

# Risk Classification of Medicare HMO Enrollee Cost Levels Using a Decision-Tree Approach

Roger T. Anderson, PhD; Rajesh Balkrishnan, PhD;  
and Fabian Camacho, MS

**Objective:** To determine whether classification tree techniques used on survey data collected at enrollment from older adults in a Medicare HMO could predict the likelihood of an individual being in a low, medium, high, or very high cost group in the subsequent year.

**Methods:** Data from comprehensive health risk assessment (HRA) screening of 11,744 new enrollees (age  $\geq 65$  years) continuously enrolled in a Medicare HMO for at least 24 months between 1997 and 2002 were combined with complete healthcare service utilization data for each enrollee in the postenrollment period to create cost groups. An original clinically devised algorithm and a Classification and Regression Tree (CART) that used the HRA data were compared with respect to their ability to correctly place an individual within the cost groups over 12 months of enrollment.

**Results:** The variables that best classified enrollees into 12-month cost groups included quality of life, age, and preenrollment health services. Classification was best for CART: a sensitivity of 39.8% and a specificity of 83% was achieved for very high cost enrollees. Compared with the clinical algorithm, CART also utilized only one third as many predictor variables to define risk categories.

**Conclusion:** Brief self-report survey data collected at enrollment can be modeled to provide moderate sensitivity in predicting very high costs in the postenrollment year. The ease of applicability of software to create empirically derived profiles of high-cost enrollees precludes the need to rely on clinically devised algorithms to identify enrollees at risk for very high costs.

(*Am J Manag Care. 2004;10(part 1):89-98*)

Providing timely healthcare to an aging population is of increasing concern for health policymakers in the United States. Persons who survive to age 65 years today have a life expectancy of nearly 18 more years. A large number will require care for chronic diseases such as arthritis, hypertension, diabetes, and chronic obstructive pulmonary disease. These highly prevalent conditions among noninstitutionalized older adults (age  $\geq 65$  years) account for a large portion of healthcare expenditures for this age group, and contribute to an overall higher rate of consumption of healthcare resources among this age group than among any other.<sup>1,2</sup> In 1996, the average older adult had nearly 12 annual physician visits and 6.5 days of annual hospital stay, and consumed 36% of all hospital stays and

49% of all hospital days.<sup>3,4</sup> At the same time, managed care enrollment of older adults, which grew nearly 4-fold in the decade from 1989 to 1999, has experienced a reversing trend, with more than 2 million enrollees involuntarily disenrolled from Medicare managed care plans that withdrew from the market because of financial constraints.<sup>3</sup> This trend brings into renewed focus the need to cost-effectively provide care by identifying high-risk older adults and providing interventions to eliminate potentially avoidable medical outcomes and associated resource utilization.

Because the health and well-being of older adults depend on both access to and timing of components and outcomes associated with care,<sup>5,6</sup> care management efforts increasingly rely on health risk assessments to characterize enrollees' healthcare needs and the likelihood of adverse medical outcomes (eg, hospitalization, emergency care). Traditionally, "risk" has been defined in the context of becoming a high-cost user of medical care, or the propensity of enrollees to consume healthcare resources in a particular time period (eg, on an annual basis), based on total payments or billed costs.<sup>2</sup> Individuals deemed to be at "high risk" may be referred to appropriate disease risk management programs.<sup>7-9</sup>

Although it is congressionally mandated that all the Medicare health maintenance organizations (HMOs) in the country conduct risk screening, there is little guidance on which strategy of identifying risk is most effective.

From the Section on Social Sciences and Health Policy, Department of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC (RTA, FC); and the Department of Management and Policy Sciences, University of Texas School of Public Health, Houston, Tex (RB).

This study was funded by an intramural grant from the Wake Forest University, Winston-Salem, NC.

Presented as a poster at the Academy for Health Services Research and Health Policy Annual Meeting, Washington, DC, June 23-25, 2002.

Address correspondence to: Roger T. Anderson, PhD, Section on Social Sciences and Health Policy, Department of Public Health Sciences, Wake Forest University Health Sciences, 2000 West First Street, Piedmont Plaza II, 2nd Floor, Winston-Salem, NC 27104. E-mail: rtanders@wfubmc.edu.

tive and beneficial to enrollees. A practical challenge is implementing a valid and reliable risk assessment process early in the enrollee's health plan tenure, when it can be most useful. Whereas claims-based strategies and clinician assessment require sufficient time in-plan to accrue patient information,<sup>10-12</sup> self-report information supports early risk classification as it enables data collection near time of enrollment and allows elements such as inclusion of disability, self-rated health, and behavioral health variables known to predict service utilization<sup>13,14</sup> that are not routinely recorded in medical information systems. Potential disadvantages of self-report information are possible low sensitivity of self-reported diagnosed and treated conditions,<sup>15,16</sup> and imprecise collection of information on disease severity.

Given the potential value of self-report screening data, a critical challenge is constructing an algorithm from the output of statistical models that maximizes sensitivity and specificity of the cost group classification. Most classification studies in the literature do not involve cost groups, but have relied on basic regression techniques to identify predictors of service utilization.<sup>8,9,11</sup> Although these procedures are sound and valuable tools for deriving a "best set of predictors" of the outcome and percent variation explained, the model parameters that relate to probability do not easily lend themselves to risk categorization because they provide no information on the accuracy of prediction of cost groups in terms of sensitivity and specificity, and because they may not address complex interactions among variables. As a result there is considerable uncertainty over the validity and precision of converting regression weighted scores into decision rules to classify new members into actionable risk-propensity groups.

This study sought to develop and test a risk classification tool for use at the time of enrollment based on its ability to correctly classify individuals into cost groups. Subjects were older adults enrolled in a Medicare managed care plan in the southeastern region of the United States.

---

## METHODS

### Study Population

A prospective cohort was drawn from 11 744 newly enrolled members in a Medicare HMO who completed the comprehensive health risk assessment battery on entry into the plan and for whom complete healthcare service utilization data were available for at least 2 years postenrollment. Claims data were used to follow this cohort annually for 2 years postenrollment. The health plan was a Medicare managed care plan in the southeastern region of United States, and was the sole

provider of medical care to enrollees ("lock-in" risk benefit plan). For the purposes of this analysis, we included only subjects who had not died, moved away, or switched to a fee-for-service plan during the study follow-up period.

Each enrollee was sent a mailed questionnaire that asked questions about demographics, clinical conditions, health service utilization in the year before joining the plan, lifestyle and health behaviors, depression, disability, and health-related quality of life. Since its inception in 1996, the response rate for this questionnaire has been more than 80%. Information from the survey data was then used to derive risk profiles of patients based on a clinician-derived algorithm, which was then provided to case managers for disease state management of high-risk patients.

Demographic characteristics and health status of this population are shown in **Table 1**. Subjects were more likely to be female (58.7%) and had an average age of 75.1 years on enrollment. About 75% rated their general health as good to excellent, while 6% reported their health worsening in the last year (pre-enrollment year).

### Health Risk Assessment Battery

The comprehensive risk assessment tool included self-reported information on conditions known ever to have been diagnosed (heart problems, stroke, chest pain, lung problems, asthma, arthritis, diabetes, poor circulation, hypertension, cancer, bladder problems, stomach/bowel problems, other); lifestyle risk (eg, smoking status, physical activity, alcohol), as adapted from the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance System; general health (the Medical Outcomes Survey 12-Item Short-Form Health Survey [SF-12])<sup>17</sup>; functional status, assessed as impairments in instrumental activities of daily living (IADLs) and activities of daily living (ADLs) as determined by the Duke Older Americans Resources and Services (OARS) screening method<sup>18</sup>; depression (Center for Epidemiological Studies Depression scale [CES-D], short form)<sup>19</sup>; number of falls in the last year; current medication status (number and conditions treated); and use of healthcare services during the last year. This battery of measures represents key aspects of health considered to indicate need or use of health services.

### Study Variables

The responses to the patient questionnaire formed the basis for the demographic and health status variables in the study and are shown in **Table 2**. The demographic variables were age, sex, and living alone (as opposed to with others, including caregivers, spouse, relatives, etc). The health status variables were indica-

tors for sedentary status (physical activity and walking for at least 30 minutes a week), falls during the previous year, smoking status, alcohol consumption, presence of depression (as indicated by a score of .06 on the CES-D short form), number of difficulties in ADLs and IADLs, number of conditions not treated before enrollment, number of comorbid conditions, perception of health status, and the physical and mental subscore components of the SF-12 questionnaire. Healthcare utilization was determined from the number of annual hospitalizations, emergency department visits, and healthcare payments in the 2-year postenrollment period; this information was obtained from enrollee claims data collected by the HMO. Total healthcare payments were obtained from administrative claims records and consisted of all recorded healthcare service utilization and prescription payments. We utilized claims paid by the HMO to calculate the costs associated with each enrollee.

The mean payment per member per month [pmpm], unadjusted for period inflation, was \$296. The cost levels were created using the following split of the distribution of total healthcare costs for 2 years postenrollment: 10% (very high), 20% (high), 30% (medium), and 40% (low). These cutoffs were chosen to produce estimates of cost risk most relevant to Medicare HMO risk-related decision making, based on consultation with Medicare HMO policymakers as well as previously conducted studies.<sup>20</sup> Enrollees were classified as low cost if they had less than \$75 pmpm postenrollment total payments, medium cost if they had payments ranging from \$75 to \$270 pmpm, high cost if they had payments ranging from \$271 to \$792 pmpm, and very high cost if they had payments equal to or more than \$793 pmpm.

### Classification Algorithms

We focused on 3 methods to classify patients into the different cost levels specified above. The first method involved applying an original cost estimation algorithm proposed by clinicians and experts in the field (see **Figure 1** for a description). The second approach involved using Salford Systems' Classification and

**Table 1.** Descriptive Statistics for Entire Sample (n = 11 744)\*

Characteristic	Mean (Standard Deviation) [Range]
<b>Demographics</b>	
Age, y	75.1 (6.6) [65-110]
Male sex, %	41.3
Sedentary, %	66.5
Living alone, %	25.9
High alcohol consumption, %	11.9
Smoker, %	14.0
Falls > 1, %	7.8
<b>Health status variables</b>	
SF-12 Physical Component Subscore	46.70 (10.50) [11.04-66.40]
SF-12 Mental Component Subscore	51.09 (8.64) [14.89-68.28]
Health the same as last year, %	79.3
ADLs, problems with	0.18 (0.80) [0-8]
IADLs, problems with	0.47 (1.25) [0-7]
Depressive symptomatology, %	10.6
Conditions not treated preenrollment	0.97 (1.16) [0-9]
Good-excellent general health perception, %	75.4
<b>Preenrollment healthcare utilization</b>	
No emergency department visits the year before, %	78.5
No hospitalizations the year before, %	83.4
At least 1 physician visit the year before, %	91.8
<b>Postenrollment dependent variable</b>	
Total per member per month payments, \$	296.01 (539.89) [0-11522.04]

\*ADLs indicate activities of daily living; IADLs, instrumental activities of daily living; and SF-12, Medical Outcomes Survey 12-Item Short-Form Health Survey.

Regression Tree (CART) software.<sup>21,22</sup> CART analysis is a nonparametric technique ideally suited to identify patterns or "groups" in complex data. It has been used in aging research,<sup>23-28</sup> when disability group status was of primary interest,<sup>23-25</sup> and by medical research to develop treatment decision rules for clinicians<sup>26-29</sup>; thus, it has the potential for a more general application to managed care risk models. For comparison purposes, the third method involved assigning cost levels at random based on the distribution of cost levels.

Candidate CARTs were first built, evaluated, and compared with each other using a set of data consisting of two thirds of the available sample (the training set); the best tree was then further evaluated for its actual performance using the remaining data (referred to in the literature as the test set).<sup>30</sup> With this approach, we could select the best-performing tree without bias and judge its true performance with new data.

CART and other tree-structured classifiers attempt to recursively partition observations into groups to maximize the classification rate.<sup>31</sup> The CART decision tree identifies groups in the optimal partition produced by

**Figure 1.** Original Risk Estimation Algorithm

Very High (4)	High (3)	Medium (2)	Low (1)
If any of the following are met:	If any of the following are met:	If any of the following are met:	If ALL of the following are met:
1. CES-D symptoms present and health now = worse, much worse; and general health = poor	1. Asthma	1. Only 1 of CV problems, diabetes, stomach/bowel conditions	1. None of: CV problems Diabetes Cancer GU/GI problems Asthma
2. Number of falls >1	2. Cancer	2. Health now = worse	2. No emergency department visit year before
3. ADLs > 3 and asthma, CV problems, or cancer	3. CES-D symptoms present	3. ADL = 1 and IADL ≤ 1	3. Hospital visit year before ≤ 1
4. Emergency department + hospital visits year before ≥ 4	4. Health now = worse, much worse	4. ADL = 0 and IADL ≥ 1	
5. Asthma or cancer and CES-D symptoms present; and health now = worse, much worse; and (ADLs ≥ 2 or (ADLs = 1 and IADLs ≥ 2)	5. General health = poor	<b>Then classify as Medium, or else proceed to next category ⇒</b>	
6. CV problems and cancer and general health = fair; poor	6. ADLs ≥ 2		
7. (ADLs ≥ 2 or (ADLs = 1 and IADLs ≥ 2) and asthma or cancer and lives alone	7. ADLs = 1 and IADLs ≥ 2		
<b>Then classify as Very High, or else proceed to next category ⇒</b>	8. CV problems, diabetes, or stomach/bowel conditions present; and health now = worse, and (ADLs = 1 and IADLs ≤ 1 or (ADLs = 0 and IADLs ≥ 1)		
	9. Two or more of CV problems, diabetes, or stomach/bowel conditions		
	<b>Then classify as High, or else proceed to next category ⇒</b>		

ADLs indicate activities of daily living; CES-D, Center for Epidemiological Studies Depression scale; CV, cardiovascular; GI, gastrointestinal; GU, genitourinary; and IADLs, instrumental activities of daily living.

the algorithm by asking a series of hierarchical, binary, Yes/No questions, and is easy to depict, as can be seen in **Figure 2**. In medical diagnosis situations, binary trees produced by CART tend to be smaller and have slightly lower error rates than multiway trees.<sup>32</sup>

The detailed methodology involved in developing the most appropriate CART can be found on the developers' Web page.<sup>21</sup> The CART algorithms used for this study is available from the authors upon request.

How unknown values are treated may affect any classification algorithm used for HRA data, because sample observations often contain missing values. The largest proportion of missing values in our data was 13% for the SF-12 Physical Component Subscore. Missing values may be easier to manage with decision trees than they are with other classification methods.<sup>33</sup> The tree-building algorithm in the Salford System CART software uses a method of "surrogate" variables, in which observations with missing predictor variables used in the tree are not dropped from the analysis; instead, the missing predictor is replaced by a proxy variable best approximating its performance.<sup>20,21</sup>

.....  
**RESULTS**

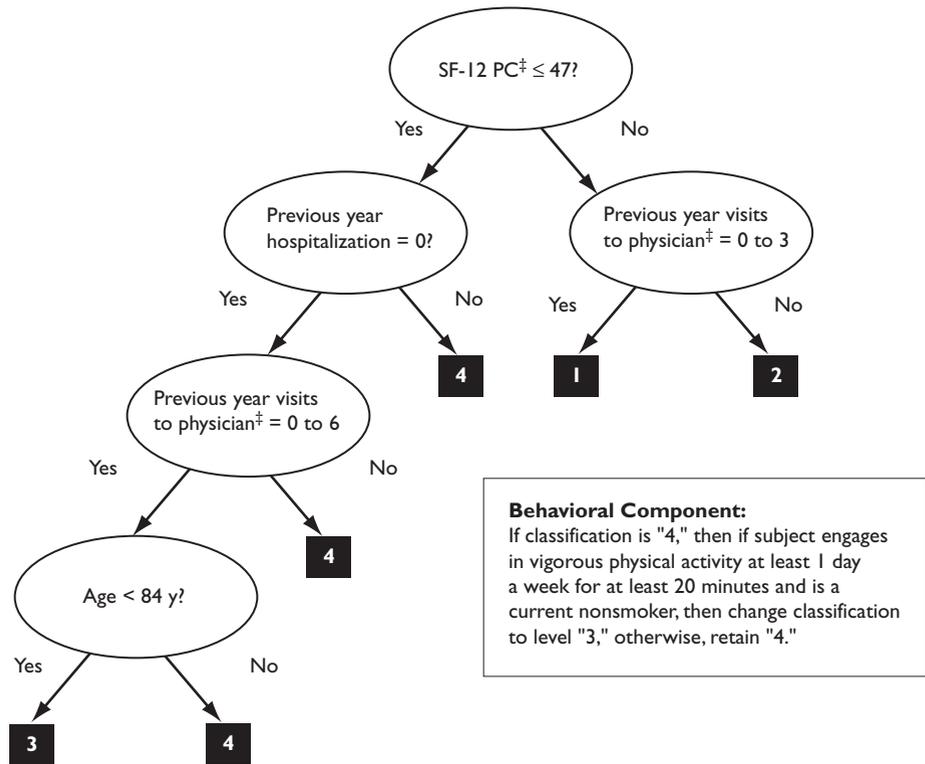
**Tables 1-4** present data on the characteristics of the sample. Table 2 and Table 3 break down the descriptive statistics in the population by the categories of risk based on the original clinical algorithm and the CART, respectively. Table 4 lists potential predictors considered in the various cost-

based risk models. The original risk estimation algorithm and the CART decision tree are shown in Figures 1 and 2, respectively.

Results from the comparisons of the 3 classification approaches are shown in Table 5. The overall classification rates show the proportion of enrollees who were correctly classified in their subsequent 12-month cost group, based on the number of enrollees actually classified in each group. There was a significant difference between the classification rate of the CART and that of the original algorithm (McNemar test  $P < .0001$ , with the CART improving overall classification by 7.1 percentage points), as well as a significant difference between the classification rates of the CART and the random classifier ( $P < .0001$ , with the CART improving overall classification by 6.8 percentage points). However, there was no significant difference between the rates of the original (clinician-based) algorithm and the random classifier ( $P = .22$ ).

Table 6 shows the sensitivity, specificity, and positive predictive values of each classifier. The probability of correct classification among high-cost patients is highest for the CART (39.8%). For enrollees assigned to the very high cost group, the CART also had the highest positive predictive value (20.4%), although this varied by only 1.2 percentage points from the original, clinician-based algorithm (19.2%). The CART had lower sensitivity than both the original algorithm and the random classifier for classifying enrollees in the middle range (high and medium) cost groups (Table 3) and was slightly more sensitive than the random classifier for enrollees in the high-cost group (21.9% vs 19.0%). In

Figure 2. Decision Tree for the Classification and Regression Tree\*



\*SF-12 PCS indicates Medical Outcomes Survey 12-Item Short-Form Health Survey–Physical Component Subscore.  
 †If the SF-12 PCS is missing, then use general health as a surrogate variable. If general health = 4 or 5 (poor), then go left.  
 ‡If Previous Year Visits to Physician is missing, then use Previous Year Hospitalization as a surrogate variable. If Previous Year Hospitalization = 0, then go left.  
 §If Previous Year Visits to Physician is missing, then use activities of daily living (ADLs) as a surrogate variable. If ADLs = 0, then go left.

contrast, the original clinician-based algorithm showed distinctly lower sensitivity for the very high risk cost group and the low-cost group compared with the CART. To increase the sensitivity of the CART for the middle-cost groups, we instructed the tree to modify the classification of “very high” to “high” risk if there was evidence of self-reported physical activity or of nonsmoking status. The modifications are described in Figure 2, and the performance of this classifier also is described in Tables 5 and 6. In general, we observed that the performance of this modified classifier was indistinguishable from that of the original tree, except that the sensitivity of the model for the medium-cost level increased by 3.7 percentage points, at the expense of a decrease of 6.5 percentage points at the very high cost level.

**Table 2.** Descriptive Statistics for Original Algorithm Categories (n = 11 744)\*

Characteristic	Low Risk	Medium Risk	High Risk	Very High Risk
<b>Demographics</b>				
Age, y	73.8 (5.6)	75.0 (6.5)	75.0 (6.6)	77.4 (7.7)
Male sex, %	41.4	43.6	40.0	38.7
Sedentary, %	59.0	63.5	68.9	78.3
Living alone, %	22.4	25.3	26.6	31.2
High alcohol consumption, %	14.1	12.2	11.6	8.1
Smoker, %	17.1	13.5	13.5	11.4
Falls > 1, %	0.0	0.0	0.0	61.1
<b>Health status variables</b>				
SF-12 Physical Component Subscore	52.4 (6.3)	49.1 (8.3)	44.9 (10.7)	36.7 (11.7)
SF-12 Mental Component Subscore	53.2 (6.2)	52.7 (6.9)	49.9 (9.6)	47.2 (10.9)
Health the same as last year, %	89.2	85.0	74.2	64.1
ADLs, problems with	0.0	0.02 (0.1)	0.2 (0.6)	0.9 (1.9)
IADLs, problems with	0.0	0.2 (0.7)	0.5 (1.2)	1.7 (2.2)
Depressive symptomatology, %	0.0	0.0	20.3	26.5
Conditions treated preenrollment	0.2 (0.5)	0.8 (0.9)	1.4 (1.3)	1.4 (1.5)
Good-excellent general health perception, %	95.8	86.9	75.3	43.7
<b>Preenrollment healthcare utilization</b>				
At least 1 emergency department visit the year before	10.5	17.6	21.9	46.7
At least 1 hospitalization the year before	6.6	13.3	17.1	38.5
At least 1 physician visit the year before	78.3	94.0	96.0	95.6
<b>Postenrollment dependent variable</b>				
Total biannual payments, \$	177.50 (347.54)	259.37 (516.81)	335.30 (586.36)	461.48 (644.27)

\*Standard deviations are indicated in parentheses where applicable. ADLs indicate activities of daily living; IADLs, instrumental activities of daily living; and SF-12, Medical Outcomes Survey 12-Item Short-Form Health Survey.

## DISCUSSION

In the literature, various methodologic approaches to risk assessment have been proposed.<sup>9-11</sup> Medical and claims-based information is sometimes viewed as providing more details regarding medical need, but these data are rarely available at the time of plan enrollment. To satisfy the need to identify subsets of newly enrolled patients who may require timely and focused care to avert negative health events, and to ensure that a new enrollee's needs do not go unrecognized in the transition from one plan to another, self-reported survey information may be a valuable option. It is well known that survey data have value in predicting cost outcomes.<sup>1,2,9,11</sup>

We focused on total payments for new enrollees of a Medicare managed care plan and included a comprehensive assessment of comorbidity and lifestyle risk factors to select our models. More importantly, we compared classification methods on their ability to cor-

rectly classify enrollees into subsequent 12-month cost groups. We know of no other reports in the literature that investigated this essential property of classification. We used CART models as an established method for group classification and sought to extend their utility in predicting specific health services utilization groups relevant to clinical operations.<sup>22,29</sup> The significant variables predicting high-cost risk differed between the original clinical algorithm and the CART. The CART predictors were a parsimonious set of 4 predictor variables (age, self-reported physician visits, hospitalization in the preenrollment year, and the SF-12 Physical Component Subscore), in contrast to the 8 complex categories (with 12 predictor variables) used to define high risk in the clinical algorithm.

Most regression-based studies of health services utilization in the literature have ended up selecting variable sets based on significant correlations with healthcare costs and categorical outcomes associated with morbidity and high-cost risk (eg, hospitaliza-

**Table 3.** Descriptive Statistics for Classification and Regression Tree Categories (n = 11 744)\*

Characteristic	Low Risk	Medium Risk	High Risk	Very High Risk
<b>Demographics</b>				
Age, y	74.2 (6.0)	74.5 (6.1)	73.8 (5.0)	79.0 (8.3)
Male sex, %	44.3	43.9	37.4	36.1
Sedentary, %	59.4	62.6	72.3	79.8
Living alone, %	24.0	27.3	25.5	29.5
High alcohol consumption, %	14.0	13.0	10.3	7.5
Smoker, %	14.9	11.7	15.6	12.1
Falls > 1, %	3.4	5.6	9.3	18.4
<b>Health status variables</b>				
SF-12 Physical Component Subscore	54.3 (2.94)	53.3 (3.0)	37.8 (7.8)	34.4 (8.4)
SF-12 Mental Component Subscore	51.8 (7.09)	50.9 (8.1)	51.5 (9.4)	49.3 (10.8)
Health the same as last year, %	85.9	75.9	80.9	65.6
ADLs, problems with	0.03 (0.3)	0.08 (0.5)	0.2 (0.8)	0.6 (1.4)
IADLs, problems with	0.1 (0.6)	0.2 (0.9)	0.5 (1.2)	1.4 (2.0)
Depressive symptomatology, %	5.2	7.5	14.4	22.1
Conditions treated preenrollment	0.8 (1.0)	1.0 (1.1)	1.2 (1.3)	1.2 (1.4)
Good-excellent general health perception, %	97.5	92.5	59.4	46.5
<b>Preenrollment healthcare utilization</b>				
At least 1 emergency department visit the year before	12.3	26.3	14.6	45.9
At least 1 hospitalization the year before	7.0	24.2	0.0	50.4
At least 1 physician visit the year before	84.7	100	93.5	97.7
<b>Postenrollment dependent variable</b>				
Total per member per month payments, \$	205.37 (468.67)	288.94 (474.48)	302.85 (523.31)	502.30 (686.63)

\*Standard deviations are indicated in parentheses where applicable. ADLs indicate activities of daily living; IADLs, instrumental activities of daily living; and SF-12, Medical Outcomes Survey 12-Item Short-Form Health Survey.

tions).<sup>1,9,11</sup> Although this work is valuable in identifying the nature of exposures and conditions that lead to utilization, the questions of how well these selected variables can classify individuals correctly into cost groups and whether risk scores produced from regression weights apply across the continuum of cost or utilization were not addressed. Advantages of CART for assessing risk include its ability to identify risk groups that lend themselves to decision rules, its ability to handle potentially complex interactions among predictor variables, the avoidance of model errors due to highly skewed data, and its relative simplicity, enabling interpretation by clinicians or nonstatisticians.<sup>22</sup>

Our CART classification models were contrasted with a model of chance and with expert opinion to enable relative comparisons to be made across defined healthcare cost groups. The results were complex. The data-derived model (based on CART) appeared to be superior in sensitivity for extremes of very high cost

and low cost, both of which are most important to Medicare HMO decision makers and care planners. For high-utilization risk classification, we obtained superior results using only 5 predictors compared with a more subjective methodology based on clinician judgment. Unlike regression models that yield parameter estimates, the decision tree results, once tabulated and depicted, are simple to interpret and use. It is noteworthy that empirically derived CART models can be obtained for care management decision rules by applying readily available software to enrollee risk survey data. The software, while technical in origin, is easy to use and may be operated by those trained to perform regression analyses.<sup>21,22</sup>

There are some important limitations in the study. First, we tested a limited set of classification approaches that we selected variously based on their use in related fields: a straightforward decision-tree format of the results and a more traditional classification approach based on clinical judgment and impression. There may

**Table 4.** Potential Predictors Used in Selecting Classifier

Predictor	Description
Age	(Continuous)
Sex	Male, female
Living with whom	Alone, with spouse, offspring, family, others
Alcoholic drinks consumed in a typical week	None, 1-3, 4-7, 8-10, 11-14, 15+
Vigorous physical activities per week	Never, 1, 2, 3, 4, 5+
Smoker	Yes, never smoked, quit
Falls in past year	0, 1, 2, 3, 4+
SF-12 Physical Component Subscore	(Continuous)
SF-12 Mental Component Subscore	(Continuous)
Health now compared with last year	1 (much better) to 5 (much worse)
ADLs	0, 1, 2, 3, 4, 5, 6, 7, 8
IADLs	0, 1, 2, 3, 4, 5, 6, 7
Depressive symptomatology present (CES-D)	Yes/No
No. of conditions not treated preenrollment	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
General health	1 (excellent) to 5 (poor)
Previous year emergency department use	Never, 1, 2, 3, 4+
Previous year hospital use	Never, 1, 2, 3, 4+
Previous year physician visits	Never, 1-3, 4-6, 7-9, 10+
Cardiovascular conditions ever present	Yes/No
Stomach/ bowel conditions ever present	Yes/No
Chest pain (angina) ever present	Yes/No
Arthritis ever present	Yes/No
Asthma ever present	Yes/No
Cancer ever present	Yes/No
Diabetes ever present	Yes/No
Lung problems ever present	Yes/No
Other health problems ever present	Yes/No

ADLs indicate activities of daily living; CES-D, Center for Epidemiological Studies Depression scale; IADLs, instrumental activities of daily living; and SF-12, Medical Outcomes Survey 12-Item Short-Form Health Survey.

be other classification models superior to those tested here, such as propensity scores<sup>34</sup> and discriminant analytic models. We are not aware of an appreciable literature on classification of risk for healthcare services utilization, and our study was designed to test a leading potential approach. Another limitation is the potential for idiosyncratic policy and practice issues in the study HMO that could have influenced use of services—either blunting or biasing our risk estimates. Our data came

from 1 HMO, and thus are not sufficiently diverse to represent variations in policy and referral encountered elsewhere. Our population of enrollees may not have been typical of other HMO populations. However, we are not aware of unique or local policy or administrative issues that could have biased our results.

The sample size for our study also was limited, but other risk assessment studies used similar numbers of patients to illustrate the effectiveness of various risk assessment approaches.<sup>35,36</sup> The results of our illustrative study, however, have to be validated with other populations before a definitive assessment of the general applicability of classification approaches like CART can be made. The self-report of some of the risk screen variables also could introduce an element of respondent bias. For example, we found that nearly three quarters of our sample reported their health to be good or excellent. This unusually positive response may be in part because these are ambulatory older adults, and the risk screens were completed at the time of enrollment. Thus, there may have been a positive response bias in subjects who feared losing coverage if they reported their health status otherwise.

In support of our sample, we found the prevalence of comorbidities and the disability status of our population to be highly similar to national estimates for older adults reported in the National Health Interview survey.<sup>3</sup> We also examined the mean annualized payments of enrollees who did not return a survey

and found them to be highly similar to those of enrollees who did return a survey. Certain limitations exist because our sample size of approximately 12 000 enrollees could be considered small, especially when dealing with relatively rare events that have very high costs. Finally, our study examined total healthcare utilization and payments, as opposed to examining disease-specific utilization and payments. This was done to avoid problems inherent in attributing all billings for healthcare

service utilization and payments to a specific condition due to miscoding of the diagnosis. These limitations do not, in our view, undermine the validity of the potential benefits of our study of alternative risk classification methodologies such as CART for comprehensive risk classification of older adults.

important early opportunity to gather health information on enrollees to aid in risk management, needs surveillance, and interventions to improve care.

CONCLUSION

We found that 5 self-reported health status variables from newly enrolled Medicare managed care enrollees can be used to classify cost groups for use in disease management and risk reduction efforts of HMOs. Clarification and regression tree analysis can be easily applied to health risk assessment data, thus providing HMO decision makers with more sensitive estimates of cost risk than those based on clinically derived risk algorithms. This method offers an

**Table 5.** Classification Rates and Other Measures for Classifiers\*

Classifier	Classification Rate on Test Set, % (n = 3915) <sup>†</sup>	No. of Predictors	Distribution of Assigned Cost Groups on Test Set, %			
			1	2	3	4
1. Original algorithm <sup>‡</sup>	29.9 (.73)	20	17.3	29.4	38.7	14.7
2. Random classifier <sup>§</sup>	30.2 (.73)	0	38.0	32.8	19.4	9.7
3. Classification and Regression Tree, best overall classification	37.0 (.77)	5	39.3	17.0	14.8	28.9
4. Classification and Regression Tree, with behavioral component	37.0 (.77)	7	39.3	23.9	14.8	22.0

\*1 indicates the low-risk category; 2, medium risk; 3, high risk; and 4, very high risk.  
<sup>†</sup>Resubstitution estimates were calculated from the test set. Missing cases were dropped. Standard deviations of estimates are in parentheses.  
<sup>‡</sup>Algorithm allows for making no decision. Proportions are only among cases in which a decision was made.  
<sup>§</sup>Random classifier assigned cost groups randomly based on the probabilities of the cost group distribution of the training sample.

**Table 6.** Additional Classifier Properties Calculated from Test Set (n = 3915)\*

Classifier	Percentage											
	Classification Rate Given Actual Cost Group (Sensitivity)				Probability of Assigning Cost Group When a Different Cost Group Usage Occurs (1 - Specificity)				Positive Predictive Value of Cost Group			
	1	2	3	4	1	2	3	4	1	2	3	4
Random classifier	39.3	34.0	19.0	11.3	37.1	32.3	19.5	9.5	42.3	31.5	18.4	11.6
Original algorithm <sup>†</sup>	26.5	29.9	40.0	24.5	16.0	31.7	35.7	11.3	55.1	29.1	20.6	19.2
Classification and Regression Tree	55.2	20.9	21.9	39.8	34.0	17.6	18.9	17.0	52.9	34.2	21.2	20.4
Classification and Regression Tree, with behavioral component	55.2	20.9	25.6	33.3	34.0	17.6	21.8	14.3	52.9	34.2	21.3	20.4

\*Resubstitution estimates were calculated from the validation set. Missing cases were dropped. 1 indicates the low-risk category; 2, medium risk; 3, high risk; and 4, very high risk.  
<sup>†</sup>Algorithm allows for making no decision. Proportions are only among cases in which a decision was made.

## REFERENCES

1. Riley G, Tudor C, Chiang YP, et al. Health status of Medicare enrollees in HMOs and fee-for-service in 1994. *Health Care Financ Rev.* 1996;17(4):65-76.
2. Garfinkel SA, Riley GF, Iannacchione VG. High-cost users of medical care. *Health Care Financ Rev.* 1988;9(4):41-52.
3. Gold MR, McCoy J. Choice continues to erode in 2002. *Monitoring Medicare+Choice Fast Facts #7.* Washington, DC: Mathematica Policy Research Inc; January 2002. Available at: [http://www.mathematica-mpr.com/PDFs/redirect\\_PubsDB.asp?strSite=fastfacts7.pdf](http://www.mathematica-mpr.com/PDFs/redirect_PubsDB.asp?strSite=fastfacts7.pdf). Accessed October 2, 2003.
4. American Association of Retired Persons. *A Profile of Older Americans.* Washington, DC: AARP; 1999.
5. Levitt L, Lundy L, Srinivasan S. *Trends and Indicators in the Changing Health Care.* MarketPlace: A Chartbook. Menlo Park, Calif: The Henry J. Kaiser Family Foundation; 2002.
6. Fincham JE. Medication compliance and the elderly. *J Pharmacoepidemiol.* 1995; 4(2):7-14.
7. Gruenberg L, Tompkins C, Porell F. The health status and utilization patterns of the elderly: implications for setting Medicare payments to HMOs. *Adv Health Econ Health Serv Res.* 1989;10:41-73.
8. Beebe J, Lubitz J, Eggers P. Using prior utilization to determine payments of Medicare enrollees in health maintenance organizations. *Health Care Financ Rev.* 1985;6:27-38.
9. Pacala JT, Boulton C, Reed RL, et al. Predictive validity of the PRA instrument among older recipients of managed care. *J Am Geriatr Soc.* 1997;45(5):614-617.
10. Meenan RT, O'Keefe-Rosetti C, Hornbrook MC, et al. The sensitivity and specificity of forecasting high-cost users of medical care. *Med Care.* 1999;37(8):815-823.
11. Hornbrook MC, Goodman MJ. Chronic disease, functional health status and demographics: a multi-dimensional approach to risk adjustment. *Health Serv Res.* 1996;31(3):283-307.
12. Fowles JB, Weiner JP, Knutson D, et al. Taking health status into account when setting capitation rates: a comparison of risk-adjustment methods. *JAMA.* 1996; 276(16):1316-1321.
13. Riley GF. Risk adjustment for health plans disproportionately enrolling frail Medicare beneficiaries. *Health Care Financ Rev.* 2000;21(3):135-148.
14. Unutzer J, Patrick DL, Simon G, et al. Depressive symptoms and the cost of health services in HMO patients aged 65 years and older. A 4-year prospective study. *JAMA.* 1997;277(20):1618-1623.
15. Raina P, Torrance-Rynard V, Wong M, et al. Agreement between self-reported and routinely collected health-care utilization data among seniors. *Health Serv Res.* 2002;37(3):751-774.
16. Fowles JB, Fowler EJ, Craft C. Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. *J Ambul Care Manage.* 1998;21(1):24-34.
17. Ware JE, Kosinski U, Keller SD. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales.* 2nd ed. Boston, Mass: The Health Institute, New England Medical Center; 1995.
18. **Multidimensional Functional Assessment: The OARS Methodology: A Manual.** 2nd ed. Durham, NC: Duke University, Center for the Study on Aging and Human Development; 1978.
19. Burnam MA, Wells KB, Leake B, et al. Development of a brief screening instrument for detecting depressive disorders. *Med Care.* 1988;26:775-789.
20. Pearlman DN, Branch LG, Ozminkowski RJ, et al. Transitions in health care use and expenditures among frail older adults by payor/provider type. *J Am Geriatr Soc.* 1997;45(5):550-557.
21. Salford Systems. CART decision tree software. Available at: <http://www.salford-systems.com/products-cart.html>. Accessed October 2, 2003.
22. Lewis RJ. An introduction to the Classification and Regression Tree (CART) analysis. Working paper presented at: 2000 Annual Meeting of the Society for Academic Emergency Medicine; May 23, 2000; San Francisco, Calif. Available at: <http://www.saem.org/download/lewis1.pdf>. Accessed October 2, 2003.
23. Wolfe F, Pincus T, O'Dell J. Evaluation and documentation of rheumatoid arthritis disease status in the clinic: which variables best predict change in therapy. *J Rheumatol.* 2001;28(7):1712-1717.
24. El-Solh AA, Sikka P, Ramadan F. Outcome of older patients with severe pneumonia predicted by recursive partitioning. *J Am Geriatr Soc.* 2001;49(12):1614-1621.
25. Kuchibhatla M, Fillenbaum GG. Assessing risk factors for mortality in elderly White and African American people: implications of alternative analyses. *Gerontologist.* 2002;42(6):826-834.
26. Hess KR, Abbruzzese MC, Lenzi R, et al. Classification and regression trees analysis of 1000 consecutive patients with unknown primary carcinoma. *Clin Cancer Research.* 1999;5:3403-3410.
27. Rainer TH, Lam PK, Wong EM, et al. Derivation of a prediction rule for post-traumatic acute lung injury. *Resuscitation.* 1999;42:187-196.
28. Selker HP, Griffith JL, Patil S, et al. A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. *J Invest Med.* 1995;43:468-476.
29. Falconer JA, Naughton BJ, Dunlop DD, et al. Predicting stroke inpatient rehabilitation outcome using a classification tree approach. *Arch Phys Med Rehab.* 1994;75:619-625.
30. Bishop CM. *Neural Networks for Pattern Recognition.* New York, NY: Oxford University Press; 1995.
31. Bratko I, Kononenko I. Learning diagnostic rules from incomplete and noisy data. In: Phelps B, ed. *Interactions in Artificial Intelligence and Statistical Methods.* Aldershot, UK: Gower Technical Press; 1987:142-153.
32. Breiman LJ, Friedman R, Olshen R, Stone C. *Classification and Regression Trees.* Monterey, Calif: Wadsworth and Brooks/Cole; 1984.
33. Buntine W. Learning classification trees. *Statistics and Computing.* 1992;2:63-73.
34. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med.* 1998;17(19):2265-2281.
35. Pacala JT, Boulton C, Urdangarin C, McCaffrey D. Using self-reported data to predict expenditures for the health care of older people. *J Am Geriatr Soc.* 2003;51(5):609-614.
36. Reuben DB, Keeler E, Seeman TE, Sewall A, Hirsch SH, Guralnik JM. Identification of risk for high hospital use: cost comparisons of four strategies and performance across subgroups. *J Am Geriatr Soc.* 2003;51(5):615-620.