

# Why Health Care Process Performance Measures Can Have Different Relationships to Outcomes for Patients and Hospitals: Understanding the Ecological Fallacy

John W. Finney, PhD, Keith Humphreys, PhD, Daniel R. Kivlahan, PhD, and Alex H. S. Harris, PhD

Relationships between health care process performance measures (PPMs) and outcomes can differ in magnitude and even direction for patients versus higher level units (e.g., health care facilities). Such discrepancies can arise because facility-level relationships ignore PPM–outcome relationships for patients within facilities, may have different confounders than patient-level PPM–outcome relationships, and may reflect facility effect modification of patient PPM–outcome relationships. If a patient-level PPM is related to better patient outcomes, that care process should be encouraged. However, the finding in a multilevel analysis that the proportion of patients receiving PPM care across facilities nevertheless is linked to poor hospital outcomes would suggest that interventions targeting the health care facility also are needed. (*Am J Public Health*. 2011; 101:1635–1642. doi:10.2105/AJPH.2011.300153)

Many health care performance measures (PMs) quantify the extent to which a process of care that has been shown to cause or predict positive outcomes among participants in empirical studies is applied to patients in health care facilities (e.g., hospitals).<sup>1,2</sup> Such process performance measures (PPMs) are implemented on the assumption that processes of care linked to positive patient outcomes in clinical trials and other research will be associated with positive facility-level outcomes when the PPM is aggregated to the facility level as the proportion of patients receiving the PPM care.

For example, it might be assumed that if coordinated care for a particular medical condition is linked to better outcomes for patients in randomized controlled trials, then health care facilities with higher levels of coordinated care for targeted patients should have higher proportions of patients with good outcomes. However, researchers who have investigated hospital- or facility-level PPM–outcome relationships sometimes have found that facility rates of PM-specified care are unrelated or only weakly related to facility-level outcomes.

One example was reported by Bradley et al.,<sup>3</sup> who examined National Quality Forum PPMs for treating patients with acute myocardial

infarction. They found that higher rates of provision of the practices recommended by the National Quality Forum were at best only modestly related to lower hospital-level, risk-adjusted 30-day mortality rates among acute myocardial infarction patients from more than 900 hospitals. Werner and Bradlow<sup>4</sup> conducted a more comprehensive analysis of data from approximately 3600 acute care hospitals. Their findings showed that facilities in the top and bottom quartiles in terms of proportion of patients receiving processes of care recommended (by the Centers for Medicare & Medicaid Services and the Joint Commission on Accreditation of Healthcare Organizations) for acute myocardial infarction, heart failure, and pneumonia differed only slightly in risk-adjusted 30-day and 1-year mortality rates. (At the facility and practice level, Lehrman et al.<sup>5</sup> and Sequist et al.<sup>6</sup> also found weak relationships between clinical care quality and patient satisfaction.)

Bradley et al.<sup>3</sup> and Werner and Bradlow<sup>4</sup> pointed to a variety of factors that may have accounted for the weak facility-level associations, including potential facility-level confounding factors (e.g., patient safety processes) and restricted variation across hospitals in the provision of certain practices (e.g., providing

aspirin at admission to patients with acute myocardial infarction symptoms). Although Werner and Bradlow called for PPMs that are more strongly related to patient outcomes, neither they nor Bradley et al. considered that, even though facility-level performance on these care processes was only weakly associated with aggregated facility outcomes, patients who received this type of care may have had significantly better outcomes than patients who did not.

We considered this apparent paradox in the context of the methodological literature on the ecological fallacy and cross-level bias that have been the focus of considerable work in such fields as epidemiology and sociology,<sup>7–15</sup> as well as the highly relevant statistical literature on multilevel analysis<sup>16–18</sup> that has been applied to address other issues in health care research,<sup>19–21</sup> including linking patient care processes to outcomes.<sup>22</sup> However, these issues have received scant attention in the quality literature on PPM–outcome relationships.<sup>3–6</sup>

We used data on a PM for treatment retention among patients with substance use disorders to examine differences in patient- and facility-level PPM–outcome relationships.<sup>23</sup> We then examined how the literature on cross-level bias and multilevel analysis can explain otherwise puzzling differences in PPM–outcome relationships at different levels of analysis. Our goal was to explain these issues in a nontechnical way that is accessible to researchers, quality managers, clinicians, and health care facility directors.

Finally, we considered the implications for health care quality management and research of variations in PPM–outcome relationships at different levels of analysis. PM-specified processes of care that are related to outcomes at the patient level, but not the facility level, nevertheless may be defensible and valuable

in a health care system's quality improvement program.

## THE ECOLOGICAL FALLACY

The assumption that a patient-level PM–outcome relationship will be replicated when data are aggregated to the facility level (or that a facility-level relationship will be replicated in patient-level data) is an example of homology, a term Hannan<sup>24</sup> used to refer to the assumption that processes function in the same way at 2 or more levels of analysis or organization. The assumption of homology in relationships between health care quality and outcomes is bolstered by the fact that facility-level PPMs and outcome variables often are simple mathematical functions (proportions, means) of corresponding patient-level variables such as individual patient receipt or nonreceipt of coordinated care aggregated to the proportion of patients receiving such care at different health care facilities and individual patient outcomes (good or poor) aggregated to the proportion of patients at different facilities with good outcomes. Nevertheless, PPM–outcome relationships are not necessarily the same at the patient and facility levels.

Discrepancies in relationships at different levels of analysis were most prominently brought to the attention of social scientists by Robinson.<sup>13</sup> Using 1930 US census data, he found that the state-level correlation between the percentage of foreign-born adult residents and the English-language illiteracy rate was  $-0.53$ . In other words, states with higher percentages of foreign-born residents had lower illiteracy rates. However, the individual-level correlation (ignoring state of residence) between foreign-born status and English illiteracy was positive ( $0.12$ ). Foreign-born individuals were more likely to be illiterate in English. The unexpected negative correlation at the state level reflected the fact that foreign-born individuals had tended to settle in states where literacy rates among native-born citizens were higher.

The problem with inferring individual-level relationships from aggregated data was eventually labeled the ecological fallacy. More recent considerations of this issue<sup>7,12,15,25</sup> have been more even-handed, addressing problems of inference in moving from both higher- to lower-level units and from lower- to higher-level units and pointing out the incompleteness of

single-level analyses when individuals are clustered within higher-level units. These discussions highlight the need for multilevel or mixed-effects (e.g., effects at the patient and hospital levels) analyses<sup>17,20,26</sup> to better capture and understand PPM–outcome relationships across and within health care facilities.

## AN EMPIRICAL, ILLUSTRATIVE EXAMPLE

Recent work in developing and validating continuing care PMs for patients with substance use disorders<sup>23,27</sup> provided us an opportunity to explore these issues empirically. Continuing care refers to a period of outpatient care following an episode of acute treatment, such as residential treatment or initial, usually more intensive, outpatient treatment. Continuing care for substance use disorders has been related to positive outcomes (e.g., a higher likelihood of abstinence from alcohol and other drugs) in observational studies and in some randomized controlled trials.<sup>28</sup> Given such findings, the US Department of Veterans Affairs (VA) adopted a continuing care PM for its facilities based on the proportion of new patients in treatment for a substance use disorder who had at least 2 outpatient visits during each of 3 consecutive 30-day periods.<sup>23</sup>

For this illustrative analysis, we focused on a subgroup of 1485 patients undergoing substance use disorder treatment at VA facilities

who had been identified by Harris et al.<sup>23</sup> as nonabstinent from alcohol or drugs in the 30 days before a baseline assessment. The patients were drawn from nonmethadone substance use disorder programs at 72 VA facilities and had been followed an average of 7.3 (SD=2.4) months subsequently, at which point abstinence for the preceding 30 days was assessed. The follow-up rate was slightly above 65%. Outcome data for those not followed were imputed from multiple variables assessed at baseline.<sup>23</sup>

Because our goal was simply to illustrate a statistical phenomenon with real data, we randomly selected 1 of 5 data sets with imputed missing data that were averaged in the analyses of Harris et al.<sup>23</sup> We ignored the fact that at some facilities patients were drawn from more than one substance use disorder program and at other facilities not all programs' patients were sampled, and we did not control for patient baseline characteristics or other confounders. As highlighted by this last point, although we occasionally refer to the effects of PM-specified care, we (similar to most researchers investigating linkages between PPMs and clinical outcomes) are not describing causal relationships.

Using these data, we present findings from 3 analyses (Table 1). Analysis 1 was a facility-level regression analysis that related facility performance on the continuing care PM (percentage of a facility's substance use disorder patients meeting the PM) to facility abstinence rate

**TABLE 1—Results of 3 Analyses Linking Facility-Level or Patient-Level (Within-Facility) Continuing Care Performance Measures to Abstinence at Follow-Up: United States, 2001–2003**

	b (SE)	P	OR (95% CI)
Analysis 1: facility-level analysis (n = 72 facilities)			
Intercept	0.447 (0.023)	<.001	...
Facility patients receiving continuing care, %	0.045 (0.065)	NS	...
Analysis 2: patient-level analysis controlling for facility clustering (n = 1485 patients)			
Intercept	-0.323 (0.063)	<.001	...
Patient continuing care (yes/no)	0.558 (0.115)	<.001	1.75 (1.40, 2.19)
Analysis 3: multilevel analysis of facility- and patient-level (within-facility) relationships			
Intercept	-0.193 (0.100)	<.056	...
Patient continuing care (yes/no)	0.647 (0.127)	<.001	1.91 (1.49, 2.45)
Facility patients receiving continuing care, %	-0.509 (0.305)	<.1	0.60 (0.33, 1.09)

Note. CI = confidence interval; NS = nonsignificant; OR = odds ratio.

(percentage of a facility's substance use disorder patients abstinent for at least 30 days at follow-up), with each facility weighted by its number of patients. The facility-level relationship was near zero and nonsignificant (Table 1). In other words, facilities with higher levels of continuing care performance were no more likely to have high abstinence rates than were facilities with lower levels of continuing care performance.

Analysis 2 was a mixed-effects logistic regression analysis. It was a mixed-effects analysis because it focused on patient data (one level) while controlling for the clustering of patients at facilities (a second level); clustering in this context refers to patients from the same facility being more likely to be similar to each other on an outcome than they are to patients at other facilities. The analysis was a logistic regression because our dependent variable, abstinence, was dichotomous at the patient level. Analysis 2 yielded a significant positive relationship (odds ratio [OR]=1.75) between receipt of PM-specified continuing care and abstinence at follow-up (Table 1). In other words, the odds of patients who received PM-specified continuing care being abstinent were 1.75 times greater than the odds of abstinence among patients who did not receive the specified continuing care.

Finally, when both individual-level receipt of continuing care and the facility percentage of patients receiving continuing care were entered as predictors of patient abstinence in a mixed-effects logistic regression analysis (analysis 3; Table 1), patient-level continuing care had a slightly stronger significant association (OR=1.91) with abstinence at follow-up than it did in analysis 2, whereas facility-level continuing care had a negative (OR=0.60), although not statistically significant ( $P<.10$ ), relationship with patient abstinence. That is, after control for the relationship of patient receipt of continuing care to abstinence, receiving continuing care at a facility at which a higher proportion of patients received continuing care was associated (not quite significantly) with a lower likelihood of patient abstinence. Thus, unlike analysis 2, analysis 3 not only controlled for the clustering (greater similarity) of patients on outcomes by facility but also controlled for the relationship between facility and provision of PM-specified care.<sup>22</sup>

The stronger positive relationship of individual patient receipt of PM continuing care to abstinence in analysis 3 than in analysis 2 reflects the negative effect of facility continuing care performance in analysis 3. In fact, the effect of patient continuing care receipt in analysis 2 is a weighted average of the within-facility (patient-level) and between-facility (proportion of patients receiving continuing care) effects in analysis 3.<sup>18,29</sup> Analysis 3 makes it possible to distinguish independent patient-level and facility-level effects of PM-specified care that remain entangled in the analysis 2 estimate of the impact of continuing care.<sup>29</sup> In doing so, analysis 3 estimates the effect of patient-level receipt of PM-specified care independent of the effect of the proportion of facility patients receiving that care. Similarly, the analysis estimates the effect of facility PPM performance independent of the patient-level effect of receiving PM-specified care. Thus, 2 separate effects are addressed in this type of analysis.

It is not possible from these results to know why, after accounting for patient receipt of continuing care, patients at facilities with higher percentages of patients receiving continuing care tended to have poorer outcomes than those at facilities with lower percentages of patients receiving continuing care. One possibility is that the facilities that expended considerable clinical staff resources to meet the continuing care PM (e.g., in trying to locate patients who had dropped out) had fewer resources to provide high-quality continuing care. Another possibility is that facilities performing better with respect to continuing care had fewer self-help groups available in the community that could have reinforced patient abstinence. To move beyond such speculations would require additional research to determine factors that are related to both facility-level continuing care performance and abstinence.

### REASONS FOR DIFFERENT RELATIONSHIPS AT DIFFERENT LEVELS OF ANALYSIS

We have shown that the relationship between a continuing care PM and an outcome differed across a facility-level analysis, a patient-level analysis that controlled for clustering of patients within facilities, and an analysis in which both patient-level continuing care receipt and

facility-level percentage of patients receiving continuing care were used to predict patient abstinence. Although incongruity in patient- and facility-level associations between health care practices and outcomes is not a new finding,<sup>30,31</sup> to our knowledge the literature on ecological (aggregation) bias and multilevel analysis has not been applied to explain why such cross-level differences in PPM–outcome relationships can arise. For example, Horn<sup>32</sup> mentioned the ecological fallacy in passing in commenting on factors that might have accounted for the weak or negative facility-level PPM–mortality relationships reported by Bradley et al.<sup>3</sup> and Werner and Bradlow,<sup>4</sup> but she did not explain why relationships might differ at different levels of analysis.

Among the reasons for incongruence in patient- and facility-level relationships are loss of information within higher-level units,<sup>8,11</sup> different confounders at different levels of analysis,<sup>14,15,33</sup> and effect modification.<sup>10</sup> We discuss each of these issues but provide more attention to the first—information loss with aggregation—because it helps in understanding and can reflect the other 2.

### Loss of Information Within Higher-Level Units

Analyses of relationships involving only aggregated (e.g., health care facility level) data do not incorporate information on the relationship of the PPM and the outcome variable (e.g., abstinence or risk-adjusted mortality) for the individuals within each health care facility.<sup>11</sup> To illustrate this phenomenon, Table 2 draws on examples from Greenland<sup>8</sup> and Greenland and Morgenstern<sup>10</sup> and provides facility-level and, within each facility, patient-level distributions for a continuing care PM and abstinence for hypothetical patients at each of 3 hypothetical facilities. These data are graphed in Figure 1. Table 2 illustrates that important information can be lost if the linkage between a PPM and outcomes for patients within facilities is not taken into account.

The top section of Table 2 depicts the absence of a relationship between facility percentages of patients receiving PM continuing care and facility abstinence rates. Performance with respect to continuing care varied from 33% to 72% of patients across these 3 hypothetical facilities, but the facility abstinence rate (25%) did not vary (see the mean flat slope in

**TABLE 2—Hypothetical Facilities With Different Levels of PM Performance and the Same Abstinence Rate and Positive Patient-Level Continuing Care PM–Abstinence Relationships**

	Facility A Abstinence			Facility B Abstinence			Facility C Abstinence		
	Yes	No	Total	Yes	No	Total	Yes	No	Total
<b>Marginal continuing care PM and abstinence distributions</b>									
Continuing care PM, %									
Yes	...	...	33	...	...	52	...	...	72
No	...	...	67	...	...	48	...	...	28
Total	25	75	100	25	75	100	25	75	100
<b>Marginal and within-facility continuing care PM and abstinence joint distributions (relationships)</b>									
Continuing care PM, No.									
Yes	26	40	66	36	68	104	44	100	144
No	24	110	134	14	82	96	6	50	56
Total	50	150	200	50	150	200	50	150	200

Note. PM = performance measure. The ellipses indicate patient-level data within programs that are either not available (if researchers only have facility-level data) or are not analyzed in facility-level-only analyses.

Figure 1 indicating no relationship between facility performance and facility abstinence rate). This analysis corresponds to analysis 1 (Table 1) and to Robinson’s<sup>13</sup> analysis of the state-level relationship between percentage of foreign-born residents and English illiteracy rates. From only these data, we do not know whether, or to what degree, a patient receiving continuing care at each of the 3 facilities is more

or less likely to be abstinent than a patient not receiving the PM-specified continuing care.

The bottom section of Table 2 presents one of many possible joint distributions—that is, the numbers of hypothetical patients in each of the 4 continuing care (yes or no) by abstinence (yes or no) cells for each facility—between receipt of PM care and abstinence. Shown is the relationship between continuing care receipt

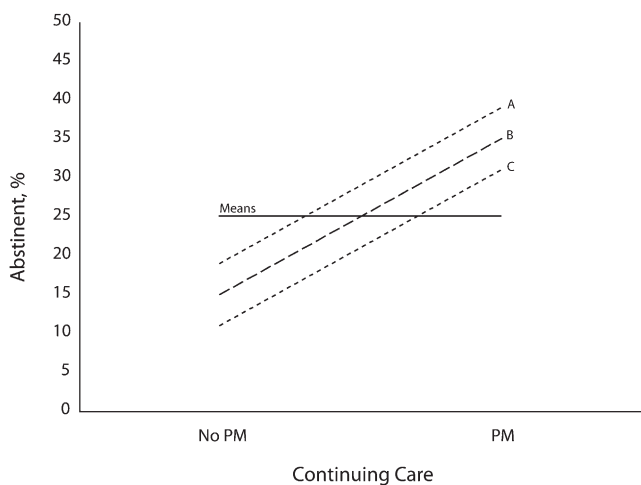
and abstinence for what are assumed to be 200 patients at each of the 3 facilities. Note that the percentage of patients receiving continuing care and the percentage of patients who were abstinent at each facility in the bottom section of Table 2 were the same as those in the top section.

Within each hypothetical facility in the bottom section of Table 2, there was a similar positive relationship between PM care and abstinence. Note that the line connecting the abstinence rate for patients receiving and not receiving continuing care within each facility in Figure 1 intersects the flat facility-level slope at the point representing the facility’s performance on the PM and the facility’s 25% abstinence rate. In facility A, which had the lowest level of PM performance (33%), the abstinence rate for patients receiving PM continuing care was the highest of the 3 facilities. Likewise, the abstinence rate in facility A for those not receiving continuing care also was the highest for the 3 facilities.

In facility C, which had the highest level of continuing care performance (72%), the abstinence rate for those receiving continuing care relative to those not receiving continuing care was higher by about the same amount as in facility A, but the abstinence rates for both the continuing care and no continuing care groups were the lowest of the 3 hypothetical facilities. Because facility-level data provide little information on PPM–outcome relationships for patients within facilities, the magnitudes and even the direction of between- and within-facility PPM–outcome relationships can vary.

**Different Confounding Variables at the Patient and Facility Levels**

PPM–outcome relationships may be different at the patient and facility levels because different factors confound the relationships at the 2 levels (a confounding variable is one that affects both the independent variable [e.g., PPM] and the dependent variable [e.g., outcome]).<sup>33</sup> For example, patients with co-occurring psychiatric disorders may be less likely to continue substance use disorder care and to be abstinent.<sup>34</sup> However, just as data on PPM and outcome percentages at the facility level provide little information on the relationship between the PPM and outcomes for patients within facilities, facility-level data on potential confounders



Note. CC = continuing care; PM = performance measure.

**FIGURE 1—Absence of a facility-level relationship of a continuing care PM to abstinence with similar positive PM–abstinence relationships for patients within 3 hypothetical facilities (A, B, and C).**

provide little information on the relationships among the PPM, outcomes, and confounders for patients within facilities. Thus, entering the proportion of substance use disorder patients at various facilities with comorbid psychiatric conditions as a covariate in a facility-level analysis would be unlikely to control for the confounding of psychiatric conditions with receipt of continuing care and outcomes at the patient level.

In addition, at the health care facility level, variables may affect the relationship between PPMs and outcomes that do not come into play, or play a different role, at the patient level. As Blalock observed,

The key to the problem may come with the realization that in shifting units we may be affecting the degree to which *other* unknown or unmeasured variables are influencing the picture.<sup>35(p99)</sup>

For example, as noted, an aggregated patient characteristic (e.g., proportion of patients with comorbid psychiatric conditions) may have different implications for facilities than the same variable would at the individual level.<sup>9</sup>

However, higher-level units, such as health care facilities, also have properties that are integral<sup>15</sup> to them (i.e., are not aggregated characteristics of the individuals who make up these units). For instance, different health care facilities have different leaders, local policies regarding care access, and structural properties, such as staff–patient ratios. They also operate in different environmental contexts that may vary in such elements as the availability of high-quality post-operative care at extended health care facilities or accessibility to self-help groups for patients with substance use disorders. These examples of additional higher-level variables associated with hospital PPM percentages are analogous to the higher state literacy rates among native-born residents that were positively related to higher percentages of foreign-born state residents in Robinson's<sup>13</sup> analyses.

These factors may confound the facility-level relationship between PPMs and outcomes but could have a quite different impact on patient-level PPM–outcome relationships. Because integral facility-level factors are less frequently assessed than are patient-level factors, the extent to which they confound facility-level or patient-level analyses is less likely to be known. Researchers interested in such integral or structural factors may assess and use them to

explain why facilities vary on PM-specified processes of care, but neither researchers nor quality managers would risk adjust for facility structural properties in a performance management system. Instead, the expectation is that facilities should create the structural environment necessary to allow specified processes of care to be applied to designated patients.

### Effect Modification

In addition to potentially reflecting information loss and different (although perhaps partially overlapping)<sup>8</sup> confounders, discrepancies between individual- and aggregate-level PPM–outcome relationships can occur when the grouping variable (e.g., health care facility) modifies PPM–outcome relationships at the patient level within facilities.<sup>10</sup> In other words, rather than unvarying PPM–outcome relationships for patients within facilities as depicted in Figure 1, these relationships could vary across facilities (Figure 2) as a result of different distributions at different facilities for one or more individual-level variables that moderate or influence the patient-level PPM–outcome relationship.<sup>8</sup>

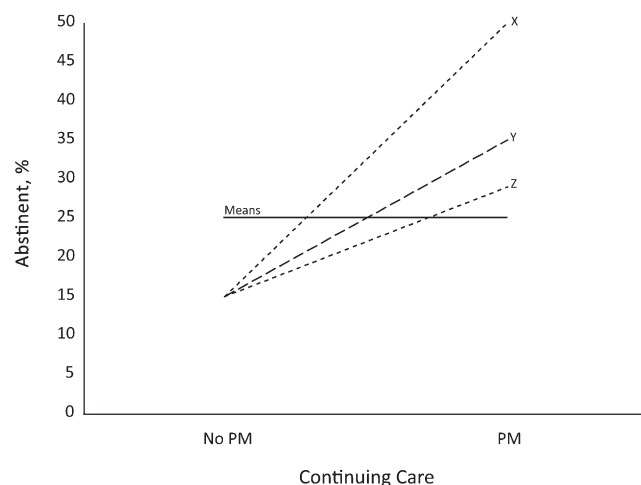
Although these moderator variables may not be related to PPM receipt or outcomes, they affect the impact of PM-specified care on the outcome. For example, a particular gene polymorphism might not be related to receipt of

a PM-specified medication for alcohol dependence or to the course of alcohol dependence among individuals not receiving the medication. However, if the gene polymorphism interacted with the medication to produce better outcomes,<sup>36</sup> the relationship of medication receipt with positive outcomes would be stronger among patients at facilities where this polymorphism was more prevalent. Regarding bias or inconsistency in group-level versus individual-level relationships as a result of effect modification, Greenland and Morgenstern noted that

[s]uch ecological bias can be severe even when the effect modification is relatively weak and there is no confounding by group. Consequently, it can be misleading to claim that a particular ecological analysis is unbiased simply because no confounding is present.<sup>10(p269)</sup>

### IMPLICATIONS FOR HEALTH CARE QUALITY MANAGEMENT

Two questions capture the inherently multilevel nature of PPMs and their relationship to clinical outcomes. Independent of the proportion of patients for whom PM-specified care is provided at their hospitals, should patients or providers prefer this type of care? Independent of the relationship of PM-specified care to within-hospital outcomes



Note. CC = continuing care; PM = performance measure.

**FIGURE 2—Absence of a facility-level relationship of a continuing care PM to abstinence with varying positive PM–abstinence relationships for patients within 3 hypothetical facilities (X, Y, and Z).**

among patients, should patients prefer receiving PM-specified care at a facility in which a higher proportion of patients receive this type of care or one in which a lower proportion of patients receive such care? These questions should be kept separate in validating, interpreting, and applying process of care PM data.

Clearly, different relationships can occur between PPMs and outcomes at the facility and patient levels as a result of, among other factors, loss of information about patients, differential confounding, and effect modification. Facility-level analyses and individual patient-level analyses (with or without control for facility-specific clustering on outcomes) produce findings on PPM–outcome relationships that are a mixture of between-facility and within-facility (patient-level) relationships and thus do not distinctly address these 2 questions.

Therefore, the optimal situation for health care systems, health services researchers, and entities that develop and monitor PPMs (e.g., National Quality Forum, Centers for Medicare and Medicaid Services, Joint Commission) would be to have data that would allow multilevel analyses of PPM–outcome relationships between and within health care facilities (as in our analysis 3). These analyses would be more useful if they included conceptually informed covariates and examination of facility-specific effect modification.

Ideally, PPM developers will identify at least some PPMs for which consistency in direction, if not magnitude, exists in patient- and facility-level associations with outcomes. PPMs that are related to positive outcomes at both the patient and facility levels are easier to defend. Because this situation will not always occur, however, we offer several practical suggestions for health care quality managers that flow from the methodological literature on cross-level inference and multilevel analysis.

Quality managers who have findings available only on facility-level PPM–outcome relationships (as in our analysis 1) should view those relationships with caution. Although they may reflect PPM–outcome relationships at the patient level, these relationships may also (as shown) differ in magnitude from or even be

in the opposite direction from patient-level relationships. Piantadosi concluded that, with aggregated data:

we not only lose all ability to extend inferences reliably to less aggregated data but we even lose the ability to estimate the direction and magnitude of bias. We cannot rely on the addition of more grouped data to eliminate the bias.<sup>37(p763)</sup>

In this context, Greenland and Morgenstern<sup>38</sup> noted that ecological analyses have produced findings indicating that radon exposure has a protective effect against lung cancer across counties in the United States<sup>39</sup> and that cigarette smoking has a protective effect against esophageal cancer across administrative areas in France.<sup>40</sup>

Overall, quality managers should remember Naylor's<sup>41</sup> caution concerning ecological analyses of treatment effects: *caveat emptor*. Without patient-level data analyzed in a multilevel framework, comprehensive analyses of PPM–outcome relationships are not possible.

If a multilevel analysis indicates that patient-level receipt of PM-specified care, but not facility performance with respect to provision of such care, is associated with positive outcomes, it seems reasonable that health care facility leaders should encourage the use of that practice through facility-level PPMs. This conclusion is especially appropriate when the process of care presumed to be captured by the PPM has been “tightly linked”<sup>42</sup> to better patient outcomes in multiple randomized controlled trials and effectiveness studies with relevant patient samples.

With respect to the tightness of PPM–outcome linkages, Kerr et al.<sup>42</sup> noted that the rate of administration of cholesterol screening, as a PPM for diabetes care, is a poorly linked PPM relative to the empirical literature that supports a relationship between patient outcomes and appropriate medical care in response to screening results. In terms of continuing care for patients with substance use disorders, it may be that provision of active, high-quality continuing care,<sup>28</sup> rather than simply provision of continuing care, is the relevant care process to track as a PPM. Furthermore, the target population to which the PPM is applied in a health care system should be tightly linked to the population in which the PM-specified process of care was shown to be beneficial.<sup>43</sup> In this regard, some

patients with substance use disorders may be able to achieve good outcomes without receiving the continuing care specified in the Veterans Affairs PPM.

Regardless of how tightly linked a PPM is to the supporting empirical evidence or a patient population, its linkage with patient-level outcomes should be validated and reexamined from time to time, including determining whether improvement on the PPM occurred over time and whether this improvement was related to improved clinical outcomes.<sup>44</sup> If facilities with higher rates of empirically supported PM-specified care have worse aggregate outcomes than facilities with lower rates of such care, quality managers should strive to understand the reasons for this situation; that is, what is happening at these facilities that may be canceling out or overriding the positive effects of more prevalent provision of PM-congruent care to patients?

By contrast, the proportion of facility patients meeting the PPM may be positively related to facility rates of desired outcomes, whereas there is a negative or no relationship between patient receipt of PM-specified care and desired outcomes.<sup>22</sup> Assume further that appropriate risk or case-mix adjustment has been applied so that, for example, the tendency for sicker patients with poorer prognoses to receive PM-specified care has been taken into account. Such findings would indicate that although PM-specified care is linked to poorer outcomes at the patient level, facilities with higher proportions of patients meeting the PPM are doing something that is linked to better patient outcomes.

For example, within each hospital, patients who receive a particular surgical procedure might be more likely to die. However, the facilities at which that surgical procedure was performed for proportionately more patients might also tend to provide better infection control and other safety measures. If so, the challenge would be to ensure a safer surgical procedure at all facilities while preserving or enhancing infection control and patient safety measures. Overall, multilevel findings can provide more useful and more nuanced information to quality managers on how patient outcomes might be improved by interventions targeting patients and their care, as well as

interventions targeting the broader health care facility or clinic.

## IMPLICATIONS FOR RESEARCH ON PERFORMANCE MEASURES

As noted by Firebaugh, “single-level analyses are subject to severe omitted-variable bias in the presence of multilevel effects.”<sup>45(p369)</sup> The obvious implication for researchers is that studies of the relationship of PPMs to patient outcomes should be guided by multilevel conceptual models and use multilevel analyses to examine them whenever possible. As noted earlier, although multilevel analyses have been applied in a variety of areas of health care research,<sup>20,21,46</sup> including linking care processes to outcomes,<sup>22</sup> they have not been used routinely to examine PPM–outcome relationships. Multilevel analyses are needed because, as we have described, facility-level relationships may be subject to the ecological fallacy and not be consistent with individual patient-level effects. Similarly, studies that focus solely on patients and ignore the facilities from which patients are drawn can be subject to the “atomistic fallacy”<sup>47</sup> and miss important contextual or ecological (facility-level) effects.

Researchers can probe causal models to better understand why PPM–outcome relationships vary across levels of analyses. In this context, it would be informative to conduct multilevel analyses to determine whether the quality indicators for the medical conditions examined by Bradley et al.<sup>3</sup> and Werner and Bradlow<sup>4</sup> have stronger relationships with lower risk-adjusted mortality at the patient level than was the case in their hospital-level analyses. (Fonarow et al.<sup>1</sup> analyzed the relationship of patient receipt of PM-specified and other care for chronic heart failure with mortality controlled for clustering of patients within hospitals but not for facility-level PM care provision.)

Potential confounders and moderators of PPM–outcome relationships at the patient and hospital levels could be explored as well. Of course, the high level of compliance on some PPMs (e.g., giving aspirin to patients who present with symptoms of acute myocardial infarction) and the lack of sensitivity of 30-day mortality as an outcome would

restrict the magnitude of relationships at the patient level, just as they likely do at the hospital level.

Some researchers and quality managers may argue that facility-level PPM–outcome relationships are of interest in their own right. This holistic approach makes sense if one is examining relationships among integral<sup>15</sup> characteristics of hospitals or facilities, that is, characteristics that cannot be disaggregated to the patient level (e.g., facility size). However, with their focus on patient outcomes (a dependent variable that can be disaggregated), multilevel analyses map much better to the phenomenon of interest than do holistic analyses.<sup>8,9</sup>

## CONCLUSIONS

Ideally, analyses at both the patient and facility levels will support the expected links between PPMs and outcomes. However, if patient-level analyses within facilities indicate reasonable relationships between receipt of PM-specified care and positive outcomes, we believe that such evidence should be considered support for encouraging this clinical practice via facility-level PPMs, even if the facility-level relationship is nonexistent or negative. This recommendation is consistent with prioritizing the provision of high-quality care to individual patients. It should be only an initial step, however, in an ongoing quality improvement process. Examining patient-level and facility-level findings in conjunction with each other can identify those facilities in which PPM performance is positively related to outcomes but outcomes for many patients are relatively poor.

Strengthening patient-level relationships (e.g., by incorporating substance use disorder continuing care interventions that are of higher quality and tailored to patient progress) while also using other means of improving system-level outcomes (e.g., increasing community access to self-help groups) would lead to better outcomes for more patients in a health care system. Overall, consideration of multilevel conceptual and methodological issues in performance measurement and management reinforces the wisdom of Donabedian’s<sup>48</sup> early call to focus on a (multilevel) system of structure, process, and outcome quality indicators. ■

## About the Authors

John W. Finney, Keith Humphreys, and Alex H.S. Harris are with the Center for Health Care Evaluation, Veterans Affairs Palo Alto Health Care System, Menlo Park, CA. At the time of the writing, Daniel R. Kivlahan was with the Substance Use Disorders Quality Enhancement Research Initiative, Veterans Affairs Puget Sound Health Care System, Seattle, WA.

Correspondence should be sent to John W. Finney, PhD, Center for Health Care Evaluation, VA Palo Alto Health Care System (152MPD), 795 Willow Rd, Menlo Park, CA 94025 (e-mail: john.finney@va.gov). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints/Eprints” link.

This article was accepted January 5, 2011.

## Contributors

J. W. Finney and A. H. S. Harris conceived the ideas underlying this article and wrote an initial draft. A. H. S. Harris conducted the data analysis for the empirical example. K. Humphreys and D. R. Kivlahan helped frame the conceptual issues and quality management implications and participated with the other authors in the writing and editing of subsequent drafts of the article.

## Acknowledgments

Preparation of this article was supported by the Health Services Research and Development Service (HSR&D), Office of Research and Development, US Department of Veterans Affairs (grants SUS 99-015 and IIR-07-092-1), and by the HSR&D Substance Use Disorder Quality Enhancement Research Initiative.

We thank Rudolf Moos and Avron Spiro for their helpful comments on an earlier version of this article.

**Note.** The views expressed are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

## Human Participant Protection

A project deemed exempt by the Stanford University and Veterans Affairs Palo Alto Health Care System institutional review board gathered the outcome data. The same institutional review board approved a project combining additional utilization data with the outcome data that included examining PPM–outcome relationships.

## References

1. Fonarow GC, Abraham WT, Albert NM, et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. *JAMA*. 2007;297(1):61–70.
2. McLellan AT, Chalk M, Bartlett J. Outcomes, performance, and quality—what’s the difference? *J Subst Abuse Treat*. 2007;32(4):331–340.
3. Bradley EH, Herrin J, Elbel B, et al. Hospital quality for acute myocardial infarction: correlations among process measures and relationship with short-term mortality. *JAMA*. 2006;296(1):72–78.
4. Werner RM, Bradlow ET. Relationship between Medicare’s Hospital Compare performance measures and mortality rates. *JAMA*. 2006;296(22):2694–2702.
5. Lehrman WG, Elliott MN, Goldstein E, Beckett MK, Klein DJ, Giordano LA. Characteristics of hospitals demonstrating superior performance in patient

- experience and clinical process measures of care. *Med Care Res Rev.* 2010;67(1):38–55.
6. Sequist TD, Schneider EC, Anastario M, et al. Quality monitoring of physicians: linking patients' experience of care to clinical quality and outcomes. *J Gen Intern Med.* 2008;23(11):1784–1790.
  7. Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *Am Sociol Rev.* 1978;43(4):557–572.
  8. Greenland S. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int J Epidemiol.* 2001;30(6):1343–1350.
  9. Greenland S. A review of multilevel theory for ecologic analyses. *Stat Med.* 2002;21(3):389–395.
  10. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol.* 1989;18(1):269–274.
  11. Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health.* 1982;72(12):1336–1344.
  12. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol.* 1988;127(5):893–904.
  13. Robinson WS. Ecological correlations and the behavior of individuals. *Am Sociol Rev.* 1950;15(3):351–357.
  14. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health.* 1994;84(5):819–824.
  15. Susser M. The logic in ecological: I. The logic of analysis. *Am J Public Health.* 1994;84(5):825–829.
  16. Goldstein H, Browne W, Rasbash J. Tutorial in biostatistics: multilevel modeling of medical data. *Stat Med.* 2002;21(21):3291–3315.
  17. Greenland S. Principles of multilevel modelling. *Int J Epidemiol.* 2000;29(1):158–167.
  18. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage Publications; 2002.
  19. Damman OC, Stubbe JH, Hendriks M, et al. Using multilevel modeling to assess case-mix adjusters in consumer experience surveys in health care. *Med Care.* 2009;47(4):496–503.
  20. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med.* 1998;46(1):97–117.
  21. Lumme S, Leyland AH, Keskimaki I. Multilevel modeling of regional variation in equity in health care. *Med Care.* 2008;46(9):976–983.
  22. Berlin JA, Kimmel SE, Ten Have TR, Sammel MD. An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty. *Biometrics.* 1999;55(2):470–476.
  23. Harris AHS, Humphreys K, Bowe T, Kivlahan DR, Finney JW. Measuring the quality of substance use disorder treatment: evaluating the validity of the Department of Veterans Affairs continuity of care performance measure. *J Subst Abuse Treat.* 2009;36(3):294–305.
  24. Hannan MT. *Problems of Aggregation and Disaggregation in Sociological Research.* Chapel Hill, NC: Institute for Research in the Social Sciences, University of North Carolina; 1970.
  25. Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: the perils of individualistic and ecological fallacy. *Int J Epidemiol.* 2009;38(2):342–360.
  26. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med.* 2001;135(2):112–123.
  27. Harris AHS, Kivlahan DR, Bowe T, Finney JW, Humphreys K. Developing and validating process measures of health care quality: an application to alcohol use disorder treatment. *Med Care.* 2009;47(12):1244–1250.
  28. McKay JR. Continuing care research: what we have learned and where we are going. *J Subst Abuse Treat.* 2009;36(2):131–145.
  29. Desai M, Begg MD. A comparison of regression approaches for analyzing clustered data. *Am J Public Health.* 2008;98(8):1425–1429.
  30. Hammermeister KE, Shroyer AL, Seithi GK, Grove FL. Why it is important to demonstrate linkages between outcomes of care and processes and structures of care. *Med Care.* 1995;33(suppl 10):OS5–OS16.
  31. Rosenheck R, Fontana A, Stolar M. Assessing quality of care: administrative indicators and clinical outcomes in posttraumatic stress disorder. *Med Care.* 1999;37(2):180–188.
  32. Horn SD. Performance measures and clinical outcomes. *JAMA.* 2006;296(2):2731–2732.
  33. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics.* 1998;54(2):638–645.
  34. Ouimette PC, Gima K, Moos RH, Finney JW. A comparative evaluation of substance abuse treatment: IV. The effect of comorbid psychiatric diagnoses on amount of treatment, continuing care, and 1-year outcomes. *Alcohol Clin Exp Res.* 1999;23(3):552–557.
  35. Blalock HM. *Causal Inferences in Nonexperimental Research.* Chapel Hill, NC: University of North Carolina Press; 1964.
  36. Oslin DW, Berrettini W, Kranzler HR, et al. A functional polymorphism of the mu-opioid receptor gene is associated with naltrexone response in alcohol-dependent patients. *Neuropsychopharmacology.* 2003;28(8):1546–1552.
  37. Piantadosi S. Invited commentary: ecologic biases. *Am J Epidemiol.* 1994;139(8):761–764.
  38. Greenland S, Morgenstern H. Neither within-region nor cross-regional independence of exposure and covariates prevents ecological bias [letter to the editor]. *Int J Epidemiol.* 1991;20(3):816–817.
  39. Cohen BL. Tests of the linear-no threshold theory of radiation carcinogenesis. *Environ Res.* 1990;53(2):193–220.
  40. Richardson S, Stucker I, Hemon D. Comparison of relative risks obtained in ecological and individual studies. *Int J Epidemiol.* 1987;16(1):111–120.
  41. Naylor CD. Ecological analysis of intended treatment effects: caveat emptor. *J Clin Epidemiol.* 1999;52(1):1–5.
  42. Kerr EA, Krein SL, Vijan S, Hofer TP, Hayward RA. Avoiding pitfalls in chronic disease quality measurement: a case for the next generation of technical quality measures. *Am J Manag Care.* 2001;7(11):1033–1043.
  43. Wachter RM, Flanders SA, Fee C, Pronovost PJ. Public reporting of antibiotic timing in patients with pneumonia: lessons from a flawed performance measure. *Ann Intern Med.* 2008;149(1):29–32.
  44. Selby JV, Uratsu CS, Fireman B, et al. Treatment intensification and risk factor control: toward more clinically relevant quality measures. *Med Care.* 2009;47(4):395–402.
  45. Firebaugh G. Is the world flat? W.S. Robinson and the ecologic fallacy. *Int J Epidemiol.* 2009;38(2):368–370.
  46. Arling G, Lewis T, Kane RL, Mueller C, Flood S. Improving quality assessment through multilevel modeling: the case of Nursing Home Compare. *Health Serv Res.* 2007;42(3):1177–1199.
  47. Alker H Jr. A typology of ecological fallacies. In: Dogan M, Rokkan S, eds. *Quantitative Ecological Analysis.* Cambridge, MA: MIT Press; 1969:69–86.
  48. Donabedian A. *Explorations in Quality Assessment and Monitoring: The Definition of Quality and Approaches to Its Assessment.* Vol. 1. Ann Arbor, MI: Health Administration Press; 1980.