo, PhD

Department of Health Systems Administration, Georgetown University, Washington, DC

## ABSTRACT

**BACKGROUND:** The use of electronic health records to conduct comparative effectiveness studies requires accurate measure of severity of patients' illness.

**OBJECTIVES:** This brief report provides data on relative accuracy of claims-based severity indices for childhood diseases.

**MEASURES:** We compared the accuracy of All Patient Refined Diagnosis-Related Groups (APR-DRG), All Payer Severity-adjusted Diagnosis-Related Groups (APS-DRG), Alemi and Walters Severity across Episodes of Illness, and count of diagnoses.

**METHODS:** The accuracy of each measure was calculated using the percent of deviance explained in mortality and percent of variation explained in length of stay (a surrogate measure of resource utilization).

**SUBJECTS:** Data were obtained from the 2006 Kid's Inpatient Database of the Healthcare Cost and Utilization Project of the Agency for Healthcare Research and Quality. We examined data on 3.1 million patients across 38 states.

**RESULTS:** Alemi and Walters' formula-based severity score explained 34% of variation in length of stay and 32% of variation in mortality. This index was more accurate than other indices, especially in predicting mortality, where it was 5-fold more accurate than APS-DRG and 3-fold more accurate than APR-DRG. The difference in accuracy was not only statistically significant but also large enough that it could change conclusions of comparative effectiveness studies.

**KEYWORDS:** Administrative data; Electronic health records; Length of stay; Pediatric illness; Resource utilization; Risk assessment

There is growing interest in conducting comparative effectiveness studies through electronic health records, where data on interventions and outcomes are readily available. An accurate measure of severity is a first step to conducting such studies. Data in electronic health records are not randomized, and more severely ill patients may be more likely to seek newer interventions that provide them with hope

*See funding, conflict of interest, and authorship disclosures at the end of this article.*

 

of relief. Inaccurate measurement of severity will inadvertently lead to the conclusion that the new interventions are associated with poorer outcomes, a case of blaming the fireman for the fire. This report is intended to help investigators be more informed about the relative accuracy of different claims-based severity indices for pediatric illnesses.

We contrast the accuracy of 4 different claims-based measures of severity of illness for children's diseases. Most claims-based measures of severity were developed for adult populations, and their utility in measuring outcome of children's illnesses has not been reported, although exceptions exist.[1]

Title IV of the Children's Health Insurance Program Reauthorization Act of 2009 required the establishment of a Pediatric Quality Measures Program. The Centers for Medicare and Medicaid Services and Agency for Healthcare Research and Quality have interpreted this requirement to mean the need to develop new measures of quality that are specific to children's conditions (eg, number of prenatal visits for pregnant patients or the annual number of emergency room visits for asthmatic children who are at least 1 year old and who have at least one asthma-related emergency department visit). Claims-based measures of severity of illness allow examination of risk-adjusted health care outcomes and therefore, can serve as an alternative to or in conjunction with the Pediatric Quality Measures.

# METHODS

## Methods of Measurement of Claims-based Severity

We examine the accuracy of 4 methods of predicting patient outcomes: All Patient Refined Diagnosis-Related Groups (APR-DRG);[2-5] All Payer Severity-adjusted Diagnosis-Related Groups (APS-DRG);[5,6] count of diagnoses;[7] and Alemi and Walters' Severity across Episodes of Illness.[8] The APR-DRG assigns 3 descriptors to each case: 1) the "base APR-DRG"; 2) the severity of illness class; and 3) the risk of mortality class. These determinations are based on the medical diagnostic group of the primary diagnoses, presence of specific secondary diagnoses, age, and selected operating and nonoperating room procedures. The APR-DRG is in use in both adult and pediatric populations.[9]

The second approach we examine is the APS-DRG. This approach classifies patients into groups that have homogenous resource use and outcomes. It uses diagnoses, procedures, and status at discharge to accomplish this task. More than 7 million patient records were examined to establish the APS-DRGs. A clinical team reviews results for logical consistency and reasonableness; when problems appear to exist because of small cell sizes, scores are imputed. The APS-DRG is in widespread use among payers for both adult and pediatric services.

The third approach is a crude measure of severity of illness based on count of the diagnoses. This measure of severity has proven more accurate than some of the other measures of prognosis, such as the Charlson Comorbidity Index, and can serve as a lower threshold for expected performance of severity indices.[8]

The fourth approach we examine is the Alemi and Walters severity index. This is a patented algorithm that assigns severity scores to a combination of diagnoses/procedure codes. In contrast to the APR-DRG or APS-DRG, no attempt is made to classify diagnoses into broad categories of disease. Each diagnostic code is scored based on its own properties, leading to approximately 14,000 variables to create the overall severity index. In the first step, codes are classified into primary and secondary or comorbidity codes. The primary diagnosis is the first diagnosis, which was the reason for admission. The remaining 4, 9, or 14 codes are the comorbidity codes during the admission. The average severity for "n" cases with primary diagnosis "p," comorbidity "c," and discharge status "d," is calculated using the following formula:

$$A_{p,c,d} = \frac{\sum_i^n \mathbf{A}_i}{n} \{i | p_i = p, c \in \mathbf{C}_i, D_i = d\}$$

Where $p_i$ is the primary diagnosis of patient "i," $D_i$ is the discharge status of patient "i," and $C_i$ is the set of comorbidities of patient "i."

Because the average resource use may exceed one; the calculated averages are standardized using the minimum and maximum functions:

$$S_{p,c,d} = \frac{A_{p,c,d} - Min(A_{p,c,d})}{Max(A_{p,c,d}) - Min(A_{p,c,d})}$$

For case "i," the overall severity, $O_i$, is calculated as:

$$O_i = 1 - \prod_c (1 - S_{p,c,d}) \ \{i | P_i = p, c \in \mathbf{C}_i, D_i = d\}$$

In predicting mortality, a slightly different formula is used. Let $D_i$ be 1 if the patient died and 0 if discharged alive. First, the average discharge status associated with different pairs of primary diagnoses and comorbidity is calculated:

$$D_{p,c} = \frac{\sum_i^n D_i}{n} \ \{i | p_i = p, c \in \mathbf{C}_i\}$$

The overall severity is calculated as:

$$O_i = 1 - \prod_c (1 - D_{p,c}) \ \{i | p_i = p, c \in \mathbf{C}_i\}$$

The following example shows how the severity of a patient with primary diagnosis of myocardial infarction (MI), congestive heart failure (CHF), and hypertension (H) is calculated. First, the average mortality of the patients who have primary diagnosis of MI and CHF is calculated; assume this is 0.6. Second, the average mortality for cases that have MI and H is calculated; assume this is 0.4. Then, the overall severity for patients who have all 3 conditions is calculated as:

$$S = 1 - (1 - .6)(1 - .4) = 0.76$$

One advantage of the Alemi and Walters measure of severity is that the scoring procedure is transparent, and investigators can easily apply it to their own data. Despite the simplicity of a formula, readers might be concerned with the inherent assumptions of the Alemi and Walters index. Why this formula and why not some other mathematical procedure? Clearly, the formula for the overall severity follows the multiplicative form of the Multi-Attribute Value models. Many investigators have used Multi-Attribute Value to model severity of illness (eg,[10-12]). Our approach differs from the traditional use of these models in 2 ways. First, we score pairs of diagnoses and not a single diagnosis; in essence, we combine pairs of attributes before using the Value model. Second, we use the multiplicative form. To understand our choices, we need to explain "preferential independence." This assumption is met if a shared diagnosis of 2 patients does not affect the order of severity between the patients. If this assumption is not met, then no mathematical formula can be used to aggregate the overall severity of multiple diagnoses from functions defined on single diagnoses.[13] There are many examples where combinations of 2 diseases (eg, MI and CHF or diabetes and renal disease) radically alter the overall severity of illness of the patient. Preferential independence is not met when there is an interaction between the shared diagnosis and one of the other patient's diagnoses. In these circumstances, the best way forward is to score the overall severity for a pair of diagnoses without using a mathematical model. This is, in fact, what we do. We score pairs of diagnoses based on their observed mortality. We then use the Value model to aggregate these scores and calculate the overall severity across pairs of diagnoses. In this fashion, we make sure that we take into account pairwise interactions and ignore higher-order interactions.

If we assume that all pairs of diagnoses meet the preferential independence assumption, then Keeney and Raiffa[13] have shown that the overall severity model must have either an additive or a multiplicative mathematical form. The additive form is not reasonable because, despite its name, it calculates the average of the severity score for each pair. So, if a patient has a pair of diagnoses indicating severe illness and another pair of diagnoses indicating low severity, then the additive model scores the overall severity somewhere between the 2 scores. This does not make logical sense. Adding other illnesses should not reduce the overall severity, even if the new diseases are relatively minor illnesses. In contrast, the multiplicative model always increases the severity score. Any illness, no matter how insignificant, increases the score of the model; less serious illness makes a small increase and more serious illness makes a larger increase. We chose the multiplicative form because it fits our assumption about how severity of illness should be scored.

## Source of Data

We used the 2006 Kids' Inpatient Database (KID) of the Healthcare Cost and Utilization Project of the Agency for Healthcare Research and Quality. This database contained data on pediatric (children 20 years of age and younger) discharges from 3739 community, nonrehabilitation hospitals in 38 states. A total of 3,131,324 un-weighted discharges were available.

## Study Design

KID includes a file where APR-DRG and APS-DRG severity indices are scored for each discharge in the database. The average severity score for a pair of primary diagnosis and comorbidity for the Alemi and Walter index was calculated from 90% of the data (2,818,101 cases). All indices were evaluated on the 10% of data (313,799 cases) set aside for validation sample.

Investigators have used different statistics to report the accuracy of severity indices; these include percent of variation explained, C-statistic, kappa, and area under the receiver operating curves. We focus on percent of variation explained. In analysis of length of stay, this article reports the percent of variation in length of stay explained. This statistic is known as $R^2$ and is calculated by regressing the length of stay on each severity index separately, as well as all indices combined. $R^2$ cannot be used to compare models fitted to different dependent variables. The best way to compare the $R^2$ reported by one index is to compare it with the $R^2$ reported for all indices combined. This comparison is a nested model comparison and provides an intuitive and statistically valid measure of performance improvement.

It also should be clear that length of stay does not have a normal distribution.[14,15] In order to improve the distribution, we ignore 5% of outliers. Nevertheless, because the assumptions of regression have not been met, it is important to consider the $R^2$ statistic as an approximate measure. Transforming the data before conducting the regression could have improved the accuracy of the $R^2$ statistics.

For analysis of mortality, we report the percent of deviance explained. Deviance measures the quality of fit and is sometimes referred to as −2 times the log-likelihood ratio. The percentage of deviance is similar to calculating $R^2$ for the logit function; it is sometimes referred to as pseudo-$R^2$, and is calculated as:

$$R^2_{logit} = (Null\ deviance - Model\ deviance)/Null\ deviance$$

Deviance was calculated from logistic regression of mortality on each severity index and from logistic regression of mortality on all indices combined. The R statistical package was used to prepare ordinary and logistic regressions.

Different analysis was done for various pediatric chronic illnesses. The Agency for Healthcare Research and Quality's Comorbidity Software, Version 3.5 was used to classify patients into a chronic illness. This classification is available in KID data files.

## RESULTS

Table 1 shows the accuracy of the 4 indices in predicting length of stay (a measure of resource use) in the validation sample. The least accurate measure of severity was the simple count of diagnoses and sometimes the APR-DRG. Surprisingly, in patients with fluid and electrolyte disorders and patients with solid tumor without metastasis, the APR-DRG was less accurate than a simple count of diagnoses. The Alemi and Walters index and the APS-DRG indices were more accurate in predicting length of stay than either the count of diagnosis or the APR-DRG. The most accurate measure was, in some disease categories, the Alemi and Walters index and, in other disease categories, the APS-DRG index. The percentage of variation in length of stay explained by the Alemi and Walters index ranged from a low of 12% for patients with an alcohol abuse problem to a high of 64% for patients with complicated diabetes. Across all diseases, the Alemi and Walters index explained 34%, the APS-DRG explained 32%, the APR-DRG explained 13%, and a simple count of diagnoses explained 8% of the variation in length of stay. Using all 4 indices improved the accuracy of the most accurate severity index, the Alemi and Walters index, by 7%.

Table 2 shows the accuracy of the 4 indices in predicting mortality in the validation sample. The Alemi and Walters index explained 21%-93% of mortality in different chronic body conditions. The APS-DRG explained from 1% to 20% of mortality. The APR-DRG explained 4% to 21% of the variation, and the count of diagnoses explained 0% to 5% of the variation in mortality. Across all chronic conditions, the Alemi and Walters index explained 32% of the variation in mortality, while the next most accurate index was APR-DRG, which explained 10% of variation in mortality. Using all 4 indices to predict mortality improved the accuracy of the Alemi and Walters index by 6%.

## DISCUSSION

This article examined the accuracy of 4 different approaches to claims-based severity measurement. We observed a great deal of variation in accuracy among these indices and for the same index in different chronic diseases. Investigators should carefully choose appropriate severity indices that explain the highest percent of variation in their particular study.

With a few exceptions, the Alemi and Walters index was the most accurate predictor of resource use, as measured by length of stay. In predicting mortality, the Alemi and Walters index was several folds more accurate than any of the other indices. Furthermore, the difference in accuracy suggests that improvements are large enough to be of not only statistical but also practical significance. It is possible that with these large differences in accuracy, the conclusion of some comparative effectiveness studies could be reversed. These data suggest that comparative effectiveness studies that wish to rely on claims-based severity indices may benefit from using the Alemi and Walters index. Of course, whether a specific comparative effectiveness study can benefit from a particular index depends on the differences between severity for the intervention and control groups. The analysis presented here was done to validate the relative accuracy of the index and has not shown that any of the indices examined could explain away differences in comparison groups.

Why is the Alemi and Walters index more accurate than the other indices? After all, it uses the same information as the other indices and therefore it could not have an advantage because it has a better source of data. We speculate that the reason for additional accuracy of the Alemi and Walters index has to do with the fact that it does not collapse the patients' diagnoses into homogenous clusters. The APR-DRG and APS-DRG

**TABLE 1: Percent of Variation in Length of Stay Explained by Severity Measures**

| Diseases Associated with the Patients | Selected from 313,131 Validation Cases | Count of Dx | APR-DRG | APS-DRG | Alemi & Walter | All 4 Indices |
|---|---|---|---|---|---|---|
| Acquired immune deficiency syndrome | 35 | 10% | 14% | 2% | 18% | 22% |
| Alcohol abuse | 2097 | 2% | 2% | 11% | 12% | 16% |
| Deficiency anemias | 8600 | 5% | 16% | 29% | 37% | 43% |
| Rheumatoid arthritis/collagen vascular | 4694 | 7% | 10% | 24% | 54% | 59% |
| Chronic blood loss anemia | 4289 | 2% | 9% | 35% | 35% | 42% |
| Congestive heart failure | 498 | 9% | 15% | 39% | 41% | 51% |
| Chronic pulmonary disease | 14,263 | 5% | 6% | 19% | 21% | 25% |
| Coagulopathy | 2272 | 9% | 11% | 31% | 36% | 43% |
| Depression | 3934 | 1% | 7% | 19% | 14% | 22% |
| Diabetes, uncomplicated | 1317 | 6% | 10% | 36% | 25% | 39% |
| Diabetes with chronic complications | 174 | 17% | 36% | 76% | 64% | 84% |
| Drug abuse | 5206 | 1% | 6% | 13% | 22% | 26% |
| Hypertension | 2882 | 5% | 14% | 40% | 35% | 48% |
| Hypothyroidism | 1239 | 8% | 12% | 37% | 36% | 44% |
| Liver disease | 645 | 5% | 12% | 26% | 20% | 30% |
| Lymphoma | 207 | 13% | 23% | 45% | 41% | 54% |
| Fluid and electrolyte disorders | 23,619 | 20% | 22% | 52% | 54% | 62% |
| Metastatic cancer | 412 | 3% | 10% | 34% | 26% | 42% |
| Other neurological disorders | 5398 | 6% | 14% | 39% | 34% | 46% |
| Obesity | 3095 | 2% | 3% | 13% | 12% | 15% |
| Paralysis | 3282 | 4% | 9% | 26% | 27% | 35% |
| Peripheral vascular disorders | 142 | 8% | 19% | 34% | 19% | 42% |
| Psychoses | 3066 | 0% | 1% | 10% | 12% | 15% |
| Pulmonary circulation disorders | 387 | 4% | 7% | 35% | 25% | 40% |
| Renal failure | 893 | 3% | 9% | 26% | 21% | 31% |
| Solid tumor without metastasis | 667 | 11% | 10% | 27% | 33% | 39% |
| Valvular disease | 992 | 9% | 15% | 28% | 32% | 38% |
| Weight loss | 1675 | 8% | 19% | 28% | 28% | 37% |
| Weighted across all patient categories | 95,980 | 8% | 13% | 32% | 34% | 41% |

APS-DRG = All Patient Refined Diagnosis-Related Groups; Dx = diagnosis.

do so. They classify diagnoses into few clusters and score the clusters. In contrast, the Alemi and Walters index scores each pair of diagnoses, without any clustering. There are literally thousands of such pairs. Thus, it allows for many more possible adjustments than scoring a few clusters. The difference of the 2 approaches is essentially the difference between traditional statistical analysis with a few variables and data mining with thousands of variables. The approach that uses more variables could be more refined and may be more accurate.

## Limitations

The Alemi and Walters index has several limitations that may restrict its use in the future. First, this is the first reported comparison of this index with the other claims-based indices. The findings from this study need to be replicated before it is more widely accepted. Second, the index is based on claims data; some investigators prefer using severity indices that are based on key clinical findings. The relative accuracy of the Alemi

**TABLE 2: Percent of Deviance in Mortality Explained by each of the Four Indices**

| | Chronic Body Condition | Selected from 187,073 Validation Cases | Count of Dx | APR-DRG Risk of Mortality | APS-DRG Mortality Weight | Alemi & Walter | All 4 Indices |
|---|---|---|---|---|---|---|---|
| 1 | Infectious & parasitic disease | 169 | NA | NA | NA | NA | NA |
| 2 | Neoplasms | 2090 | 2% | 5% | 5% | 33% | 38% |
| 3 | Endocrine, nutritional, & metabolic & immunity disorders | 6336 | 3% | 10% | 7% | 27% | 33% |
| 4 | Diseases of blood & blood-forming organs | 4234 | 5% | 15% | 12% | 32% | 39% |
| 5 | Mental disorders | 12,753 | 1% | 4% | 4% | 36% | 39% |
| 6 | Diseases of the nervous system & sense organs | 5536 | 1% | 16% | 15% | 40% | 46% |
| 7 | Diseases of the circulatory system | 3505 | 3% | 16% | 12% | 29% | 37% |
| 8 | Diseases of the respiratory system | 13,325 | 1% | 5% | 5% | 30% | 33% |
| 9 | Diseases of the digestive system | 4724 | 2% | 5% | 6% | 61% | 61% |
| 10 | Diseases of the genitourinary system | 1293 | 2% | 11% | 3% | 27% | 33% |
| 11 | Complications of pregnancy, childbirth, & the puerperium | 2749 | NA | NA | NA | NA | NA |
| 12 | Diseases of the skin & subcutaneous tissue | 300 | 0% | 8% | 20% | 54% | 57% |
| 13 | Diseases of the musculoskeletal system | 1152 | 2% | 6% | 12% | 35% | 38% |
| 14 | Congenital anomalies | 15,304 | 2% | 17% | 1% | 21% | 32% |
| 15 | Certain conditions originating in the perinatal period | 518 | 2% | 21% | 0% | 93% | 93% |
| | Overall | 73,988 | 2% | 10% | 6% | 32% | 38% |

APR-DRG = All Patient Refined Diagnosis-Related Groups; Dx = diagnoses.
NA = Sample set for this chronic condition did not have any mortality. Chronic body condition as defined in the 2006 Kid's Inpatient Database of the Healthcare Cost and Utilization Project (HCUP). The reported percentages are based on:
$R^2_{logit} = (Null\ deviance - Model\ deviance)/Null\ deviance$

and Walters index compared with the more clinical indices has not been established. Third, the Alemi and Walters index also has not been compared with other statistical techniques for measuring severity, such as propensity scoring. Finally, while the Alemi and Walters index was more accurate than other indices, it may still not be capturing all the differences in severity of the illness of patients. It was possible to construct 6%-7% more accurate indices by using a combination of the Alemi and Walters index and the other 3 indices. This shows that there is room for improvement in the Alemi and Walters index. We encourage additional comparative studies to establish the accuracy of various claims-based indices.

The availability of accurate claims-based indices for pediatric illness should encourage use of these measures in the Agency for Healthcare Research and Quality's Pediatric Quality Measures program. Sometimes investigators choose an index because it is simpler to use as opposed to more accurate. In the coming years, a major change in claims data is being implemented. The International Classification of Diseases is changing from version 9 (ICD-9) to version 10 (ICD-10). This change will increase diagnoses and procedure codes from the current 14,000 to more than 170,000. These changes require the re-estimation of the parameters for all claims-based databases. A change to ICD-10 codes is expected to further increase the accuracy of the Alemi and Walters index.

*Corresponding Author: Farrokh Alemi, PhD, Department of Health Systems Administration, Georgetown University, 3700 Reservoir Rd., Washington, DC 20007.*
*E-mail address: alemi@cox.net*
*The author of this article holds a patent (George Mason University patent 7,702,526 B2) in the Alemi and Walters Severity Index and may benefit financially from more widespread use of this index. Scientists, students, and government agencies are welcome to use the algorithms described in this article without royalties, but commercial use requires licensing.*
*Farrokh Alemi was supported by NIH grant number 1R01HD055208 "Service Delivery Models to Reduce Maternal and Child Health Disparities."*

## REFERENCES

1. Payne SM, Schwartz RM. An evaluation of pediatric-modified diagnosis-related groups. *Health Care Financ Rev*. 1993;15(2):51-70.
2. Goldfield N. Averill R On "risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool"? *Health Serv Res*. 2000;34(7): 1491-1495. discussion 1495-1498.
3. Romano PS, Chan BK. Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? *Health Serv Res*. 2000;34(7): 1469-1489.
4. Muldoon JH. Structure and performance of different DRG classification systems for neonatal medicine. *Pediatrics*. 1999;103(1 Suppl E):302-318.
5. Averill RF, Goldfield NI, Muldoon J, Steinbeck BA, Grant TM. A closer look at all patient refined DRGs. *J AHIMA*. 2002;73(1):46-50.
6. Leary RS, Johantgen ME, Farley D, Forthman MT, Wooster LD. All-payer severity-adjusted diagnosis-related groups: a uniform method to severity-adjust discharge data. *Top Health Inf Manage*. 1997;17(3):60-71.
7. Wang PS, Walker A, Tsuang M, Orav EJ, Levin R, Avorn J. Strategies for improving comorbidity measures based on Medicare and Medicaid claims data. *J Clin Epidemiol*. 2000;53(6):571-578.
8. Alemi F, Walters SR. A mathematical theory for identifying and measuring severity of episodes of care. *Qual Manag Health Care*. 2006;15(2):72-82.
9. Sedman AB, Bahl V, Bunting E, et al. Clinical redesign using all patient refined diagnosis related groups. *Pediatrics*. 2004;114(4):965-969.
10. Krahn M, Ritvo P, Irvine J, et al. Construction of the Patient-Oriented Prostate Utility Scale (PORPUS): a multiattribute health state classification system for prostate cancer. *J Clin Epidemiol*. 2000; 53(9):920-930.
11. Revicki DA, Leidy NK, Brennan-Diemer F, Thompson C, Togias A. Development and preliminary validation of the multiattribute Rhinitis Symptom Utility Index. *Qual Life Res*. 1998;7(8):693-702.
12. Revicki DA, Leidy NK, Brennan-Diemer F, Sorensen S, Togias A. Integrating patient preferences into health outcomes assessment: the multiattribute Asthma Symptom Utility Index. *Chest*. 1998;114(4): 998-1007.
13. Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley & Sons; 1976.
14. Faddy M, Graves N, Pettitt A. Modeling length of stay in hospital and other right skewed data: comparison of phase-type, gamma and log-normal distributions. *Value Health*. 2009;12(2):309-314.
15. Marazzi A, Paccaud F, Ruffieux C, Beguin C. Fitting the distributions of length of stay by parametric models. *Med Care*. 1998;36(6):915-927.