Use of risk-adjusted change in health status to assess the performance of integrated service networks in the Veterans Health Administration

ALFREDO J. SELIM1,2,3, DAN BERLOWITZ1,3, GRAEME FINCKE1,3, WILLIAM ROGERS1,5, SHIRLEY QIAN1, AUSTIN LEE1,4, ZHONGXIAO CONG1, BERNARDO J. SELIM3, XINHUA S. REN1,3, AMY K. ROSEN1,3 AND LEWIS E. KAZIS1,3

1Center for Health Quality, Outcomes, and Economic Research, A Health Services Research and Development Field Program, VA Medical Center, Bedford, Massachusetts, 2Section of General Internal Medicine and Emergency Services, Boston VA Health Care System, West Roxbury, Massachusetts, 3Boston University Schools of Medicine and Public Health, Boston, Massachusetts, 4Boston University, Mathematics, Boston, Massachusetts, and 5Senior Research Scientist, The Health Institute, New England Medical Center

Abstract

Objective. Health outcome assessments have become an expectation of regulatory and accreditation agencies. We examined whether a clinically credible risk adjustment methodology for the outcome of change in health status can be developed for performance assessment of integrated service networks.

Study design. Longitudinal study.

Setting. Outpatient.

Study participants. Thirty-one thousand eight hundred and twenty-three patients from 22 Veterans Health Administration (VHA) integrated service networks were followed for 18 months.

Main measures. The physical (PCS) and mental (MCS) component scales from the Veterans Rand 36-items Health Survey (VR-36) and mortality. The outcomes were decline in PCS (decline in PCS scores greater than $-6.5$ points or death) and MCS (decline in MCS scores greater than $-7.9$ points).

Results. Four thousand three hundred and twenty-eight (13.6%) patients showed a decline in PCS scores greater than $-6.5$ points, 4322 (13.5%) had a decline in MCS scores by more than $-7.9$ points, and 1737 died (5.5%). Multivariate logistic regression models were used to adjust for case-mix. The models performed reasonably well in cross-validated tests of discrimination ($c$-statistics = 0.72 and 0.68 for decline in PCS and MCS, respectively) and calibration. The resulting risk-adjusted rates of decline in PCS and MCS and ranks of the networks differed considerably from unadjusted ratings.

Conclusion. It is feasible to develop clinically credible risk adjustment models for the outcome of decline in PCS and MCS. Without adequate controls for case-mix, we could not determine whether poor patient outcomes reflect poor performance, sicker patients, or other factors. This methodology can help to measure and report the performance of health care systems.

Keywords: health-related quality of life, outpatient care, patient outcomes, quality of care

There is increasing emphasis on ambulatory care and a corresponding need to evaluate its outcomes [1]. One particularly useful outcome is health status, which can be determined by patient self-report using well established and carefully validated questionnaires such as the Medical Outcomes Study Short Form 36 (MOS SF-36) [2]. Studies have shown that the MOS SF-36 is sensitive to changes in health in general populations [3]. Ware et al. [4] have linked changes in MOS SF-36 scores to the performance of systems of care. Efforts to infer quality from such assessments may be misleading, however, if we do not account for differences in patient’s clinical characteristics, or case-mix, that can also influence changes in health status.

Risk adjustment is a methodology to control for patient characteristics that may affect health care outcomes. A variety

Address reprint requests to Alfredo J. Selim, Center for Health Quality, Outcomes and Economic Research (CMQOER), Edith Nourse Rogers Memorial Hospital (152), Building 70, 200 Springs Road, Bedford, MA 01730, USA. E-mail: selim.alfredo_j@boston.med.va.gov
of measures are currently applied to adjust for risk across ambulatory populations [5]. Although these applications represent significant advances in the measurement of case-mix, they focus primarily on cost and care utilization. Studies have shown that both the socioeconomic background and individual clinical status of patients influence performance measurements [4,6]. However, methods for risk adjustment to predict change in health status are yet to be tested.

In this article, we aimed to develop and validate a comprehensive approach for risk adjustment of change in health status to assess the performance of integrated service networks in the Veterans Health Administration (VHA). These networks are designed to provide coordinated and integrated care with an emphasis on outpatient treatment. Given that they vary in geography and patient characteristics, they offer a unique opportunity to study case-mix adjustment for change in health status. In this study, we specifically addressed three research objectives: (i) to determine whether clinically credible and statistically reliable risk-adjusted models can be developed, (ii) to examine whether case-mix differences exist across integrated service networks, and (iii) to assess whether risk adjustment alters judgements of network performance.

Methods

Study population

This study used data from the National Survey of Ambulatory Care Patients [7]. Any veteran who received ambulatory care in VHA integrated service networks between 1 January 1997 and 17 December 1997 was eligible to be sampled. A typical network encompasses a geographic region with an average of 7–10 VA medical centers, 25–30 ambulatory care clinics, 4–7 nursing homes, 1–2 domiciliaries, and 10–15 counseling centers [8]. Among 43,965 patients randomly recruited, 31,823 (72.3%) patients completed a mailed survey in February 1998. Of those patients, 1,737 (5.5%) died during the 18 months follow-up and 21,378 (67.2%) responded to a mailed follow-up survey that was administered 18 months after the baseline survey. We identified 8708 patients with missing follow-up information.

Outcome measures

We measured outcomes using the Veterans Rand 36-items Health Survey (VR-36) [9], a reliable and valid measure of health-related quality of life modified from the MOS SF-36 and mortality data. The SF-36 scales were summarized into physical (PCS) and mental component (MCS) scales [10]. The two summaries, PCS and MCS, were scored using a linear t-score transformation that was normed to a general US population. Mortality was determined using the VHA Beneficiary Identification and Record Locator Subsystem File. This file is a fairly complete and accurate (95%) database because it is used to determine benefits to survivors of veterans [11].

The outcomes were the decline in PCS and MCS at the 18-month follow-up. The rationale for choosing ‘decline’ was because poor outcomes reflect potentially poor quality of care [12]. The cut points for ‘decline’ were based on two standard errors of a point-in-time score, a similar strategy used in the Medical Outcomes Study [3]. Decrements of more than −6.5 points in PCS scores have been associated with increased risk of both hospitalization and mortality [12]. Declines in MCS by more than −7.9 have been observed in those with clinical depression compared to those without depression [13]. Patients who died were aggregated in the outcome of decline in PCS but were excluded from MCS [3,14].

Choice of risk adjustors and data sources

We used three domains of risk adjustors: demographic and socioeconomic information, diagnoses, and baseline health status [15]. However, these concepts are not, strictly speaking, independent domains of risk. The specifications of the relationships among the different domains of risk and change in health status will be addressed in the rest of this section.

Sociodemographic data are considered proxies for pre-existing physiological reserve as well as for preferences and education, which may affect patients’ perception of illness and how they use the health care system. Information about age and gender was obtained from the VHA Outpatient Clinic File. Race, marital status, level of education, and employment status, on the other hand, were patients’ self-reported data. We also included a VHA administrative proxy for income, the means test, because the VHA uses it to classify patients based on their incomes (<$20,000 versus >$20,000) for consideration of (co)payments for some services such as medications and transportation to VHA facilities.

Diagnoses are the basis for understanding pathophysiology, choosing therapy, and predicting health outcomes [16]. We categorized diagnoses using the Comorbidity Index, which was developed to predict health status [17]. This diagnosis-based case-mix measure uses conditions that are commonly encountered in outpatient clinic visits. We calculated the Comorbidity Index scores as the unweighted count of 30 medical and 6 mental conditions. We used ICD-9-CM diagnoses from the VHA administrative files that were recorded in 1 year (between January 1997 and January 1998).

Baseline health status is a significant predictor of change in health status. Investigators have noted that high baseline functioning was a predictor of worsened functioning [18], whereas poor baseline health status was associated with subsequent improvement [19]. We used tertiles of the baseline PCS and MCS scores (‘high’, ‘intermediate’, and ‘low’) to minimize correlation errors with the outcome of change in health status and to improve the interpretation of the resulting beta coefficients.

Statistical analysis

Our first objective was to create clinically credible and statistically reliable risk-adjusted models to predict decline in PCS and MCS. We used a random sample of 2/3 of the study population (derivation sample of 21,215 subjects) to develop the models. This was done in several stages to examine how the selection of different dimensions of risk affects the
performance of the models. We only retained variables significant at the $P \leq 0.05$ level in the final risk adjustment models for the derivation sample. We then applied regression coefficients from those models onto the remaining 1/3 of the sample (validation sample of 10,608 patients). We used two measures to assess the performance of the models [15]. First, we calculated the c-statistics test, which reflects the predictive power of the models to discriminate among patients by ordering them according to rates of the outcome event. A c-statistic value of less than 0.5 indicates poor discriminatory power of the model. Second, we used the Hosmer–Lemeshow statistic test to evaluate the calibration of the model. Patients were divided into deciles, based on the expected risk for decline in PCS or MCS. Within each decile, the expected rate of worsening was compared with the observed rate of worsening. A $P$-value greater than 0.05 indicates a good fit.

For our second objective, we calculated the case-mix as the expected rates of decline in PCS and MCS at the network level. This was accomplished by using the multivariate regression models to calculate the expected outcomes for each patient in every network. Since patients who died were classified as ‘decline’ in PCS, we used a previously developed model to calculate the probability of death [20]. This was done because the population on which death can be assessed consists of all persons who completed the baseline survey, but the population on which MCS and PCS can be assessed consists of a different population of patients who completed a follow-up survey. The following formula was used to combine both probabilities to compute the expected decline rates: \[
\text{probability (death) + probability (decline in PCS)} \times (1 \text{ – probability (death)})
\]. We applied analysis of variance to test for differences in case-mix or expected rates among the 22 integrated service networks.

For our third objective, we first examined whether networks had higher or lower ‘worse’ rates than expected. To do so, we compared each network’s observed (unadjusted) rate of decline with the expected rate of decline. To calculate the observed decline in PCS rate, we used the following formula: mortality rates + (1 – mortality rates) \times \text{(decline in PCS rates)}. Second, we calculated adjusted rates of decline in PCS and MCS at the network level. The adjusted rate for a network was its observed rate divided by its expected rate, multiplied by the mean of the rates observed for all networks. Next we examined the number of networks that changed their rank after adjustment with special attention to the identification of outliers. We defined an outlier as being at least two standard errors above (better) or below (worse) the mean predicted change for the network average. Networks were blinded in this analysis for purposes of confidentiality.

Results

Similar to the VA outpatient clinical population, the study participants had a mean age of 64 (SD ± 12) years, 95% were male, 80% were white, 64.5% had less than a high school education, and 63% were married. The number of comorbid conditions was on average 4 (SD ± 2). The average baseline PCS and MCS scores were 32.7 (SD ± 11) and 42.8 (SD ± 13), respectively. These scores are substantially lower than the general US population in which the mean PCS and MCS are 50.

Among 31,823 veterans that were followed for 18 months, 1737 patients died (5.5%), 21,378 (67%) responded to a mailed follow-up survey, 4328 (13.6%) showed a decline in PCS scores greater than −6.5 points and 4322 (13.5%) had a decline in MCS scores by more than −7.9 points.

Table 1 summarizes the associations of individual patient characteristics and decline in PCS scores in the derivation sample. Older patients, those with a higher number of comorbidities, and those with a high baseline PCS were more likely to have a decline in PCS. In contrast, unmarried people, those employed, and those with high baseline MCS scores were less likely to have a decline in PCS. Race, gender, educational level, and income were not significant predictors.

Table 2 summarizes the associations of individual patient characteristics and decline in MCS in the derivation sample. Older patients, those employed, and those with higher education, income, and baseline PCS were less likely to have decline in MCS. In contrast, patients with higher Comorbidity Index scores (more comorbidities) and those with higher baseline MCS were more likely to have worse MCS. Gender, race, and marital status were not statistically significant predictors.

Table 3 summarizes the stepwise multidimensional risk analysis to predict decline in PCS scores greater than −6.5 points and in MCS scores greater than −7.9 points. The performance of the models improved by adding different dimensions of risk.

The final resulting model for decline in PCS scores included only the statistically significant variables from Table 1 (age, marital status, employment, Comorbidity Index, and baseline PCS and MCS). The c-statistic values of the model for decline in PCS scores were 0.73 and 0.72 in the derivation and validation samples, respectively. Model calibration was confirmed with a non-significant Hosmer–Lemeshow statistic test ($\chi^2 = 10.86$, $P = 0.21$, and $\chi^2 = 3.47$, $P = 0.90$ in the derivation and validation samples, respectively). The final resulting model for decline in MCS included only the statistically significant variables from Table 2 (age, marital status, education, income, employment, Comorbidity Index, and baseline PCS and MCS). The c-statistic values of the model for decline in MCS were 0.69 and 0.68 in the derivation and validation samples, respectively. Model calibration was confirmed with a non-significant Hosmer–Lemeshow test ($\chi^2 = 14.4$, $P = 0.07$) in the derivation sample. Although the Hosmer-Lemeshow test showed a $P$-value of 0.01 in the validation sample, the expected and the observed rates of decline in MCS were similar, indicating an acceptable fit.

Figures 1 and 2 show the networks’ case-mix, as measured by the expected decline in PCS and MCS rates (dots). Expected rates of decline in PCS by network ranged from 22.8 to 29.6%. The range of expected rates of decline in MCS was from 20.0 to 23.2%. Analysis of variance confirmed that these 22 integrated service networks differed significantly in their rates of expected decline in PCS and MCS ($P < 0.0001$), which indicates significant differences in case-mix. The bars in Figures 1 and 2 denote observed rates and the vertical lines with anchors indicate two
standard errors. When the expected rates (dots) are above or below the anchored line, the observed rates are significantly better or worse than expected. Most of the networks showed differences between the observed and expected rates of decline in PCS and MCS. Three networks (B, C, and U) had significantly \( P < 0.05 \) better than expected rates of decline in PCS and one (K) had significantly \( P < 0.05 \) better than expected rates of decline in MCS. In contrast, there were four networks
Table 3 The C-statistic and Hosmer–Lemeshow results for sequential models predicting decline in Physical Component Scale (PCS) scores and Mental Component Scale (MCS) scores in the derivation sample

<table>
<thead>
<tr>
<th>Models</th>
<th>Decline in PCS scores</th>
<th>Decline in MCS scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographics</td>
<td>C-statistic 0.54</td>
<td>C-statistic 0.54</td>
</tr>
<tr>
<td></td>
<td>$P = 0.15$</td>
<td>$P = 0.15$</td>
</tr>
<tr>
<td>Sociodemographics + Comorbidity Index</td>
<td>C-statistic 0.55</td>
<td>C-statistic 0.54</td>
</tr>
<tr>
<td></td>
<td>$P = 0.84$</td>
<td>$P = 0.18$</td>
</tr>
<tr>
<td>Sociodemographics + Comorbidity Index +</td>
<td>C-statistic 0.72</td>
<td>C-statistic 0.68</td>
</tr>
<tr>
<td>baseline PCS and MCS</td>
<td>$P = 0.91$</td>
<td>$P = 0.06$</td>
</tr>
</tbody>
</table>

1Covariates used in the models: Sociodemographics [(age per years, <50, 50–64, and 65>), gender (male/female), race/ethnicity (Whites/African Americans/Hispanics/Others), marital status (married/single/divorced-separated/widowed), level of education (<12 years/>12 years), employment status (employed/not), and income (<$20 000/>$20 000)], Comorbidity Index, baseline PCS (high, intermediate, high), and baseline MCS (high, intermediate, high).

Figure 1 Observed and expected rates of decline in Physical Component Scale (PCS) by integrated service network. The bars are two standard errors. *, $P < 0.05$.

Figure 2 Observed and expected rates of decline in Mental Component Scale (MCS) by integrated service network. The bars are two standard errors (SE). *, $P < 0.05$. 
A. J. Selim et al.

Table 4  Comparisons of unadjusted and adjusted decline in Physical Component Scale (PCS) and Mental Component Scale (MCS) rates: ranked from the lowest to the highest unadjusted rates of decline

<table>
<thead>
<tr>
<th>Networks</th>
<th>Unadjusted rates (%)</th>
<th>Adjusted rates (%)</th>
<th>Networks</th>
<th>Unadjusted rates (%)</th>
<th>Adjusted rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>23.8</td>
<td>27.3</td>
<td>B</td>
<td>19.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Q</td>
<td>24.4</td>
<td>25.3</td>
<td>U</td>
<td>19.4</td>
<td>6.7</td>
</tr>
<tr>
<td>H</td>
<td>24.5</td>
<td>25.9</td>
<td>F</td>
<td>19.6</td>
<td>19.9</td>
</tr>
<tr>
<td>R</td>
<td>24.5</td>
<td>25.5</td>
<td>A</td>
<td>19.9</td>
<td>7.8</td>
</tr>
<tr>
<td>B</td>
<td>25.0</td>
<td>23.4</td>
<td>K</td>
<td>19.9</td>
<td>18.8</td>
</tr>
<tr>
<td>T</td>
<td>25.2</td>
<td>25.9</td>
<td>T</td>
<td>20.0</td>
<td>19.3</td>
</tr>
<tr>
<td>U</td>
<td>25.3</td>
<td>23.7</td>
<td>L</td>
<td>20.4</td>
<td>19.6</td>
</tr>
<tr>
<td>F</td>
<td>25.6</td>
<td>29.2</td>
<td>S</td>
<td>20.6</td>
<td>19.9</td>
</tr>
<tr>
<td>V</td>
<td>25.6</td>
<td>24.7</td>
<td>V</td>
<td>20.7</td>
<td>19.3</td>
</tr>
<tr>
<td>K</td>
<td>25.9</td>
<td>26.5</td>
<td>D</td>
<td>20.7</td>
<td>21.1</td>
</tr>
<tr>
<td>L</td>
<td>26.0</td>
<td>24.4</td>
<td>J</td>
<td>20.9</td>
<td>20.8</td>
</tr>
<tr>
<td>O</td>
<td>26.0</td>
<td>27.0</td>
<td>C</td>
<td>21.4</td>
<td>22.1</td>
</tr>
<tr>
<td>D</td>
<td>26.1</td>
<td>25.0</td>
<td>E</td>
<td>21.5</td>
<td>19.8</td>
</tr>
<tr>
<td>G</td>
<td>26.1</td>
<td>29.7</td>
<td>N</td>
<td>21.5</td>
<td>20.1</td>
</tr>
<tr>
<td>C</td>
<td>26.4</td>
<td>23.3</td>
<td>G</td>
<td>22.3</td>
<td>21.3</td>
</tr>
<tr>
<td>A</td>
<td>26.6</td>
<td>24.8</td>
<td>H</td>
<td>22.3</td>
<td>23.1</td>
</tr>
<tr>
<td>J</td>
<td>26.7</td>
<td>27.4</td>
<td>Q</td>
<td>22.3</td>
<td>22.5</td>
</tr>
<tr>
<td>I</td>
<td>27.0</td>
<td>29.6</td>
<td>M</td>
<td>22.7</td>
<td>21.1</td>
</tr>
<tr>
<td>N</td>
<td>27.1</td>
<td>27.5</td>
<td>I</td>
<td>23.2</td>
<td>22.8</td>
</tr>
<tr>
<td>S</td>
<td>27.9</td>
<td>27.6</td>
<td>R</td>
<td>23.5</td>
<td>22.5</td>
</tr>
<tr>
<td>E</td>
<td>28.1</td>
<td>28.3</td>
<td>O</td>
<td>23.8</td>
<td>23.2</td>
</tr>
<tr>
<td>M</td>
<td>31.5</td>
<td>30.1</td>
<td>P</td>
<td>24.90</td>
<td>22.4</td>
</tr>
</tbody>
</table>

1 Covariates used in the model of decline in PCS scores: age (per years <50, 50–64, and 65+), marital status (married/single/divorced-separated/widowed), employment status (employed/not), comorbidities, baseline PCS (high, intermediate, high) and baseline MCS (high, intermediate, high). Covariates used in the mortality model: age (per years <50, 50–64, and 65+), gender (male/female), comorbidities, baseline PCS and MCS scores.

2 Covariates used in the model of decline in MCS: age (per years <50, 50–64, and 65+), marital status (married/single/divorced-separated/widowed), level of education (<12 years/>12 years), employment status (employed/not), and income (< $20,000/>$20,000), comorbidities, baseline PCS (high, intermediate, high) and baseline MCS (high, intermediate, high).

* Good outlier: below 2 standard errors.
¶ Bad outlier: above 2 standard errors.

(F, G, I, and M) with significantly (P < 0.05) worse than expected rates of decline in PCS and one (P) with significantly (P < 0.05) worse than expected rates of decline in MCS.

Table 4 summarizes the networks’ unadjusted and adjusted rates. Regarding adjusted rates of decline in PCS, three networks (B, C, and V) were identified as good outliers, whereas networks G, N, and I were identified as bad outliers. After risk adjusting rates of decline in MCS, two networks (G and P) were identified as bad outliers.

Discussion

Accurate information on health outcomes has become an expectation of regulatory and accreditation agencies. Important decisions, such as reimbursements and accreditations, will be based on perceived performance. Our study showed that it is feasible to develop clinically credible risk adjustment models with good statistical properties for the outcomes of decline in PCS and MCS in outpatient care. The resulting models produced an expected rate for each integrated service network, which we compared with its actual rate. There were significant case-mix differences across the 22 networks. Risk adjustment altered the assessment of network performance when compared to their unadjusted rates of decline in PCS and MCS.

The rates of decline in PCS and MCS in the VHA were similar to those in other population-based studies such as the Health Outcome Survey [21]. However, we found that the rates across networks differed significantly even after comprehensive risk adjustment. The fact that we found networks with lower-than-expected rates of decline in PCS and MCS, for example, suggests that some networks are doing better than expected. This opens the possibility of examining these networks to identify processes of care or management practices.
that may serve as models of best practices. In contrast, greater rates of decline are often attributed to poor quality of care and may serve to identify those networks that need to improve health care services through activities such as disease management programs or behavioral health practice guidelines.

In highlighting differences in case-mix across the integrated service networks, our study provides further evidence for the importance of risk adjustment, which made a difference for ranking most networks when compared with unadjusted rates. Some networks moved to lower ranks, meaning that their performance improved relative to other networks, whereas others moved to higher ranks, suggesting that they were not doing so well. In addition, we found that several networks had risk-adjusted rates of decline in PCS and MCS that were two standard errors below (better) or above (worse) the overall mean. Since this rate is a strong indicator of a problem, such findings could give networks a much better understanding of their performance and greater certainty about the need for quality improvement.

The associations among sociodemographic characteristics, diagnoses, and decline in PCS and MCS were consistent with the literature [3]. However, controlling for sociodemographics and comorbid illnesses explained only a fraction of the variance in the outcomes measured. The same is true in other studies [22]. Health status at baseline was a strong significant predictor of decline in PCS and MCS scores. Patients with higher baseline physical and mental health were more likely to have a decline in PCS and MCS scores compared with those with lower baseline physical and mental health, respectively.

We should note some limitations of this study. First, in patients who have low PCS or MCS scores at baseline, the VR-36 questionnaire may be insensitive to further decline, which would bias against documenting worsening health status in patients who are already severely ill (‘floor effect’). Against this is the finding of other investigators that over half of the patients with low health status reported that their health status subsequently declined [23]. In addition, our findings did not change when we altered the cut-points of change using those from the Medicare Health Outcomes Survey (decline in PCS and MCS scores greater than −5.7 and −6.7 points for PCS and MCS, respectively), thus leaving more room for decline. Second, we are missing follow-up information on some patients. This is unlikely to have affected our findings since baseline PCS and MCS scores between those with and without missing follow-up data showed only small differences (PCS = 33.4 ± 11 versus 32.6 ± 11 and MCS = 40.9 ± 13 versus 42.8 ± 13, respectively). The patients without follow-up data tended to be younger and therefore probably were less likely to have a decline in health status, which would bias our findings conservatively. Third, the c-statistics values suggest a modest predictive power of the models [15]. This represents an opportunity for future work that may lead to improved model performance and possibly further change in judgements of network performance.

In summary, this study advances our understanding of the clinical predictors of decline in PCS and MCS. We have developed a risk-adjustment model for decline in health status that can be used to assess this important outcome of ambulatory care. VHA as well as non-VHA initiatives can incorporate this methodology into their process of measuring and reporting performance of health care systems. Although not the focus of this study, future efforts can begin to identify the processes of care that may affect patient centered outcomes.

References


3. Bayliss EA, Bayliss MS, Ware JE, Steiner JF. Predicting declines in physical function in persons with multiple chronic medical conditions: what we can learn from the medical problem list? Health Qual Life Outcomes 2004; 2: 47.

4. Ware JE, Bayliss MS, Rogers WH et al. Differences in 4-year health outcomes for elderly and poor, chronically ill patients treated in HMO and fee-for-service systems. JAMA 1996; 276: 1039–1047.


Accepted for publication 10 September 2005