

Variations in risk-adjusted outcomes in a managed acute/long-term care program for frail elderly individuals

DANA B. MUKAMEL¹, DERICK R. PETERSON², ALINA BAJORSKA³, HELENA TEMKIN-GREENER³, STEPHEN KUNITZ³, DIANE GROSS³ AND T. FRANKLIN WILLIAMS⁴

¹University of California, Department of Medicine, Division of General Internal Medicine and Primary Care, Irvine, CA, and ²Department of Biostatistics, ³Department of Community and Preventive Medicine, and ⁴Department of Medicine, University of Rochester, Rochester, NY, USA

Abstract

Objective. To develop and investigate the properties of three performance measures based on risk-adjusted health outcomes for a frail, elderly, community-dwelling population enrolled in a managed, acute, and long-term care program.

Design. Retrospective analyses of an administrative dataset containing individual level records with information about socio-economics, health, functional and cognitive status, diagnoses, and treatments. We estimated risk-adjustment models predicting mortality, decline in functional status, and decline in self-assessed health. Each model includes individual risk factors and indicator variables for the program site in which the individual enrolled. Sites were ranked based on their performance in each risk-adjusted outcome, and the properties of these performance measures were investigated.

Setting. Twenty-eight sites of the Program of All-Inclusive Care for the Elderly (PACE) that provide primary, acute, and long-term care services under capitated Medicare and Medicaid payment to a nursing home certifiable, and functionally and cognitively frail community-dwelling elderly population.

Study participants. Three thousand one hundred and thirty-eight individuals who were newly enrolled between 1 January 1998 and 31 December 1999. The average age of these enrollees was 78 years, 27% were male, 50% were diagnosed with dementia, and they had approximately 4 Activities of Daily Living limitations and 7.4 Instrumental Activities of Daily Living limitations.

Main outcome measures. Risk-adjustment models, performance ranking for each site, and correlations between performance rankings.

Results. We present risk-adjustment models for mortality, change in functional status, and self-assessed health status. We found substantial variation across sites in performance, but no correlation between performance with respect to different outcomes.

Conclusions. The variations in outcomes suggest that sites can improve their performance by learning from the practices of those with the best outcomes. Further research is required to identify processes of care that lead to best outcomes.

Keywords: elderly, long-term care, PACE, quality indicators, risk-adjustment

Assessment of quality of medical care based on risk-adjusted health outcomes has become an important component of efforts to evaluate, regulate, and inform consumers about quality [1]. Examples include the New York State Cardiac Surgery reports, which provide risk-adjusted mortality rates for hospitals and surgeons, and the Nursing Home Compare website, which reports on 14 risk-adjusted outcomes. Much of the development work and reporting to date has focused on institutionalized patients in hospitals or nursing homes. In this

study we apply the risk-adjustment methodology that has been used in those settings to a frail, disabled population, living for the most part in the community with the support of a comprehensive system of care—the Program of All-Inclusive Care for the Elderly (PACE). We demonstrate that administrative data collected by each PACE site can be used to measure relative performance of sites based on risk-adjusted health outcomes. We examine properties of these measures and the implications for the quality of care that PACE provides its enrollees.

Address reprint requests to Dana B. Mukamel, Department of Medicine, University of California, Division of General Internal Medicine and Primary Care, Irvine, CA, USA. E-mail: dmukamel@uci.edu

Description of PACE

PACE offers comprehensive care to frail elderly individuals who are nursing home certifiable, but prefer to remain in the community. It is a managed care program, combining Medicare and Medicaid capitated payments to cover all the health care needs of its enrollees: primary, acute, and long-term care. For detailed descriptions of PACE see Bodenheimer [2] and Eng *et al.* [3]. There are currently >30 sites around the country serving ~10 000 individuals [4]. All sites share the goal of maintaining patients in the community as long as possible, through provision of in-home and day care services. All follow the same philosophy of care, emphasizing close relationships between the primary care providers and the patient, with care given by an interdisciplinary team.

Methods

Risk-adjusted health outcomes as measures of quality

Health outcomes are important measures of performance of medical care providers because they offer patients the information that is of most concern to them, i.e. how likely they are to get better if they receive care from a particular provider. Not all outcomes, however, are informative about quality. For example, mortality for a terminally ill patient is not a useful outcome measure because all patients are expected to die. For these patients, a measure based on pain control may be more relevant. To be useful, performance measures based on outcomes should meet several criteria [5]: the outcome should be of importance to patients and should contribute to their well-being; the outcome can be affected by care and thus influenced by the medical care provider; the outcome-based performance measure should account for differences in patient risks that the provider has no control over, such that programs that have worse outcomes because they are treating sicker patients are not penalized; and they should be calculated over a large number of patients, thus increasing the precision of the measure.

In this paper we investigate three health outcomes: mortality, change in functional status and self-assessed health status. These were chosen because they are central to the well-being of enrollees and because high quality medical care can be expected to impact favorably upon each of them. In fact, measures based on mortality [6] and changes in functional status [5,7] have been developed for use in nursing homes, which serve populations similar to PACE.

Patient health outcomes depend on two types of factors: patient risks and the quality of the care the patient receives [8]. Therefore, to be able to make inferences about quality of care from information about outcomes, one needs to separate the effect of patient risks and the effect of quality. This is typically done by statistically modeling the effect of patient risks and provider effect on the outcome, through estimation of risk-adjustment models. These models generally assume that differences in outcomes across providers that cannot be explained by differences in patient risks (i.e. the residual) reflect differences in quality of care. In this paper we estimate

hierarchical risk-adjustment models of the following general form:

$$\text{Outcome}_{i,j} = f(\alpha + \beta \text{PR}_{i,j} + \sum \gamma_j D_j + \epsilon_i) \quad (1)$$

where $\text{outcome}_{i,j}$ is the outcome experienced by patient i when treated in program j (e.g. functional status at 3 months post-admission), $\text{PR}_{i,j}$ is a vector of risk factors for patient i in program j (e.g. age, comorbidities) and D_j are indicator variables for each program j . The functional relationship between the outcome and patient risks may vary, as indicated by the function f . In this model, the estimated coefficients for the program indicator variables, i.e. the γ_j , can be interpreted as the effect of each program, relative to a baseline PACE program, on the outcome for a patient with the same risk factors at enrollment. Thus, the γ_j provide a measure of the quality of program j relative to other programs in the domain that is relevant to the specific outcome.

Sample

The study included participants from 28 of the 33 existing sites who enrolled between 1 January 1998 and 31 December 1999. The sample used for the mortality outcome included 3138 participants out of 4006 (78%). Excluded were those who left and re-entered (100), those who did not have health and functional assessment done within 31 days from enrollment (598), participants whose data indicated chemotherapy and no cancer (14), and those with more than three risk factors missing (156). When less than three risk factors were missing, values were imputed by the site-specific means of the observed values for each variable.

The measures based on changes in functional status and self-assessed health were based on a subsample of 1874 (68%) new enrollees who had a functional and health assessment done at 3 months following enrollment. This subsample was very similar to the full sample in terms of all risk factors, as can be seen in Table 1.

Data and variables

We obtained information about each participant from a patient level administrative database collected by all sites. All sites used the same variable definitions and all received training in data collection [9]. The dataset included information about enrollees' demographics, socioeconomic, health status and disability, medical history, utilization of health services, and date of death.

Activities of daily living (ADL) limitations included: bathing, dressing, grooming, toileting, transfer, walking, and feeding. Instrumental activities of daily living (IADL) included: meal preparation, shopping, housework, laundry, heavy chores, managing money, taking medications, and transportation. Cognitive status was measured by the number of errors in responding to the Short Portable Mental Status Questionnaire [10], which can range from 0 to 10. The self-assessed health information was based on a question asked of the participant with five possible responses: excellent (coded as 1), good, fair, poor (coded as 4), and not answered or missing.

Table 1 Risk factors hypothesized to influence the outcome and included in the initial analyses (means and standard deviations of continuous variables are shown in parentheses)¹

Risk factor present at enrollment	Percent of sample that experiences the risk factor	
	Sample for mortality outcomes	Sample for functional change and self-assessed health outcomes
Age (years)	78.57 (9.37)	78.19 (9.45)
Male	26.96%	27.37%
Living arrangements		
Living alone	33.15%	31.64%
Spouse	13.30%	12.77%
Nursing home	1.98%	1.95%
Home with relatives	41.05%	42.09%
Other	10.52%	11.55%
Ethnicity		
White	46.67%	43.61%
Black	30.95%	33.19%
Hispanic	14.78%	14.28%
Asian	6.54%	7.90%
Other	1.06%	1.02%
Education (years)	8.69 (4.21)	8.65 (4.20)
Activities of daily living limitations		
Bathing	82.88%	82.18%
Dressing	68.29%	68.35%
Grooming	71.69%	71.50%
Toileting	47.90%	48.61%
Transfer	44.82%	46.05%
Walking	50.54%	51.54%
Feeding	29.13%	29.60%
Instrumental activities of daily living limitations (<i>n</i>)	7.40 (1.23)	7.38 (1.27)
Incontinence		
Urinary	58.77%	58.82%
Bowel	27.59%	26.42%
Short Portable Mental Status Questionnaire (number of errors)	4.48 (3.13)	4.51 (3.10)
Diseases		
Cardiac	77.60%	77.21%
Respiratory	19.98%	17.93%
Cerebrovascular	29.13%	29.77%
Infections	3.25%	3.09%
Dementia	50.51%	50.11%
Depression	32.50%	31.96%
Cancer	10.77%	9.71%
Other	95.86%	95.41%
Behavior disorders		
Wandering	12.52%	12.17%
Verbal disruption	9.79%	9.50%
Physical aggression	5.73%	6.12%
Regressive behavior (e.g. childish behavior)	7.97%	7.97%
Treatment		
Parenteral medication or fluids	10.39%	—
Nursing treatment such as tube feeding, parenteral fluids, dialysis, and suctioning	—	8.70%
Oxygen daily	4.85%	—

continued

Table 1 *continued*

Risk factor present at enrollment	Percent of sample that experiences the risk factor	
	Sample for mortality outcomes	Sample for functional change and self-assessed health outcomes
Chemotherapy	0.42%	—
Dialysis	1.28%	—
Suctioning	0.22%	—
Self-assessed health		
Excellent	6.88%	7.47%
Good	38.18%	37.09%
Fair	29.38%	30.58%
Poor	11.85%	11.85%
Not reported	13.70%	13.02%
Sample size	3138 (430 deaths)	1874

¹We report separately for mortality and the two other outcomes because the sample for functional change and self-assessed health were a subset of the mortality sample.

²—, values are not shown because these variables were not considered to be a risk factor for the respective outcomes and were not included in the initial analyses.

For the model of functional status outcome, the outcome variable was defined as the change in the number of ADL limitations between enrollment and 3 months following enrollment. For the self-assessed health model, the outcome was self-assessed health at 3 months following enrollment, adjusted for self-assessed health at enrollment. This is similar to modeling the change in self-assessed health, but is more efficient in that it allows use of data on participants missing self-assessed health at enrollment, but not at 3 months after enrollment, without making any assumptions about where missing self-assessed health should fall on the ordered scale of self-assessed health at enrollment. For the mortality model, the outcome was the number of days from enrollment till death, censored at 31 December 1999.

Independent variables included individual characteristics (listed in Table 1) that were chosen by the project clinicians as those likely to be important determinants of the probability of each outcome. Also included were indicator variables for each one of the sites. These are the variables capturing the performance of each site, as discussed above. All risk factors were measured at the time of enrollment in the program.

Analyses

We estimated three regression models, one for each outcome. All models were based on individual level data (see equation 1 above). The probability of mortality was modeled with a semi-parametric Cox proportional hazards model, assuming a common but non-parametric baseline hazard function, so that the effects of site and participant level risks can be easily summarized by the relative hazards. The method of maximum partial likelihood was used to estimate the Cox regression parameters. Self-assessed health at 3 months and change in functional status were modeled linearly, using ordinary least squares.

We initially fit models that included all risk factors hypothesized to influence the outcome. For those predictors that were significant at the two-sided significance level of 0.15, we tested for potential interactions with all other risk factors. The final models we present are those that include only risk factors that were significant at the two-sided significance level of 0.05. An *F*-test, or a likelihood ratio test in the case of the Cox model, was used to test the hypothesis that all non-significant variables were also jointly non-significant, and can therefore be excluded from the models.

Results

Table 1 shows averages for all variables used in the initial analyses. Most variables were included in the initial estimation of all three models. The only exceptions were several treatment variables. We have included only treatment variables that the project clinicians considered unlikely to be discretionary and therefore not subject to variations in practice styles.

Tables 2–4 present the estimated models for each of the outcomes, showing only the estimates for the individual risk factors. Table 2 presents mortality hazard ratios for each risk factor. For example, male gender has a hazard ratio of 1.50, indicating that the risk of death of a male enrollee is 50% higher than for a female, the reference group. The risk of mortality increases with age, being male, being Asian or white, limitations in toileting or walking, more than seven IADL limitations, cognitive dysfunction as measured by a score >8 on the Short Portable Mental Status Questionnaire, poor self-assessed health, bowel incontinence, cancer, diabetes, renal failure, and daily use of oxygen. The risk of mortality is lower

Table 2 Mortality risk-adjustment model¹

Risk factors measured at enrollment	Hazard ratio ²	P-value
Age (years)	1.03	0.001
Male	1.50	0.001
Ethnicity (reference group includes blacks, hispanics, and other)		
Asian or white	1.54	0.001
Activities of daily living limitations: toileting and/or walking	1.86	0.001
Instrumental activities of daily living limitations >7	1.26	0.070
Short Portable Mental Status Questionnaire >8	1.33	0.040
Self-assessed health ³	1.13	0.006
Bowel incontinence	1.22	0.068
Patient with diabetes	1.28	0.021
Patient with renal failure	1.71	0.001
Cancer diagnosis and chemotherapy (reference group has no cancer and no chemotherapy)		
Patients with cancer, no chemotherapy	1.89	0.001
Patients with cancer and chemotherapy	11.71	0.001
Daily oxygen	2.59	0.001
Patient is wandering	0.68	0.019

¹Estimated model included site indicator variables, not shown.

²The hazard ratio is the relative risk of mortality per unit increase in the variable. For example, a male is 50% more likely to die than a female.

³Self-assessed health in this model is defined as a variable that can take the values of 1 = excellent, through 4 = poor and 5 = missing.

Table 3 Functional decline (within 3 months of admission) risk-adjustment model¹

Risk factor measured at enrollment	Coefficient ²	P-value
Age (years)	0.01	0.008
Living arrangement (reference group = living alone):		
With spouse or other relatives	0.23	0.002
In a nursing home	0.76	0.002
Total number of activities of daily living limitations	-0.28	0.001
Total number of instrumental activities of daily living limitations (reference group = 8 instrumental activities of daily living limitations)		
<7	-0.43	0.001
7	-0.26	0.004
Short Portable Mental Status Questionnaire	0.05	0.001
Patient is wandering	-0.25	0.019
Bladder incontinence	0.16	0.034
Bowel incontinence	0.24	0.004
Patients with cancer diagnosis	0.23	0.037

¹Estimated model included site indicator variables, not shown. The dependent variable is defined as the difference between number of activities of daily living at 3 months and enrollment. Thus, higher values indicate decline in functional status.

²The coefficient indicates the average change in number of activities of daily living limitations per unit increase in the variable. For example, the average participant with bowel incontinence will have an average increase of 0.24 activities of daily living limitations compared with a participant who is continent.

for enrollees who exhibit wandering behavior. The mortality model explained 12.1% of the variation (based on a normalized pseudo R^2).

Table 3 shows the model predicting changes in functional status within 3 months. The model explained 16.7% of the variation. A positive coefficient indicates an increase in

Table 4 Self-assessed health at 3 months post-enrollment risk-adjustment model¹

Risk factor measured at enrollment	Coefficient ²	P-value
Self-assessed health at admission (reference group has self-assessed health = excellent)		
Self-assessed health = good	0.33	0.001
Self-assessed health = fair	0.66	0.001
Self-assessed health = poor	1.02	0.001
Self-assessed health = missing	0.54	0.001
Age (years)	-0.01	0.003
Female	0.10	0.013
Education (years)	-0.01	0.048
Dementia	-0.09	0.018
Depression	0.10	0.014

¹Estimated model included site indicator variables, not shown.

²The coefficient indicates the average self-assessed health status at 3 months (on a scale from 1 = excellent to 4 = poor) per unit increase in the variable. For example, the average participant with self-assessed health = 2 (good) at admission will have a self-assessed health state at 3 months of 0.33 higher compared with an individual entering with self-assessed health = 1 (excellent). An individual entering with depression will have an increase of 0.10 in his/her self-assessed health status (i.e. they will perceive their health status as worse at 3 months) compared with a participant who is not diagnosed with depression.

number of ADL limitations, i.e. a deterioration in functional status. For example, living in a nursing home increases ADL limitations at 3 months by 0.76 compared with an enrollee who lives alone. Functional status at 3 months is likely to deteriorate for older individuals, those living with others or in a nursing home, and those with cognitive dysfunction, bowel or bladder incontinence, or cancer. Individuals entering the program with a higher number of ADL limitations are less likely to experience decline (a ceiling effect), as are those who wander.

Table 4 presents the model predicting self-assessed health at 3 months after enrollment. The model explained 18.5% of the variation. Higher values for self-assessed health indicate poorer health status. Thus, the positive coefficient of 0.10 for depression indicates that an enrollee diagnosed with depression at enrollment will report worse health status at 3 months compared with an enrollee without this diagnosis. Self-assessed health is likely to be better for those reporting a better health status at admission, those who are older, males, have higher education, and are diagnosed with dementia. Those with depression are likely to report a worse self-assessed health state at 3 months. Those with missing self-assessed health at enrollment are also likely to report worse self-assessed health at 3 months compared with those with excellent health at admission. Their outcome is similar to those reporting their self-assessed health at admission as fair. Self-assessed health status at enrollment has a much larger effect on the outcome than other individual risk factors.

In all three models, the hypothesis that program indicators can be excluded from the model was rejected ($P = 0.0001$ for

mortality, $P = 0.0057$ for functional status, and $P = 0.014$ for self-assessed health). The incremental variation explained by introducing program indicator variables to a model with patient risks only (as measured by R^2 or the partial likelihood) was 22% for mortality, 14% for functional status, and 13% for self-assessed health. These data suggest that outcomes depend not only on individual patient risks, but also on the program in which they are enrolled.

Table 5 reports the risk-adjusted outcome-based performance measure for each site for all three outcomes (based on the coefficient for the site indicator variables in the respective models). The table also indicates which are significantly different from the average.

Several interesting observations emerge from inspection of Table 5. Firstly, there is substantial variation in performance across sites. Mortality hazard ratios vary from a low of 0.12 to a high of 2.37, implying that risk of death is lower by 88% compared with the average in the site with the lowest rate, and exceeds the average by 137% in the site with the highest rate. The range for changes in functional status is from an average improvement of almost half an ADL (-0.42) to a decline of more than half (+0.72). The variation in changes in self-assessed health ranges from -0.27 at the program with the most improvement to +0.31 in the program with the most decline.

Secondly, the number of statistical outliers is greater than might be expected by chance alone. With 28 sites, and defining statistical outliers at the 0.1 (two-tailed) significance level, we would expect 2.8 programs to be flagged as potentially false outliers. We find, however, seven outliers for mortality, six for functional status, and six for self-assessed health. This also suggests that sites indeed differ in their performance and that the variation is not solely statistical in nature.

Thirdly, the site with the best (or worst) rate in one outcome is not the same as the site with the best (or worst) rate for another outcome. In fact, there is no association between performance in one area and that in another. Table 5, part II demonstrates this by showing the ranking of each program based on its performance measures reported in Table 5, part I. Spearman's rank correlation (which ranges between -1 and 1, with -1 indicating perfect inverse correlation, 0 indicating no correlation, and 1 indicating perfect correlation) between the mortality ranking and the functional performance ranking was -0.065 with a P -value of 0.734, between mortality and self-assessed health it was 0.274 with a P -value of 0.155, and between functioning and self-assessed health it was -0.304 with a P -value of 0.113.

Discussion

In this paper we present three measures of performance of PACE programs based on risk-adjusted health outcomes. It is important to first consider the validity of these measures. Validity ideally should be determined by comparing these measures to a gold standard of care. However, a gold standard does not exist. When a gold standard does not exist we typically rely on assessment of the face validity and content validity of

Table 5 Site average outcomes and ranking, based on models adjusting for individual participant risks

Site ¹	Mortality ²	Functional status change ³	Self-assessed health at 3 months ⁴
I. Site average outcomes ⁵			
A	0.12 ⁷ (best)	-0.004	0.145
B	0.29	-0.128	-0.085
C	0.35	0.410 ⁷	-0.078
D	0.43 ⁶	0.716 ⁸ (worst)	-0.114
E	0.62	0.002	-0.168 ⁶
F	0.68	-0.109	0.000
G	0.83	-0.164	-0.114
H	0.88	-0.196 ⁷	0.140 ⁷
I	0.88	0.250	0.020
J	0.95	-0.002	-0.210
K	0.99	0.015	0.061
L	1.04	0.067	0.102
M	1.11	-0.203	0.037
N	1.11	-0.007	-0.26 ⁷
O	1.17	0.167	-0.270 ⁸ (best)
P	1.19	-0.118	0.061
Q	1.24	-0.418 ⁶ (best)	0.092
R	1.34	0.037	-0.006
S	1.35	-0.194	0.307 ⁸ (worst)
T	1.36	-0.148	-0.171 ⁷
U	1.44	0.087	-0.006
V	1.48	0.020	-0.075
W	1.48	-0.319 ⁷	0.031
X	1.83 ⁶	-0.089	0.015
Y	2.13 ⁸	0.145	0.072
Z	2.16 ⁸	0.324 ⁶	0.028
AA	2.16 ⁸	-0.260	0.195
BB	2.37 ⁸ (worst)	0.119	0.066
II. Site rankings ⁵			
A	1 ⁷ (best)	14	26
B	2	9	7
C	3	27 ⁷	8
D	4 ⁶	28 ⁸ (worst)	6
E	5	16	4 ⁶
F	6	11	13
G	7	7	5
H	8	5 ⁶	25 ⁷
I	9	25	15
J	10	15	10
K	11	17	19
L	12	20	24
M	13	4	18
N	14	13	2 ⁷
O	15	24	1 ⁸ (best)
P	16	10	20
Q	17	1 ⁶ (best)	23
R	18	19	12
S	19	6	28 ⁸ (worst)
T	20	8	3 ⁷
U	21	21	11
V	22	18	9

continued

Table 5 *continued*

Site ¹	Mortality ²	Functional status change ³	Self-assessed health at 3 months ⁴
W	23	2 ⁷	17
X	24 ⁷	12	14
Y	25 ⁸	23	22
Z	26 ⁸	26 ⁶	16
AA	27 ⁸	3	27
BB	28 ⁸ (worst)	22	21

¹Sites are sorted by ranking for mortality outcome.

²Mortality hazard ratio for the average participant in a site relative to the average for all sites.

³Change in activities of daily living limitations for the average participant in a site relative to the average for all sites.

⁴Self-assessed health at 3 months post-enrollment for the average participant in a site relative to the average for all sites.

⁵As all the outcomes are adverse (e.g. decline in functional or self-assessed health status), high positive values indicate worse outcomes than lower values.

Ranking significantly different from the average: ⁶ $0.1 > P > 0.05$; ⁷ $0.05 > P > 0.01$; ⁸ $P < 0.01$.

the measures [11]. Face validity relates to whether an indicator seems to measure the domain of interest. Content validity addresses the question of whether a measure contains all the relevant items of interest. The measures we developed here have face validity because they meet the four criteria discussed at the beginning of the Methods section [5]. Their content validity is derived from the large number of risk factors that were considered in the initial analyses, although some risk factors may be limited (e.g. depression tends to be underdiagnosed). Another aspect of validity, construct validity, which is based on demonstrating the link between the measures and processes of care, is outside the scope of this study. Thus, these measures should not be viewed as fully validated and should not be used in public reporting of quality of individual sites. They can, however, be useful in guiding efforts to understand better the performance of PACE and to identify opportunities for improvement.

The variations we find in these three health outcomes suggest that there is room for improvement in the care practiced in different sites. Programs with the best risk-adjusted outcomes can serve as best practice models for others, thus offering the opportunity for all to achieve better outcomes.

Translating the variations in outcomes that we find into practical guidelines that sites can adopt in order to bring about improvement requires an understanding of the structure and process of care, and of the relationships between them and outcomes. Several approaches to making these linkages exist. Firstly, clinical and statistical studies of the relationships between structure, process, and outcomes can identify best practices. For example, studies of the relationships between the characteristics of the care team, considered to be one of the hall marks of the PACE model, and outcomes would help identify ways in which teams can change their operations to improve enrollees' outcomes.

A second approach, which has been in place for some time now, is to encourage sharing of successful experiences. The National PACE Association holds meetings twice a year. Most sites send individuals, including the administrators,

medical directors, directors of nursing, and others. These meetings offer opportunities to share experiences, both formally through presentations at scientific meetings, and informally through interactions between meeting participants. To enhance these activities and focus them on quality improvement possibilities, the National PACE Association may wish to consider specific sessions featuring sites with the best outcomes.

Another intervention that has been shown to be effective in improving outcomes for coronary artery bypass graft surgeries is to learn by observation. The Northern New England Cardiovascular Disease Study Group conducted an intervention study, where participating hospitals sent teams to observe the process in other participating hospitals [12]. These visits, in conjunction with training in quality improvement techniques and feedback of outcome data to each cardiac surgeon, have led to a decrease in risk-adjusted mortality rates in these hospitals. PACE could adopt a similar approach, sending personnel to visit and observe practices in those sites with the best outcomes.

Implementing a program of learning from the best may be complicated by the fact that, as suggested by the data, there is no single best program. Those that excel in terms of mortality are not the ones that excel in terms of functional decline or in terms of self-assessed health. This lack of correlation between performance measures for the different outcomes has been found in other settings as well: in nursing homes [13], hospitals [14], and ambulatory care clinics [15]. Recognizing this multidimensionality of quality, efforts at improvement also need to be multifaceted, with lessons learned from the best in each dimension of performance. Furthermore, research is needed to understand the interactions between processes of care that lead to different outcomes and the reasons that they may not be correlated. For example, for mortality outcomes, information about advance directives may be important.

One additional caveat is noteworthy. In this study we evaluate short-term outcomes, within 2 years of enrollment for mortality and within 3 months for the two other outcomes. Programs' ability to maintain health over longer periods may

be different, as different processes of care might be required for maintenance as opposed to affecting change within a short time after enrollment. Thus, this analysis does not provide information about long-term outcomes in PACE and a separate assessment is required to evaluate this type of care.

Acknowledgements

The Authors gratefully acknowledge financial support from the National Institute on Aging (grant no. RO1 AG1755).

References

1. Donabedian A. The quality of care: how can it be assessed? *JAMA* 1988; **260**: 1743–1748.
2. Bodenheimer T. Long-term care for frail elderly people—the On Lok Model. *New Engl J Med* 1999; **341**: 1324–1328.
3. Eng C, Pedulla J, Eleazer GP, McCann R, Fox N. Program of All-inclusive Care for the Elderly (PACE): an innovative model of integrated geriatric care and financing. *J Am Geriatr Soc* 1997; **45**: 223–232.
4. Gross D, Temkin-Greener H, Kunitz S, Mukamel DB. The growing pains of integrated health care for the elderly: lessons from the expansion of PACE. *Milbank Mem Fund Q* 2004; **28**: 257–282.
5. Mukamel DB. Risk adjusted outcome measures and quality of care in nursing homes. *Med Care* 1997; **35**: 367–385.
6. Mukamel DB, Spector WD. Nursing home costs and risk adjusted outcome measures of quality. *Med Care* 2000; **30**: 78–95.
7. Rosen A, Berlowitz DR, Anderson JJ, Ash AS, Kazis L, Moskowitz MA. Functional status outcomes for assessment of quality in long-term care. *Int J Qual Health Care* 1999; **11**: 37–46.
8. Mukamel DB, Dick A, Spector WD. Specification issues in measurement of quality of medical care using risk adjusted outcomes. *J Econ Soc Measure* 2000; **26**: 267–281.
9. Mukamel DB, Temkin-Greener H, Clark M. Stability of disability among PACE enrollees: financial and programmatic implications. *Health Care Financ Rev* 1998; **19**: 83–100.
10. Pfeiffer EA. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc* 1975; **23**: 433–441.
11. Iezzoni LI. *Risk Adjustment for Measuring Health Care Outcomes*, 1st edition. Ann Arbor, MI: Health Administration Press, 1994.
12. O'Connor GT, Plume SK, Olmstead EM *et al*. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA* 1996; **275**: 841–846.
13. Mukamel DB, Brower CA. The influence of risk adjustment methods on conclusions about quality of care in nursing homes based on outcome measures. *Gerontologist* 1998; **38**: 695–703.
14. Rosenthal GE. Weak association between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Publ Health* 1997; **87**: 429–433.
15. Gandhi TK, Cook EF, Puopolo AL, Burstin HR, Haas JS, Brennan TA. Inconsistent report cards. assessing the comparability of various measures of the quality of ambulatory care. *Med Care* 2002; **40**: 155–165.

Accepted for publication 14 April 2004

