

# THE CHALLENGE OF ASSIGNING SURPLUSES OR DEFICITS TO INDIVIDUAL PHYSICIANS "AT RISK" FOR MANAGED CARE PHARMACY EXPENSES

M-1549

November 2000

Mark Cowen  
and  
Robert L. Strawderman

**Keywords:** risk adjustment, managed care, pharmaceutical utilization, mixed regression models, physician resource utilization.

**Abstract:**

**Context:** A strategy for controlling pharmacy costs in managed care is to provide incentives for physicians to prescribe efficiently. Little is known about how the choice of method for assigning budgetary surpluses or deficits impacts individual physicians.

**Objective:** To develop alternative statistical approaches for risk adjustment of pharmacy utilization at the physician level, and compare projected surpluses or deficits with those generated by traditional methods.

**Design:** Use of two basic statistical modeling techniques: multivariable, fixed-effects, linear regression as the traditional approach, and mixed models as the alternative. Explanatory variables consisted of age, gender, and patient diagnoses.

**Setting:** Two managed care populations in 1999, one from the Midwestern US in which all physicians were in a risk arrangement, the other from three Western States where two-thirds of physicians took risk.

**Main Outcome Measures:** The range of projected surpluses or deficits for individual physicians, and the intra-class correlation coefficient (the proportion of variability in expenses attributable to differences among physicians).

**Results:** The intra-class correlation coefficients were very small, 0.04 or less in both populations. However, the projected financial performance of individual physicians varied tremendously under the traditional approach, from a maximum deficit of \$104,086 to a maximum surplus of \$116,403. In comparison, an alternative approach projected a maximum deficit of \$26,341 and a maximum surplus of \$14,526 using the same data.

**Conclusions:** In the two populations studied, traditional regression models led to substantial differences in surpluses and deficits among physicians, despite almost negligible statistical differences in physician performance. The alternative approaches generated considerably smaller surpluses or deficits, underscoring the practical implications and importance of the choice of method when individual physicians are at risk.

29NOV2000

SUBMITTED TO  
MEDICAL CARE

BOK -  
1549

A153

**The Challenge of Assigning Surpluses or Deficits to Individual Physicians  
"At Risk" for Managed Care Pharmacy Expenses**

Mark E. Cowen, M.D., S.M.<sup>1</sup>

Robert L. Strawderman, Sc.D.<sup>2</sup>

<sup>1</sup> Departments of Medicine, St. Joseph Mercy Hospital and the University of Michigan Medical School, Ann Arbor, MI.

<sup>2</sup> Departments of Biometrics and Statistical Science, Cornell University, Ithaca, NY.

Corresponding Author: Mark E. Cowen, M.D., c/o Allegiance L.L.C., Suite 9, 2000 Hogback Rd., Ann Arbor, MI 48105; Fax (734) 677-4847; Email: [cowenm@allegiancellc.com](mailto:cowenm@allegiancellc.com)

Word Count, abstract: 263

Word Count, text: 4,243

Key words: risk adjustment, managed care, pharmaceutical utilization, mixed regression models, physician resource utilization

## **The Challenge of Assigning Surpluses or Deficits to Individual Physicians "At Risk" for Managed Care Pharmacy Expenses**

### **ABSTRACT**

**Context.** A strategy for controlling pharmacy costs in managed care is to provide incentives for physicians to prescribe efficiently. Little is known about how the choice of method for assigning budgetary surpluses or deficits impacts individual physicians.

**Objective.** To develop alternative statistical approaches for risk adjustment of pharmacy utilization at the physician level, and compare projected surpluses or deficits with those generated by traditional methods.

**Design.** Use of two basic statistical modeling techniques: multivariable, fixed-effects, linear regression as the traditional approach, and mixed models as the alternative. Explanatory variables consisted of age, gender, and patient diagnoses.

**Setting.** Two managed care populations in 1999, one from the Midwestern U.S. in which all physicians were in a risk arrangement, the other from three Western States where two-thirds of physicians took risk.

**Main Outcome Measures.** The range of projected surpluses or deficits for individual physicians, and the intra-class correlation coefficient (the proportion of variability in expenses attributable to differences among physicians).

**Results.** The intra-class correlation coefficients were very small, 0.04 or less in both populations. However, the projected financial performance of individual physicians varied tremendously under the traditional approach, from a maximum deficit of \$104,086 to a maximum surplus of \$116,403. In comparison, an alternative approach projected a maximum deficit of \$26,341 and a maximum surplus of \$14,526 using the same data.

**Conclusions.** In the two populations studied, traditional regression models led to substantial differences in surpluses and deficits among physicians, despite almost negligible statistical differences in physician performance. The alternative approaches generated considerably smaller surpluses or deficits, underscoring the practical implications and importance of the choice of method when individual physicians are at risk.

## **The Challenge of Assigning Surpluses or Deficits to Individual Physicians "At Risk" for Managed Care Pharmacy Expenses**

Introduction. A common strategy <sup>1</sup> to control rising pharmacy costs in managed care is to hold physicians financially responsible or "at risk" for these expenditures.<sup>2</sup> If patient expenses fall below the target, physicians receive a bonus. If expenses exceed the target, physicians suffer the loss. A number of concerns arise if surpluses or deficits are allocated to individual physicians. There is an ethical concern if the dollars involved represent a significant portion of a physician's income.<sup>3</sup> There are also technical concerns regarding the validity of the methods used to calculate these incentives.

Usually in managed care settings, physician "efficiency" is measured only indirectly, that is, as the ratio of the actual patient costs to their expected costs, <sup>4</sup> or alternatively, as the difference between the patients' expected and actual expenditures. <sup>5</sup> If either formulation is used to determine the financial incentive, then the magnitude of a physician's bonus or penalty will depend on how closely his or her patients' costs approximate their expected amounts. Because comprehensive criteria are not readily available for identifying what prescriptions individual patients should receive, a physician's expected costs are often based on those of the "average" physician, an example of a practiced-based norm. <sup>6</sup> Use of this type of norm presents two challenges: properly accounting for the illness burden of the patients, <sup>7</sup> and appropriate methods for the statistical comparison of physicians.

Illness burden is addressed through risk adjustment. Although sophisticated methods have been used in studies of quality of care provided by hospitals <sup>8</sup> and physicians, <sup>9</sup> risk adjustment methods used in managed care settings are generally less developed. Sometimes no adjustment is performed, <sup>10</sup> or at best, multivariable linear regression models are used controlling for age, gender, and at times diagnostic information. <sup>7, 11, 12</sup> A common measure of how well the risk has been adjusted at the patient level is the proportion of the variability in expenses accounted for by

the explanatory variables, a quantity known as the  $R^2$ .<sup>13</sup> The higher reported values for managed care pharmacy utilization have been in the 0.20 – 0.30 range.<sup>12, 14, 15</sup> The extent to which these low values reflect variability in physician prescribing habits and or inadequate risk adjustment is not known.

Comparing physicians also presents statistical challenges, the financial consequences of which are not well understood either. Sometimes, researchers measure "efficiency" by including the "provider" as an explanatory variable in linear regression models.<sup>16, 17</sup> However, this may overstate the provider effect, particularly for physicians caring for small numbers of patients,<sup>8, 9, 18</sup> and consequently exaggerate the surplus or deficit for an individual physician. Moreover, this technique does not distinguish between pure sampling variability and true inter-physician variability, the latter being an important aspect of the profiling process.<sup>8</sup> There are other concerns with using traditional regression methods to compare physicians. For example, such methods assume that each patient's costs are unrelated to those of every other patient, that is, patient costs are "statistically independent." However, this assumption presents a contradiction because patients served by a physician who systematically prescribes efficiently or inefficiently share, by definition, a common practice style. Failure to properly account for these clusters of patients is tantamount to assuming all physicians prescribe homogeneously. If an organization really believed all physicians prescribed the same, then there would be no need to compare physicians in the first place. This flaw also has some important implications for statistical inference.

The major goal of this study is to derive physician-level surpluses or deficits by approaches that are more defensible from a methodological point of view than those based on traditional linear regression. The practical consequences of the choice of statistical method will be illustrated by comparing the physician-level surpluses or deficits generated by these different approaches. The key difference between the alternative and traditional approaches studied is the use of a statistical technique known as a "mixed" model. With this, heterogeneity in physician prescribing,

and hence clustering of patient costs by physician, is captured through the introduction of a random physician effect. The simplest approach assumes heterogeneity is reflected entirely by physician-level differences in the baseline cost for the "average" patient. This allows prescriber efficiency to be assessed more explicitly by partitioning the overall variability in prescribing costs into two components: random error and inter-physician variability. Our emphasis will be on the choice of statistical models rather than on the identification of the best risk-adjusting variables available, which would necessitate a separate study. Thus our analyses utilized a set of clinically-meaningful diagnostic categories which we developed for the public domain.

Methods. We studied two distinct patient populations from different HMOs, referenced by geographical location within the United States. To simplify interpretation of the findings, the analyses were confined to patients 18 to 64 years old with total pharmacy expenditures under \$20,000 per year, enrolled for at least 11 months of the year, and attended by primary care physicians with at least 10 HMO patients. This reduced variability in expenditures due to catastrophic conditions, differences between adults and children, enrollment times, or practitioner type. We also assumed that the organizations as a whole performed at their target levels. This implies "physician efficiency" is being assessed relative to the average physician within his or her organization, an appropriate point of view from the perspective of a managed care organization. Primary care physicians were defined as family physicians, internists, and pediatricians. Pharmacy expenditures were defined as the "ingredient costs", or contracted average wholesale price of the pharmaceuticals minus a network discount. As is common in profiling, <sup>4,19</sup> costs were assigned to the designated primary care physician even if the prescriptions were written by others.

The physicians of the "Midwestern" population belonged to a physician-hospital organization at risk for medical and pharmacy expenditures as previously reported. <sup>20</sup> The "Western" population was drawn from a three-State region HMO in which most physicians were independent practitioners. Approximately two-thirds of these physicians participated in a risk arrangement.

Initial analyses, and in particular variable selection, were performed using 1996 data from the Midwest. Minor modifications to the list of explanatory variables were made subsequently, but the same models were used with 1998 and 1999 Midwestern data, and with 1999 Western data.

The primary outcome of interest is the derived surplus or deficit at the physician level expressed in two forms, as absolute dollars and as a proportion of the actual expenditures of the panel of patients. These are reported in aggregate, that is, as the range of physician surpluses or deficits determined by the various methods. Of secondary interest is the extent of physician influence on pharmacy costs, measured here through the "intra-class correlation coefficient" (ICC).<sup>9, 21</sup> The ICC measures the proportion of variability in a patient's expenditures attributable to the physician effect. Values of ICC close to 1.0 indicate physician-level factors (eg. prescribing habits) explain most of the observed variability in patient pharmacy costs; values close to 0.0 indicate the opposite.

All statistical analyses were performed using SAS Version 6.12.<sup>22</sup> Two basic statistical modeling techniques were used. The first or traditional approach used fixed-effects, multivariable linear regression models, generated through SAS PROC REG. The second employed similar models but also included a random intercept term, a simple example of a mixed effects model, as generated through PROC MIXED. Given an arbitrary patient with covariates  $X$ , the traditional approach assumes the pharmacy costs for this type of patient are, on average, the same for every physician:

$$\text{predicted cost for physician } i \text{ for a patient with covariates } X = a + f(X),$$

where  $a$  is an intercept term that represents the cost of the average patient and  $f(.)$  is some known function of the patient-level covariates  $X$ . Importantly, neither  $a$  nor  $f(.)$  depend on the specific patient or physician under consideration. With this approach, a physician's surplus or deficit is calculated as the sum of the actual expenditures of the patients minus the sum of their

predicted costs. Thus, a positive difference yields a deficit owed by the physician and a negative difference a surplus.

In contrast, the mixed effects model assumes

$$\text{predicted cost for physician } i \text{ for a patient with covariates } X = a + b(i) + f(X).$$

Here,  $a$  and  $f(.)$  are as above, and  $b(i)$  represents a physician-specific contribution to the per-patient cost independent from the specific patient under consideration. This model assumes the unobserved deviations  $b(i)$  are normally distributed with mean zero and constant variance. Theoretically, the parameter  $a$  is the same in both the fixed and mixed effects regression models, and represents the cost of the average patient in the entire managed care organization. Because we have assumed the organization performs at its target level, the expected contribution of any physician to this average cost is therefore equal to zero. Physician heterogeneity is incorporated by allowing the model intercept to be physician-specific, which in practical terms means we allow a given physician's actual contribution to patient costs to deviate from the expected contribution of zero. The "empirical best linear unbiased predictor" (EBLUP) of the deviations  $b(i)$  constitute model-derived, direct estimates of each physician's per patient contribution to cost (hereafter, the "physician effect"). An estimated surplus or deficit for a given physician may then be derived by multiplying his or her EBLUP by the number of his or her patients. From an organizational perspective, a negative value constitutes a surplus owed to an "efficient" physician, and positive value a deficit owed by an "inefficient" physician. The use of the EBLUP effectively "shrinks" the estimated surplus or deficit towards zero, relative to what it would have been if computed under the traditional approach. The amount of shrinkage for a given physician decreases as one or both of the ICC and patient panel size increases.<sup>9</sup> We refer to this alternative approach as the "intercept method", and compare this to the traditional fixed-effects approach referenced as "linear regression".

Because the notion of an "efficient" physician is inexact, and point estimates of the surplus (or deficit) are subject to sampling variability, we also present an example of a decision rule that could be used to modify the EBLUP further in the calculation of a specific incentive. We refer to this as "the confidence interval rule". We first compute an estimate of the standard deviation (SD) of each physician's estimated surplus or deficit (SURDEF), then subsequently a crude confidence interval of the form  $SURDEF \pm SD$ . The SD calculation reflects two sources of variation: that due to inter-physician heterogeneity and that due to sampling variability. Mathematical details underlying the SD computation are available as supplemental material. Financial risk is subsequently determined as follows. If the confidence interval contains zero, then the physician's surplus or deficit is set to zero. If  $SURDEF - SD$  exceeds zero, then this amount is taken as the deficit owed by the physician. Similarly, if  $SURDEF + SD$  is less than zero, then this is taken to be the surplus owed to the physician. This decision rule sets a limit on the amount of inter-physician heterogeneity that will be tolerated, accounting for the uncertainty in the estimated surplus/deficit. For a given patient panel, an efficient physician is defined as one who operates within  $\pm 1$  SD of the "ideal" physician whose discretionary contribution to patient costs are exactly zero. Generally, this rule leads to surplus and deficits smaller than those produced by the "intercept" method, the extent being governed by the magnitude of the SD.

We also studied the impact of using both the intercept method and its confidence interval application for separate pharmacy categories. We defined 28 mutually-exclusive pharmaceutical groups based on their American Hospital Formulary Service classification number.<sup>23</sup> Categories were defined to be clinically relevant, for example, "antibacterials", "antihistamines", "GI", "cardiovascular", and "antidepressants". Details are available as supplemental material. The total cost for each patient is correspondingly broken down into 28 mutually exclusive parts, with each component of total cost modeled separately as described above. This approach yields 28 EBLUP-derived surplus (or deficit) estimates per physician, which are then summed to produce an overall surplus or deficit estimate. The "28 models" approach makes significantly less restrictive assumptions regarding the relationship between patient covariates and total pharmacy

costs, and allows a physician's influence on costs to differ across the categories. Confidence intervals are also constructed for each category, as previously described.

In all, 5 basic approaches were compared: "traditional linear regression", "intercept method single model", "intercept method 28 models", "confidence interval rule single model", and "confidence interval rule 28 models". For all five approaches, explanatory variables consisted of patient age and gender at a minimum, with patient diagnoses utilized as indicated. Because diagnoses were not assigned to particular prescriptions, they were obtained from the medical care claims of the patients and merged with their pharmacy data. The Clinical Classifications for Health Policy version 2 (now known as the Clinical Classifications Software) <sup>24, 25</sup> was used to group ICD9 diagnoses into 260 categories, <sup>12</sup> which were in turn collapsed into 37 groups of clinically similar conditions, examples of which are "infections", "diabetes", "affective disorders", and "hypertension ". Further details are available as supplemental material.

The baseline scenarios capped pharmacy costs per category at \$2500, that is, costs higher than \$2500 were assigned the value of \$2500, whether costs were examined with single or with 28 models. This level accounted for approximately 95% of pharmacy expenditures, and modulated the influence of the higher cost patients. Sensitivity analyses were performed by varying the level of the cap for pharmacy costs within each category, and also the minimum physician panel size.

Results. Characteristics of the two study populations are shown in Table One. Despite similar age and gender compositions, the groups differed more than two-fold in per-patient pharmacy expenditures. Costs were highly skewed in both groups. When costs were capped at a maximum of \$750 per pharmacy category, the most expensive patient incurred \$ 6,836 in the Midwestern population and \$5,703 in the Western population (Table 2). When the ceiling of costs was raised to \$20,000 per pharmacy category, the most costly patient incurred \$19,782 in the Midwestern population, and \$18,914 in the Western.

With traditional linear regressions for the Midwestern population, the adjusted  $R^2$  ranged from 0.10 (costs per pharmacy category capped at \$750) to 0.06 (costs capped at \$20,000), controlling for patient age and gender (results not displayed in tables). When patient diagnoses were also considered, the adjusted  $R^2$  improved to 0.39 and 0.30 respectively. The regression models performed less well in the Western population, where the corresponding measures were 0.06 and 0.03 controlling for age and gender, and 0.26 and 0.20 using diagnoses.

Only a small proportion of the overall variability in both populations could be attributed to the physician (random) effect, expressed as the intra-class correlation coefficient (Table 2), as derived from the mixed models. In the Midwestern population, the coefficients ranged from a high of 0.018 to a low of 0.002, depending on the modeling technique, level of capped expenditures and use of diagnoses. There was relatively more variability among physicians in the Western population, where the intra-class correlation coefficient ranged from 0.042 to 0.003. Estimates of the 209 Midwest physicians' EBLUPs, (their specific contributions to patient costs), are shown in Figure 1 and expressed as dollars per patient (single mixed model, costs per category capped at \$2500). The one standard deviation confidence intervals surrounding these EBLUPs are very broad, with an average width of \$112.

The projected financial performance of the physicians varied greatly depending on the modeling method used (Figure 2). Only one scenario from the Midwest population in 1999 is shown (per category costs capped at \$2500), but the pattern was similar in both populations using varying levels of capped costs. The outer dark bars encompass the entire range of physicians. The inner lighter bars represent the range between the fifth and ninety-fifth percentile of physicians, that is, the middle 90%. The largest individual surpluses or deficits were projected from the standard regression models, particularly those controlling only for patient age and gender. With this scenario, individual physician performance ranged from a deficit of \$104,086 to a surplus of \$116,403 in the Midwest population, and a deficit of \$89,964 to a surplus of \$83,199 in the Western. This range was compressed somewhat with the "intercept method", especially when

the 28 models employing diagnoses were considered. As expected, the smallest surpluses or deficits were projected using the "confidence interval rule" for each of the 28 pharmacy categories. Here, the maximum physician deficit in the Midwestern population was \$26,341 and the maximum surplus \$14,526. The corresponding figures in the Western population were \$24,773 and \$5,494. The difference in estimates for the Midwest population compared to the comparable "intercept" method is appreciable, and largely reflects the substantial variability observed in Figure 1.

The differences among the various methods appear more striking when the surpluses or deficits are expressed as proportions of the actual expenditures of the patient panels (Figure 3). The largest proportional surpluses or deficits were generated through linear regression models controlling only for patient age and gender. One physician was projected to owe 60% of his patient's expenditures, while another would receive a bonus nearly 6 times the pharmacy costs of his patients. In the West, the physician with the worse deficit would have to pay back 80% of his patient's expenditures, while a peer would receive a surplus 97 times what his or her patients spent. The smallest range in performance was that determined by the confidence interval rule for each of the 28 pharmacy categories. In the Midwest, the physician with the proportionally largest deficit would owe 10% of his patient's expenditures with this method, while the physician with the largest surplus would receive 19% of the pharmacy costs of his panel. The corresponding figures in the West were 16% and 15%, respectively.

At the organizational level, it is important for budgetary reconciliation that the physician-level surpluses or deficits sum to zero. Such agreement is assured if a traditional linear regression model is used. However, this did not occur under the various mixed models we fit, primarily because of the way in which the physician-level surpluses (or deficits) are derived. With the use of the "intercept" method, there was a projected net aggregated physician surplus in the Midwest of between 0.2 % and 1.8 % of the total pharmacy expenditures, depending on the level of capped costs, models and covariates considered. In the West there was not a consistent

pattern, with net surpluses or net deficits projected depending on the scenario used, however these were within 3% of the total pharmacy costs. Using the confidence interval allocation rule, there was a projected net summed physician deficit ranging between 0.4% and 1.0% of total pharmacy expenditures in the Midwest, and between 0.4% and 2.2% in the West, indicating money would be owed by the aggregated physicians to the managed care organizations.

A sensitivity analysis was performed (Figure 4) to determine whether the wide range in physician performance was due to high cost patients and small physician panel sizes. Consequently, a restricted sub-population of patients was studied whose costs were capped at a maximum of \$750 per pharmacy category, and who were assigned to physicians who had at least 200 patients per panel. Under these conditions, the range in physician performance based on linear regression tightened considerably over that of the baseline scenario shown in Figure 2. However, the range in physician performance determined by the confidence interval rule, 28 mixed models, did not change substantially from the baseline scenario. This suggested the latter approach was less sensitive to the inclusion of high cost patients and small physician panel sizes.

Comment. We developed and studied some alternatives to traditional linear regression for measuring and rewarding physician prescribing efficiency. The magnitude of the projected physician-level surpluses or deficits varied greatly depending on the statistical modeling method used. We found variability among physician prescribing was small relative to other sources of variability in the data, a finding consistent with that reported recently regarding the management of diabetes.<sup>9</sup> Although the findings were similar in two populations differing more than two-fold in pharmacy expenditures, they might not be generalizable to others. Our methodological considerations would continue to remain applicable, however.

We believe that the major criterion for selecting a preferred approach is methodological validity. Explicit control of an individual physician's profit or loss is secondary. As indicated in the Introduction, using linear regression to derive expected costs for individual physicians rests on

questionable, possibly even contradictory, foundations. The use of mixed models has a more defensible methodological basis because they allow for the systematic aspect of a physician's practice, a specific objective of profiling. A useful way to think about the key difference between the traditional and mixed regression methods is as follows. Under the traditional approach, the "residual cost", or actual – predicted cost, is assumed to be entirely due to the physician effect, with no allowance for random (e.g., yearly) variability in prescribing costs for patients of a given type. In contrast, the mixed model assumes this residual further consists of two components: a systematic contribution due to the physician, and random variability. The EBLUP estimates the former, and its use at least acknowledges that there is an important difference between these two types of contributions for the purposes of determining financial responsibility.

Once the physician's contribution to costs has been quantified, some mechanism is needed for generating a fair incentive. The confidence interval allocation rule appears to be a reasonable approach. It defines an efficient physician as one whose likely contribution to costs is within 1 standard deviation of the expected contribution of zero. Statistically, its construction acknowledges that point estimates of the physician effect are still subject to sampling variability (see Figure 1), thus reducing the possibility of disproportionate rewards or penalties by the vagaries of chance. Methodologically, it allows for some inter-physician heterogeneity in the definition of "efficiency", and for some uncertainty regarding physician-level risk factors that are neither related to quality of care nor under their direct control. This allocation rule is certainly not the only way in which one could incorporate the various sources of variability into the incentive process. Nevertheless, it balances methodological, practical, and statistical concerns in a reasonably straightforward manner.

There are some limitations with this study that should be acknowledged. Our diagnostic categories might not be the optimal set of explanatory variables. However, any improvement over these would be expected to decrease the range of surpluses or deficits even further. There are also some limitations connected to our use of mixed regression models. For example, we

assume the variation in physician practice patterns is completely captured by the inclusion of a random intercept term, and that the variation in this term reflects only "controllable" variation in practice styles. Both of these assumptions represent substantial oversimplifications. Our use of 28 different models, hence 28 random intercept terms for the various cost categories, may partially offset these drawbacks. These assumptions could be relaxed with the incorporation of additional data on factors thought to influence physician practice style but having no explicit connection to pharmacy costs. Such data are presently unavailable. Another limitation is that physician performance continues to be defined in terms of how the average physician prescribes rather than by criteria on how they should prescribe. However, comprehensive criteria-based guidelines linked directly to organizational expenditures are presently unavailable. In the meantime, studying pharmaceutical categories separately, as done with the 28 models, permits some flexibility in the assessment process. Because the categories were clinically defined, some type of weighting, or a mixed reimbursement scheme such as that proposed by Schokkaert<sup>26</sup> could be developed in the future. For example, physicians could assume a greater share of risk for antihistamine utilization than for cardiovascular drugs. A final, more practical limitation is that of budgetary reconciliation. As indicated, the sum of the physicians' surpluses and deficits obtained under the various mixed model approaches generally do not net out exactly to zero. Although the imbalance represented a very small percentage of the pharmacy expenditures in all cases, an organization would still need a mechanism for the reconciliation process, particularly if it exceeded its budget. We view this concern as being more administrative than methodological, and consequently will not attempt to make further recommendations here.

Physicians within each of the two organizations practiced similarly on the most part to other physicians within their organization. If physician inefficiency were the sole factor responsible for the difference in costs between the Midwestern and Western population, then most physicians in the Midwestern population were prescribing inefficiently. Thus, if organizations wish to control their pharmacy costs, we suggest the focus be directed away from penalizing its outlier physicians,<sup>27</sup> towards identifying and addressing the factors responsible for the shared high-cost

culture.<sup>9, 28, 29</sup> Such explanations might include unmeasured case mix differences, administrative structures, benefit designs, co-pay influence,<sup>2</sup> physician supply,<sup>29</sup> activity of pharmaceutical detailing and other cultural factors.<sup>30</sup> Unfortunately, sufficient data were not available to us to examine these influences at this time.

A fair allocation of surpluses or deficits to individual physicians is not an easy task, nor can it be done without error. The methodological concerns regarding the traditional approaches appear to have practical consequences. Given the continued rise in pharmaceutical costs<sup>31</sup> and the real potential for large profits or losses,<sup>32</sup> those currently under risk arrangements may wish to scrutinize how their own surplus or deficit is being calculated.

## **Acknowledgements**

The authors are grateful for the technical assistance of Doug Clark, Dave Dusseau, and Lori Kostoff PharmD, and the helpful comments on an earlier draft of this manuscript by Michael Chernew PhD, Timothy Hofer MD, J. William Thomas PhD, and Paul N. Valenstein MD.

## References

1. *Novartis Pharmacy Benefit Report: 1999 facts and figures*. East Hanover, NJ; 1999:11.
2. Hillman AL, Pauly MV, Escarce JJ, Ripley K, Gaynor M, Clouse J, Ross R. Financial incentives and drug spending in managed care. *Health Aff* 1999; 18:189–200.
3. Pearson SD, Sabin JE, Emanuel EJ. Ethical guidelines for physician compensation based on capitation. *N Engl J Med* 1998;339:689–93.
4. Tucker AM, Weiner JP, Honigfeld S, Parton RA. Profiling primary care physician resource use: examining the application of case mix adjustment. *J Ambulatory Care Manage* 1996;19:60–80.
5. Roblin DW. Physician profiling using outpatient pharmacy data as a source for case mix measurement and risk adjustment. *J Ambulatory Care Manage* 1998; 21:68–84.
6. Lasker RD, Shapiro DW, Tucker AM. Realizing the potential of practice pattern profiling. *Inquiry* 1992; 29:287–97.
7. Fowles JB, Weiner JP, Knutson D, Fowler E, Tucker AM, Ireland M. Taking health status into account when setting capitation rates. *JAMA* 1996;276:1316–21.
8. Normand S–L, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc* 1997;92:803–14.

9. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs of quality of care of a chronic disease. *JAMA* 1999; 281:2098–2105.
10. Kerr EA, Mittman BS, Hays RD, Siu AL, Leake B, Brook RH. Managed care and capitation in California: how do physicians at financial risk control their own utilization? *Ann Intern Med* 1995;123:500–4.
11. Salem–Schatz S, Moore G, Rucker M, Pearson SD. The case for case–mix adjustment in practice profiling. *JAMA*. 1994;272:871–4.
12. Cowen ME, Dusseau DJ, Toth BG, Guisinger C, Zodet MW, Shyr Y. Casemix adjustment of managed care claims data using the Clinical Classification for Health Policy Research Method. *Med Care* 1998;36:1108–13.
13. Schwartz M, Ash A. Evaluating the performance of risk–adjustment methods: continuous outcomes. In: Iezzoni LI, ed. *Risk Adjustment for Measuring Healthcare Outcomes*. 2nd ed. Chicago, IL: Health Administration Press; 1997:391–426.
14. Fowles JB, Weiner JP, Knutson D, Adlis S, Craft C, Fowler E, et al. A comparison of alternative approaches to risk management, Final report to the Physician Payment Review Commission, November 30, 1994.
15. Popovian R, Johnson KA, Nichol MB, Liu G. The impact of pharmaceutical capitation to primary medical groups on the health care expenditures of Medicare HMO enrollees. *J Managed Care Pharm* 1999;4:14–9.

16. Bryce CL, Engberg JB, Wholey DR. Comparing the agreement among alternative models in evaluating HMO efficiency. *Health Serv Res* 2000; 35:509–28.
17. Feinglass J, Martin GJ, Sen A. The financial effect of physician practice style on hospital resource use. *Health Serv Res* 1991;26:183–205.
18. Morris, CN and Christiansen CL. Hierarchical models for ranking and for identifying extremes, with applications. In Bernardo JM, Berger JO, Dawid AP, Smith AFM, ed. *Bayesian Statistics 5*, Oxford University Press; 1996:277–296.
19. Chang W, McCracken SB. Applying case mix adjustment in profiling primary care physician performance. *J Health Care Finance* 1996;22:1–9.
20. Chernew M, Cowen ME, Kirking DM, Smith DG, Valenstein P, Fendrick AM. Pharmaceutical cost growth under capitation: a case study. *Health Aff* 2000; 19:266–76.
21. Crowder MJ, Hand DJ, *Analysis of Repeated Measures*. New York, Chapman and Hall; 1990.
22. SAS Version 6.12, SAS Institute Inc., Cary, NC.
23. McEvoy, GK, ed. *AHFS 97 Drug Information*. Bethesda, MD: American Society of Health–System Pharmacists; 1997.
24. Elixhauser A, Andrews RM, Fox S. Clinical Classifications for Health Policy Research: Discharge statistics by principal diagnosis and procedure, Division of Provider Studies Research Note 17, Agency for Health Policy and Research, Rockville, MD: Public Health Service, 1993. (AHCPH Publication No. 93–0043).

25. Elixhauser A. Clinical classifications for health policy research, version 2: software and user's guide. Healthcare cost and utilization project (HCUP-3), Research Note 2, 1996. (AHCPR Publication No. 96-0046).
26. Schokkaert E, Dhaene G, van de Voorde C. Risk adjustment and the trade-off between efficiency and risk selection: an application of the theory of fair compensation. *Health Econ* 1998;7:465-80.
27. Berwick DB. Continuous improvement as an ideal in health care. *N Engl J Med* 1989;320:53-56.
28. Sandy LG. [letter; comment]. Unreliability of physician "report cards" to assess cost and quality of care. *JAMA* 2000; 283:51-2.
29. Wennberg JE. [letter; comment]. Unreliability of physician "report cards" to assess cost and quality of care. *JAMA* 2000; 283:52-3.
30. Lexchin J. Improving the appropriateness of physician prescribing. *Int J Health Serv* 1998; 28:253-67.
31. Mehl B, Santell JP. Projecting future drug expenditures 2000. *Am J Health Syst Pharm* 2000;57:129-38.
32. Page L. One in 10 California HMO practices are predicted to fail in 1999. *AMNews* September 20, 1999.

**Table 1. Demographics of the populations in the analyses**

	<b>Mid-Western</b>	<b>Western</b>
<b>Number of adult covered lives with continuous enrollment in 1999</b>	52,913	70,979
<b>Mean age in years</b>	41.5 (sd 12.5)	41.1 (sd 11.7)
<b>% Female</b>	52.6 %	53.7 %
<b>Mean Ingredient Costs per member month*</b>	\$ 37.63 (sd 83.05)	\$ 17.87 (sd 49.37)
<b>Median Ingredient Costs per member month*</b>	\$ 6.43	\$ 2.22
<b>% members with no pharmacy costs in year</b>	27.0 %	35.9 %
<b>Number of physicians with minimum panel of 10 patients</b>	209	477
<b>Mean number of patients per physician</b>	253 (sd 264)	149 (sd 197)
<b>Median number of patients per physician</b>	172	70

\*maximum patient costs \$20,000 per year

**Table 2.** Proportion of overall variability attributed to physician random effect. Minimum Physician Panel Size of 10.

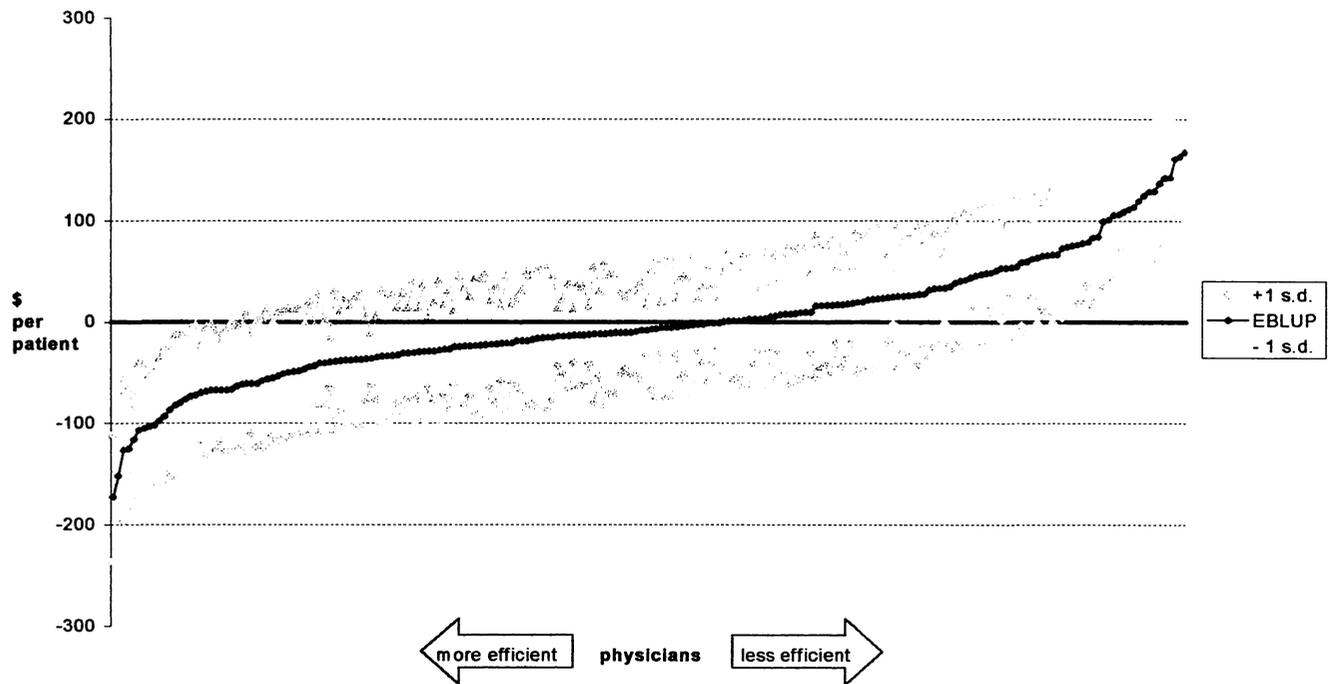
**1999 Mid-Western Population**

	Level of capped costs per pharmacy category	Maximum patient pharmacy expenditures per year	Intra-Class Correlation Coefficient	
			Age-Gender Covariates	Diagnoses and Age-Gender Covariates
1 Mixed Model	\$750	\$6,836	0.018	0.016
	\$2,500	\$11,067	0.014	0.012
	\$20,000	\$19,782	0.010	0.009
28 Mixed Models	\$750	\$6,836	0.006	0.005
	\$2500	\$11,067	0.004	0.004
	\$20,000	\$19,782	0.003	0.002

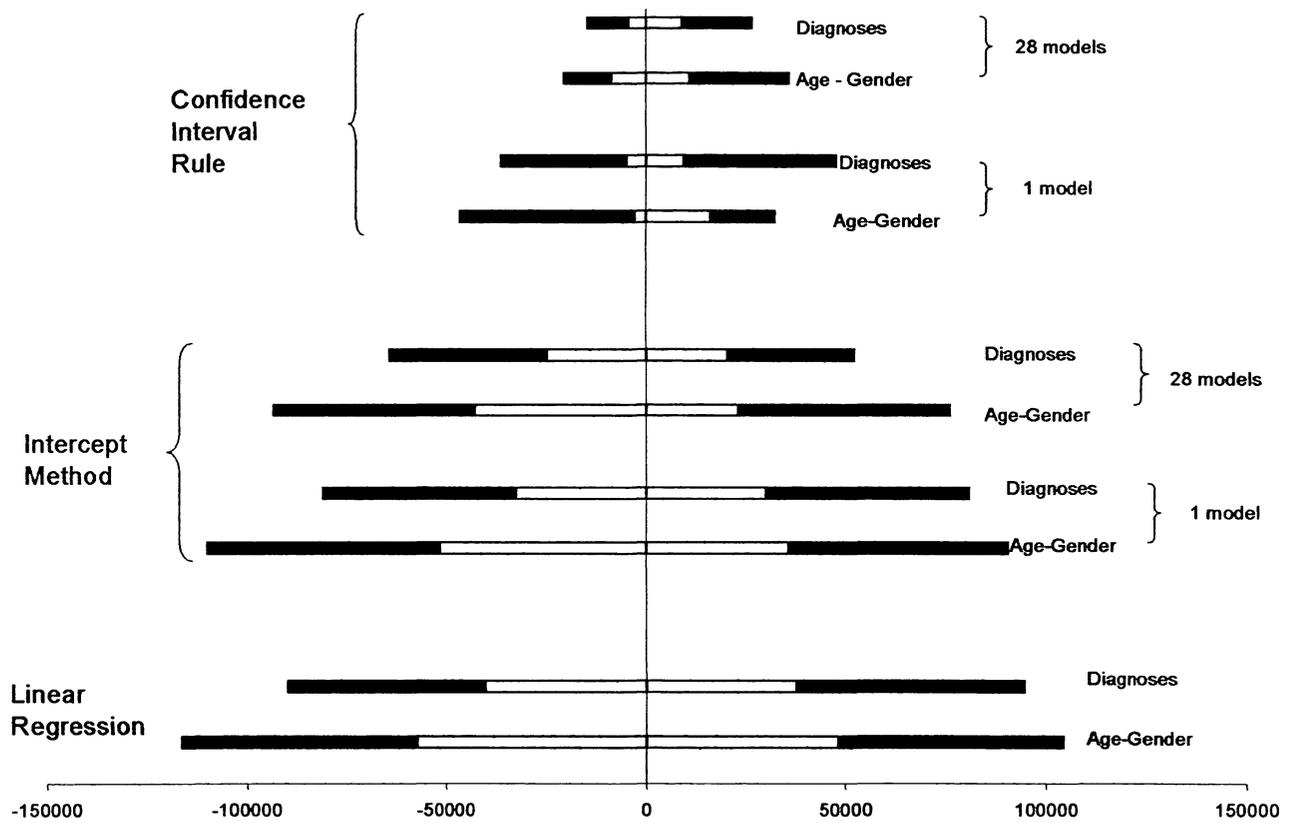
**1999 Western Population**

	Level of capped costs per pharmacy category	Maximum patient pharmacy expenditures per year	Intra-Class Correlation Coefficient	
			Age-Gender Covariates	Diagnoses and Age-Gender Covariates
1 Mixed Model	\$750	\$ 5,703	0.042	0.029
	\$2,500	\$10,033	0.035	0.024
	\$20,000	\$18,914	0.025	0.017
28 Mixed Models	\$750	\$ 5,703	0.007	0.006
	\$2500	\$10,033	0.005	0.004
	\$20,000	\$18,914	0.004	0.003

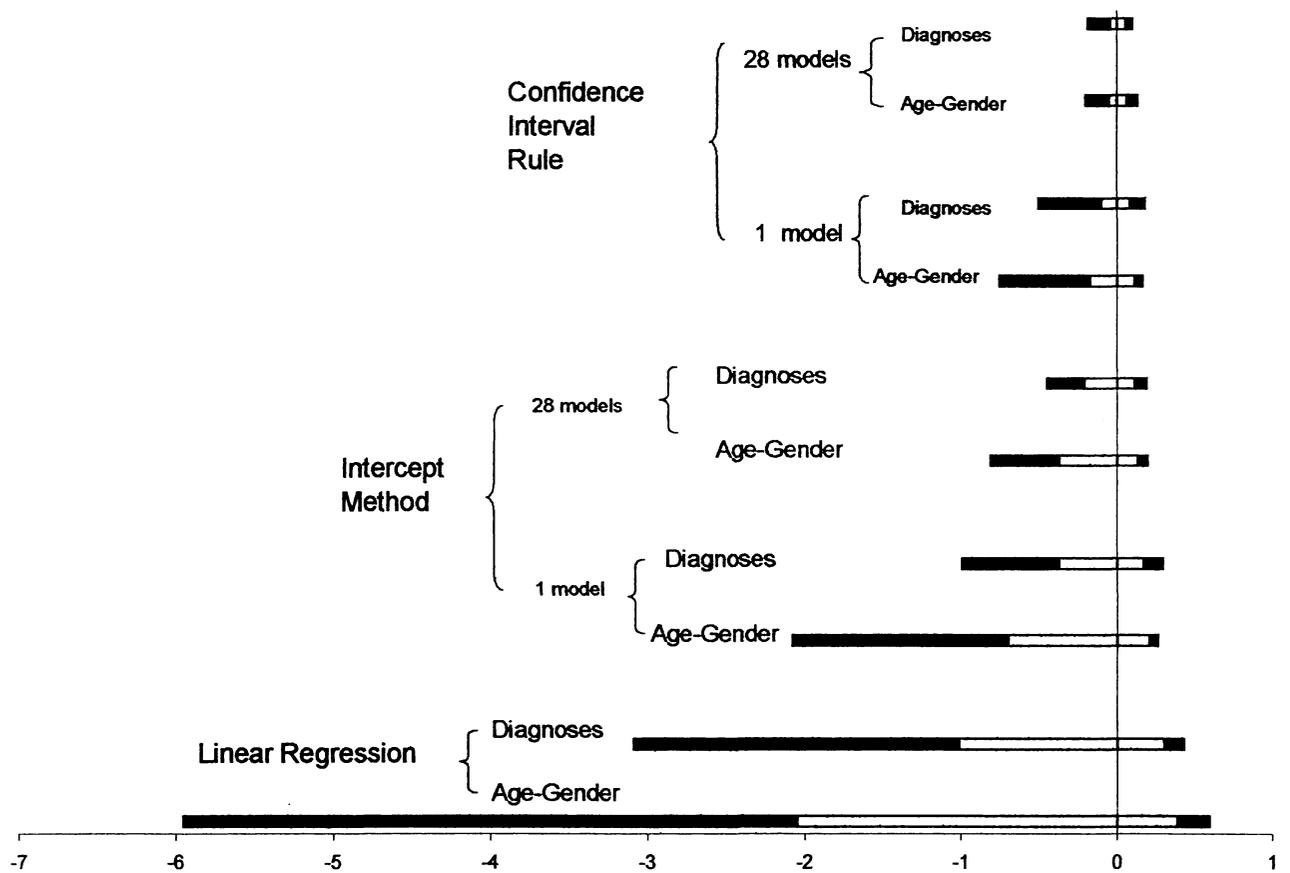
**Figure 1. Individual physicians' empirical best linear unbiased predictors (EBLUPs).** Midwestern population of physicians (n = 209), patient costs per pharmacy category capped at \$2500. This shows the estimate of each of the physicians' effect on his or her patients' costs, surrounded by a one standard deviation confidence interval.



**Figure 2. Range of individual physicians' surplus or deficits projected by the various methods.** Midwestern population of physicians (n = 209), patient costs per pharmacy category capped at \$2500. Negative numbers represent surpluses to physicians, positive numbers are deficits. The dark outer bars encompass the entire range of physician performance. The inner lighter bars represent the range of the middle 90% of physicians.



**Figure 3. Range of individual physicians' surpluses or deficits projected by the various methods, expressed as a proportion of their patients' expenditures. Midwestern population of physicians (n = 209), patient costs per pharmacy category capped at \$2500. Negative numbers represent surpluses to physicians, positive numbers are deficits. The dark outer bars encompass the entire range of physician performance. The inner lighter bars represent the range of the middle 90% of physicians.**



**Figure 4. Range of individual physicians' surpluses or deficits, expressed as a proportion of their patients' expenditures. Restricted scenario where the maximum costs per category were capped at \$750, and each physician had a minimum of 200 patients in his or panel. Negative numbers represent surpluses to physicians, positive numbers are deficits. The dark outer bars encompass the entire range of physician performance. The inner lighter bars represent the range of the middle 90% of physicians.**

