May, 1993

# Small area variation analysis. Methods for comparing several diagnosis-related groups.

Paula Diehr, *University of Washington*

MEDICAL CARE
Volume 31, Number 5, pp YS45–YS53, *Supplement*


73

# Small Area Variation Analysis

## Methods for Comparing Several Diagnosis-Related Groups

PAULA DIEHR, PHD,* KEVIN CAIN, PHD,* ZHAN YE, MD,†
AND FAYEZ ABDUL-SALAM, MS*

**In small-area variation analysis, the variation of health care utilization rates, e.g., admission rates, among small areas is calculated. Frequently, the variation of one diagnosis, diagnosis-related group (DRG), or procedure is compared with the variation of another. Unfortunately, the methods generally used to make these comparisons are not consistent. They differ on whether they 1) adjust for the prevalence of the DRGs, 2) distinguish between variation among areas and variation within areas, 3) weight all areas equally, and 4) adjust for multiple admissions per person. None has an associated confidence interval. These discrepancies occur in part because there is no statistical model of small area variation. Without such a model, it is not known how to measure variation, and thus, it is not known how to compare different DRGs. Here, the authors use data on 473 DRGs from 28 counties in Washington state to study the nature of variability. The variation was higher for the more prevalent DRGs, suggesting that adjusting for prevalence may be reasonable. The true coefficient of variation appears to be a "natural" measure of variation, but the usual small area variation statistics do not provide good estimates of the true coefficient of variation. A new estimate is proposed that can be used to compare and test the variability of several DRGs. (Med Care 1993; 31:YS45–YS53)**

Small area variation analysis (SAVA) is a popular method in health services research. In a typical study, one calculates the hospitalization rate for a particular diagnosis or procedure (we will refer to it as a diagnosis-related group [DRG]) in each of several geographic areas (we will refer to them as counties) and computes some descriptive statistic, such as the coefficient of variation, to show how much the rates vary. These descriptive statistics are calculated for many different DRGs, and factors associated with the higher- and lower-variability DRGs are discussed. One popular hypothesis is that higher variability is associated with DRGs for which there is more uncertainty about the appropriate treatment. A thorough exposition on this approach and its implications has been published recently.[1,2]

The methods for comparing DRGs are not consistent. Wennberg,[1] for example, often shows scatter plots of the admission rate for each small area. On these scatter plots, the

From the Departments of *Biostatistics and †Epidemiology, University of Washington, Seattle, Washington.

Supported in part by the Department of Biostatistics, University of Washington (Seattle, WA), by grant HS06344 to the Back Pain Outcome Assessment Team from the Agency for Health Care Policy and Research and by grant R48/CCR002181-05 from the Centers for Disease Control, Centers for Chronic Disease Prevention and Health Promotion (Atlanta, GA).

Address correspondence to: Dr. Paula Diehr, Department of Biostatistics SC-32, University of Washington, Seattle, WA 98195.

TABLE 1. Descriptive Statistics for Measures of [illegible] for 197 Washington State Diagnosis-Related Groups

| Variable | Average | SD | Minimum | Maximum | Correlation With Preval | Correlation With CVA |
|---|---|---|---|---|---|---|
| Prevalence per 100 k | 47.66 | 93.47 | 10.08 | 906.35 | 1.0000 | −.0977 |
| CVA | 0.34962 | 0.18910 | 0.08625 | 1.52114 | −0.0977 | 1.0000 |
| CVW | 0.37861 | 0.16553 | 0.18414 | 1.44070 | −0.1865[a] | 0.9807[b] |
| Root SCV | 0.47399 | 0.26134 | 0.14972 | 2.08221 | −0.0653 | 0.8857[b] |
| CVU | 0.54944 | 0.20106 | 0.30029 | 1.54560 | −0.2235[b] | 0.8361[b] |

[a] 1-tailed significance—0.01.
[b] 1-tailed significance—0.001.

variation is not adjusted in any way for the prevalence (number of admissions per person per year) of the procedures. However, when Wennberg compares variation numerically, descriptive statistics, such as the coefficient of variation, are presented, which are all adjusted for the prevalence of the procedure (Tables 1 and 2).[1,3] Adjusting for prevalence can have serious consequences. For example, in the data explained subsequently, DRG 373 (uncomplicated vaginal delivery) has the highest standard deviation of the 199 DRGs studied; its coefficient of variation, however, is the 18th smallest. If DRGs are to be labeled as having a "high" or "low" variation, this discrepancy needs resolution.

To our knowledge, there has been no discussion of whether SAVA statistics should be adjusted for prevalence. To address this question, we must 1) define the usual small area variation statistics, 2) propose a statistical model for variation, 3) examine the characteristics of the model using real data, and

TABLE 2. Top and Bottom 20 Diagnosis Related Groups (DRGs)

| | | CVA | | CVW | | Root SCV | | CVU | |
|---|---|---|---|---|---|---|---|---|---|
| DRG | Mean[a] | CV | Rank | CV | Rank | CV | Rank | CV | Rank |
| 53 | 15 | 1.52 | 197 | 1.44 | 197 | 2.08 | 197 | 1.55 | 197 |
| 382 | 27 | 1.19 | 196 | 1.12 | 196 | .97 | 184 | 1.06 | 189 |
| 384 | 46 | 1.17 | 195 | 1.11 | 195 | 1.18 | 193 | 1.09 | 190 |
| 435 | 62 | 0.92 | 194 | 0.87 | 193 | 1.29 | 195 | 1.11 | 192 |
| 436 | 45 | 0.91 | 193 | 0.86 | 192 | 1.07 | 189 | 1.18 | 193 |
| 467 | 24 | 0.90 | 192 | 0.88 | 194 | 1.09 | 192 | 1.30 | 196 |
| 434 | 14 | 0.80 | 191 | 0.78 | 191 | 1.23 | 194 | 1.04 | 188 |
| 231 | 29 | 0.78 | 190 | 0.74 | 189 | 1.41 | 196 | 1.26 | 194 |
| 12 | 12 | 0.77 | 189 | 0.76 | 190 | 1.09 | 190 | 1.03 | 187 |
| 229 | 14 | 0.65 | 188 | 0.65 | 188 | 1.09 | 191 | 1.01 | 186 |
| 266 | 10 | 0.18 | 10 | 0.30 | 63 | 0.24 | 22 | 0.53 | 114 |
| 449 | 35 | 0.17 | 9 | 0.21 | 9 | 0.20 | 12 | 0.34 | 16 |
| 421 | 15 | 0.16 | 8 | 0.26 | 24 | 0.40 | 95 | 0.57 | 135 |
| 261 | 19 | 0.15 | 7 | 0.23 | 12 | 0.29 | 37 | 0.49 | 86 |
| 24 | 34 | 0.15 | 6 | 0.20 | 7 | 0.23 | 18 | 0.37 | 24 |
| 211 | 11 | 0.15 | 5 | 0.28 | 43 | 0.26 | 25 | 0.53 | 113 |
| 378 | 29 | 0.14 | 4 | 0.20 | 5 | 0.26 | 24 | 0.42 | 50 |
| 34 | 11 | 0.12 | 3 | 0.26 | 34 | 0.23 | 19 | 0.52 | 99 |
| 295 | 26 | 0.12 | 2 | 0.31 | 79 | 0.37 | 80 | 0.53 | 109 |
| 204 | 32 | 0.09 | 1 | 0.20 | 6 | 0.22 | 16 | 0.35 | 18 |

[a] Admissions per 100,000 population.

4) compare the performance of three commonly used SAVA statistics. A method for comparing the small area variation of several DRGs is then proposed.

## Descriptive Statistics

One commonly used descriptive statistic, the ratio of the maximum to the minimum rate, has been shown to have undesirable properties and is not considered here.[4–6] Three other descriptive statistics used in SAVA studies—the systematic component of variance (SCV) and the weighted (CVW) and unweighted coefficients of variation (CVU)—are discussed here, and defined more formally in the Appendix. The CV (standard deviation of rates over mean rate) is used, both unweighted and weighted by population. The SCV is an estimate of the (squared) CV among counties after the variation within counties has been removed.[1,6,7] None of these measures has an associated reference distribution. One statistic with a known distribution is a 2 × k chi-square (procedure yes/no by k counties, see Appendix). This test statistic is calculated based on the assumption that the number of admissions in a community has a Poisson distribution, which is approximately true if each procedure represents a different person; this assumption is also made in deriving the SCV. (If a person can have multiple admissions, it is unlikely that the Poisson assumption holds because the Poisson process requires that admissions be independent.[8]) A "multiple admissions factor" (MAF, Appendix) was proposed to correct the SCV and the chi-square statistics for multiple admissions.[6,8]

## A Model for Variation Among Small Areas

There are several types of variation inherent in the problem of SAVA. Rather than modeling the number of admissions per area, we consider a simple model for the number of admissions per person in a single DRG as a function only of the area in which the person lives.

$$\text{Let } Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where $Y_{ij}$ = the number of admissions for person j in area i, $\mu$ =grand mean (prevalence), $\alpha_i$ = the effect of area i, and $\epsilon_{ij}$ = the effect of person j in area i.

There are k counties. Both $\alpha_i$ and $\epsilon_{ij}$ are random effects, with a mean of 0 and variances of $\sigma_A^2$ and $\sigma_e^2$, respectively. The parameter $\sigma_A^2$ is the true variance among counties, and $\sigma_A/\mu$ is the true CV. If each person can have at most one admission and $\sigma_A^2 = 0$, then $\sigma_e^2 = \mu\ (1 - \mu)$, a binomial distribution. If multiple admissions per person are possible, then $\sigma_e^2$ will be larger than $\mu\ (1 - \mu)$, but will probably be approximately proportional to $\mu$. (If $\sigma_A^2 > 0$, then each area will have a different mean and probably a different variance. Here, we assume that $\sigma_e^2$ is approximately the same for each area).

The expectation of the unweighted sample variance $S_u^2$ is $\sigma_A^2 + (\sigma_e^2/k)\ \Sigma(1/n_i)$ (Appendix). If the $n_i$ (population of area i) is very large, then $S_u^2$ is essentially an estimate of $\sigma_A^2$. Otherwise, $E(S_u^2)$ is greater than $\sigma_A^2$. Similarly, $E(S_w^2)$ is greater than $\sigma_A^2$. An estimate of $\sigma_A^2$ can be obtained from the analysis of variance model, using the usual moment estimate ($\hat{\sigma}_A^2$, Appendix). Although the statistical model used to derive the SCV is different from the model used here, the SCV is an estimate of $(\sigma_A/\mu)^2$, or the squared true CV, if each person can be admitted only once. If multiple admissions per person are possible, then the SCV adjusted by the MAF (Cain K et al. Unpublished data, 1993.) is an estimate of $(\sigma_A/\mu)^2$.

The three common SAVA statistics have some important differences, i.e., the CVU and CVW do not account correctly for variation within the counties; the SCV accounts for it, but only under the assumption that there are no readmissions; and the SCV and CVU are unweighted (rates from large and small counties have equal importance), but the CVW gives more weight to the larger

counties. All three statistics, however, adjust the observed variation for the prevalence of the DRG.

There is no literature about the nature of county-level variation and, in particular, about the relationship of $\sigma_A$ to $\mu$. As noted previously, the Wennberg scatter plots imply that the variation among counties should be measured by $\sigma_A$, without adjustment for $\mu$. This suggests either that $\mu$ and $\sigma_A$ are independent or that a difference between two areas of, for example, 20 admissions per 100,000 is equally meaningful whether the prevalence is 10 per 100,000 or 100 per 100,000.

A different relationship is implied by the fact that all three statistics (CVW, CVU, and the square root of SCV) are forms of the CV, which does adjust $\sigma_A$ for $\mu$. The reason for choosing such measures is never stated. One reason could be that the resulting measure is unitless. A more likely reason is based on a feeling that more variation is acceptable around a high mean than around a low mean or that we think of variation in terms of percentages. The use of a CV implies that there is a relationship of $\mu$ to $\sigma_A$ but that it is not interesting. Therefore, $\mu$ should be adjusted out. If $\sigma_A$ is strictly proportional to $\mu$, then the true CV will be independent of the prevalence.

Without theory to guide us, we cannot decide whether or not to adjust for prevalence in comparing several DRGs. We can, however, examine actual data to determine if variation is related to prevalence at all. If it is, we can look for a "natural" way to adjust variation for prevalence and, in particular, to determine whether the coefficient of variation is a reasonable measure of small area variation.

## Methods

We studied the relationship of $\sigma_A$ to $\mu$ for 473 DRGs in 28 counties in Washington state in 1987.[9] Unique identifiers were available for the patients. Therefore, the means and variances of the number of admissions per person could be computed, and from these, the MAFs were calculated. Eleven (of the original 39) counties were excluded because they were very small (population, $< 10{,}000$) or because they were border counties, meaning that some out-of-state hospital admissions were not in our data set. We did this to make the rates more stable so that the variation could be examined independent of the county's size and border problems. For each DRG, we calculated $\hat{\mu}$ and $\hat{\sigma}_A^2$ (Appendix) and plotted $\log_{10}(\hat{\sigma}_A)$ against $\log_{10}(\hat{\mu})$. Consider the equation: $\log_{10}(\sigma_A) = a + b * \log_{10}(\mu)$. Exponentiating and dividing both sides by $\mu$ gives $\sigma_A/\mu = 10^a\mu^{b-1}$. If $b = 1.0$, then the CV is independent of $\mu$. If b is less than 1, then the CV tends to be smaller for more prevalent DRGs than for the less prevalent DRGs, inappropriately making more prevalent DRGs appear less variable.

We studied the association of $\hat{\sigma}_A$ and $\hat{\mu}$ for this data set and for another described subsequently. We computed $\hat{\sigma}_A$, $\hat{\sigma}_A/\hat{\mu}$ = CVA (the analysis of variance estimate of the CV), $S_u$, $S_w$, CVU, CVW, and the square root of the SCV (Appendix). These statistics were then compared using correlations and ranks.

## Findings

### Relationship of $\hat{\sigma}_A$ to $\hat{\mu}$

Figure 1 is a plot of $\log_{10}(\hat{\sigma}_A)$ against $\log_{10}(\hat{\mu})$ for the 199 DRGs with $\hat{\mu}$ greater than one per 10,000. This cutoff was chosen to ensure that at least one admission per year was expected in the smallest county. The $\hat{\mu}$ is based on 4.5 million Washington residents and can be considered known without error. Each point in the plot represents a different DRG. The two highest points are for DRGs 373 and 391, which represent normal births: one for the mother and one for the baby. The straight line is the least-squares regression line. The other line is a "lowess" fit to the data, which does not make any assump-

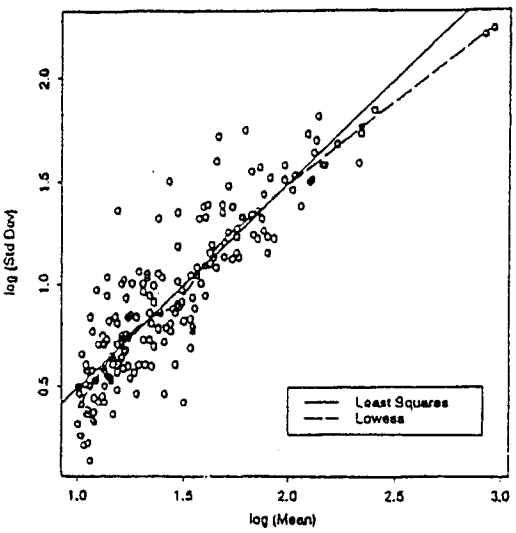


FIG. 1. Log mean versus Log standard deviation for 199 DRG's (min Exp >1).

tions about the form of the relationship.[10] The lowess and the straight line agree well, except possibly for the two extreme points. The regression parameters are estimated as a = −0.50, and b = 0.99895 (standard error, 0.04), and the correlation coefficient is 0.88. If the two highest DRGs are removed, then b = 1.026 (± 0.04). If all 425 DRGs with positive variance estimates are included, b = 0.84 (± 0.02), with the lower slope probably caused by the omission of the negative variance estimates for the lower prevalence DRGs. There is a strong relationship between $\hat{\sigma}_A$ and $\hat{\mu}$. Furthermore, b is remarkably close to 1.0, which means that the true CV is essentially independent of the prevalence. Because DRGs are not strictly independent (for example, DRGs 373 and 391 occur together), the standard errors may be somewhat low.

Similar calculations were made for the regression of $\log_{10}(S_u)$ on $\log_{10}(\bar{Y}_u)$. For all DRGs, b = 0.70 (± 0.01); for the minimum expected number of cases = one, b = 0.87 (± 0.02); and after removing the two highest DRGs, b = 0.87 (± 0.03). In all cases, the slope is significantly different from 1.0. The results were similar for the weighted mean and standard deviation, $S_w$, and $\bar{Y}_w$.

Thus, although the true CV is independent of the mean, the sample CV (CVU or CVW) will be negatively correlated with the mean (because b < 1.0). The results were similar if all 39 counties were used, although, because of negative variance estimates, only 175 DRGs could be used for that analysis.

To test whether this surprising relationship is unique to Washington state DRGs, we analyzed data published by the Health Care Financing Administration (HCFA) for 12 diagnoses in people older than 65 years of age, which studied the variation among 49 states (not Alaska) plus the District of Columbia.[11,12] In this different data set, there was again a strong linear relationship between $\log_{10}(\hat{\sigma}_A)$ and $\log_{10}(\hat{\mu})$. The slope was 0.915, which was not significantly different from 1.0 because there were only 12 data points. We did not have individual-level data and had to assume that $S_u = \hat{\sigma}_A$. This seemed reasonable because states are large, but it may also explain why the slope was a little less than 1.0. The HCFA data also included state-level data for hospital admissions for 15 procedures (e.g., cholecystectomy rates); the slope was significantly less than 1.0.

### What Is "Too Much" Variation?

Ideally, this question would be answered from randomized clinical trials that could provide firm estimates of the percent of a population who should have a particular type of admission. Lacking such standards, one approach has been to compare the variation for several different DRGs, with the implication being that DRGs that have more variation than typical DRGs may have "too much" variation.[2] We have shown here that it may be justifiable to correct a DRG's variation for its prevalence. If this is to be done, then an appropriate way to identify outliers is to examine a plot such as Figure 1 and

label points (DRGs) that are "too far" from the line as outliers. Because b = 1, it is equivalent to use the CV from analysis of variance (CVA) to describe the variability among counties. In the current study, we calculated the CVA, CVU, CVW, and the square root of the SCV for each DRG. We then ranked the DRGs by their CVA value. Two DRGs were dropped because the SCV was negative, leaving 197 for analysis.

Table 1 shows descriptive statistics for measures of interest. The prevalence ranged from ten to 906 per 100,000 population. The CVA averaged 0.35 with a standard deviation of 0.19. The CVW, root SCV, and CVU had means of 0.38, 0.47, and 0.55, respectively. Table 1 also shows the correlation of the four measures with prevalence and with the CVA. As expected, CVW and CVU were significantly negatively correlated with the prevalence; the CVA and the root SCV were not. The three usual CV estimates (CVW, root SCV, and CVU) were highly correlated with the CVA; the CVW was best and the CVU, worst.

Table 2 shows the agreement of the estimates in a different way. It lists the ten most variable DRGs and the ten least variable DRGs, based on the CVA. The ranks for the other CV measures are also shown. The first line shows that DRG 53, "sinus and mastoid procedures, age ≥ 18," had a prevalence of 15 per 100,000. It had the highest CVA, 1.52. The CV was estimated as 1.44 by the CVW, as 2.08 by the SCV, and as 1.55 by the CVU. Although the CV estimates differed, all four ranked DRG 53 as the most variable (rank, 197 of 197). Only one of the top 20 DRGs (DRG 383, not shown) was not in the top 20 of the CVW values (it was ranked 21st highest). The SCV did not identify four of the top 20, and the CVU did not identify six of them. However, the agreement was reasonably good. The ten DRGs with the lowest variation are also shown in Table 2. The lowest is DRG 204, "disorders of the pancreas except malignancy," which had a CVA of 0.09. The other three measures ranked DRG 204 as sixth, 16th, and 18th, respectively. The CVW did not identify nine of the lowest 20, and the SCV and CVU did not identify 14. The agreement in rank was thus worse for the DRGs that had low variation.

A histogram of the CVA (not presented) shows nine reasonably high outliers, corresponding to the first nine DRGs in Table 2. These DRGs can be said to have too much variation because they are far from the regression line shown in Figure 1. That is, they do not behave like most of the other DRGs.

### The Chi-Square Statistic

We have noted elsewhere that tests for too much variation based on the chi-square statistic had better performance than those based on the SCV, CVU, or CVW.[4–6] The chi-square statistic also is useful in the current context because the $CVA^2$ can be written as:

$$1)\ (CVA)^2 = [\chi^2 - (k-1)^*MAF]/[\mu^*(k-1)^*n_o].$$

(See Appendix for definitions. The average MAF for DRGs in Washington state is 1.1, with a maximum value of 4.1. Researchers who have data with personal identifiers can, of course, calculate the CVA directly without using these formulas.)

Equation 1 yields an approximate 95% confidence interval for $(\sigma_A/\mu)^2$, based on the noncentral chi-square distribution as follows:

$$2)\ CVA^2 \pm (1.96)^*(4^*\chi^2 - 2^*(k-1))^{0.5}/(\mu^*n_o).$$

This confidence interval, which is valid when either k or $\sigma_A/\mu$ is "large," can be used to test whether two DRGs have significantly different variation. In our data, for example, the confidence intervals for DRGs 53, 382, and 384 were (1.44, 1.60); (1.12, 1.25); and (1.13, 1.22), respectively. The small area variation of DRG 53 was thus significantly

greater than that of the other two DRGs, which were similar. A small simulation study suggests that this confidence interval behaves well when the minimum number of expected cases (under the alternative hypothesis) is greater than 1. When the DRGs were ranked on the lower limit of their confidence intervals, the rankings changed somewhat.

## Discussion and Summary

We have demonstrated that, for reasonably prevalent DRGs in 28 "large" (population > 10,000) nonborder Washington counties, the natural measure of variation among counties for a particular DRG is the CVA. The results also showed that the usual measures of the CV (CVU, CVW, and root SCV) are highly correlated with the CVA but tend to be larger, more variable, and negatively correlated with the prevalence. We recommend that the CVA be used as the statistic of choice in studying DRGs with "reasonably high" prevalence rates. However, our results do not clearly delineate whether or not it is a good statistic for the lower prevalence DRGs or for procedures. The CVA can be calculated using the chi-square approximation, which also yields an approximate confidence interval. The SCV, if corrected for multiple admissions, is also an estimate of $(\sigma_A/\mu)^2$. The SCV does not perform well for the low prevalence DRGs, probably because it is influenced by the smallest communities. The version of the SCV used by Wolfe's group[3,13] should provide a more stable estimate of $\sigma_A/\mu$, which can then be corrected for multiple admissions. Although the methods have some intrinsic differences, the relative rankings of all methods are similar for the most variable DRGs. Furthermore, if all the small areas have large populations, all the estimates will be essentially the same because differences in weighting and multiple admissions will have little effect.

There are, of course, limitations to these findings. Some have to do with the methods results and some, with the substantive results. Although we used it as a "gold standard," the CVA is, of course, only an estimate of the true CV. It attempted to account for all the important factors, but it did require some assumptions. Although we did not assume that the number of admissions per county had a Poisson distribution, we did assume that the variance of the number of visits per person was the same for each county, which is probably not strictly true. However, because of the limited amount of data available in the smallest counties, any estimate that allowed each county to have a different variance would have required other strong assumptions, which could not be verified in the smaller counties. It will be interesting to repeat the analysis for some of the DRGs using other estimates of the true CV.

We do not take the substantive results for particular DRGs very seriously for several reasons. The data were from a subset of counties in Washington state with a lower bound on prevalence. Analyses based on different DRGs, in larger or smaller counties, or in other settings might provide different findings (although the results from the HCFA data were encouraging). Patient age and sex were not controlled. In analyses (not shown) of age- and sex-adjusted rates, the slope (b) is still 1.00, but the mean CVA is 0.30 rather than 0.35. Furthermore, we used single DRGs rather than meaningful groups of DRGs or International Classification of Disease, ninth revision codes as the unit of analysis. We made these choices to ensure common calculation methods for each DRG, but they probably are not the choices that would have been made in a study of any particular diagnosis.

We have found an apparent "natural law" that seems to hold when the minimum expected number of admissions is at least one. On average, the true standard deviation among counties is approximately one third of the mean (prevalence). It may be important that the slope of $\log_{10}(\hat{\sigma}_A)$ against

$\log_{10}(\hat{\mu})$ is so close to 1.0 (Fig. 1). There are probability distributions, such as the exponential and log-normal, in which the mean and standard deviation are proportional. If, for each DRG, m counties have a high mean admission rate and the others have a low mean and if the group means are far enough apart, the CV will be approximately the same for all DRGs. These or other models may lend insight into the nature of small area variation.

We recommend that the CVA be used as a measure of variability when DRGs are compared and the expected number of admissions in the smallest county is at least one. Our findings may also suggest a class of models that will be useful in modeling small area variation. New research is needed to extend these results to procedures (versus diagnoses) and to different data sets to examine the effect of age and sex and to study the properties of the confidence interval suggested for the CVA in Equation 2 of this study.

## References

1. Wennberg J. Small area analysis and the medical care outcome problem. In: Sechrest L, Perrin E, Bunker J, eds. AHCPR Conference Proceedings, Tucson, Arizona, April 8–10, 1987: 177. Bethesda, MD: Public Health Service, publication no. (PHS) 90-3454, 1990.

2. Roos N, Wennberg J, McPherson K. Using diagnosis-related groups for studying variations in hospital admissions. Health Care Financing Review 1988;9:53.

3. McMahon L, Wolfe R, Tedeschi P. Variation in hospital admissions among small areas: A comparison of Maine and Michigan. Med Care 1989;27:623.

4. Diehr P, Cain K, Connell F, et al. What is too much variation? the null hypothesis in small area analysis. Health Serv Res 1990;24:741.

5. Diehr P, Grembowski D. A small area simulation approach to determining excess variation in dental procedure rates. Am J Public Health 1990;80:1343.

6. Diehr P, Cain K, Kreuter W, et al. Can small-area analysis detect variation in surgery rates? The power of small-area variation analysis. Med Care 1992;30:484.

7. McPherson K, Wennberg J, Hovind O, et al. Small-area variations in the use of common surgical procedures: An international comparison of New England, England, and Norway. N Engl J Med 1982;307:1310.

8. Cain K, Diehr P. Testing the null hypothesis in small area analysis. Health Serv Res 1992;27:267.

9. Washington State Department of Health Comprehensive Hospital Abstract Reporting System (CHARS). 1987.

10. Chambers JM, Cleveland WS, Kleiner B, et al. Graphical Methods for Data Analysis. North Scituate, MA: Duxbury Press, 1983.

11. Department of Health and Human Services. Special report: Hospital data by geographic area for aged Medicare beneficiaries. Vol. 1. Baltimore, MD: Health Care Financing Administration, HCFA publication number 03300, 1990.

12. Department of Health and Human Services. Special report: Hospital data by geographic area for aged Medicare beneficiaries. Vol. 2. Baltimore, MD: Health Care Financing Administration, HCFA publication number 03301, 1990.

13. Wolfe R, Petroni G, McLaughlin C, et al. Empirical evaluation of statistical models for counts or rates. Stat Med 1991;10:1405.

# Appendix

## Definitions of Symbols

### Parameters

$\mu$ = grand mean (prevalence)

$\sigma_A^2$ = variance among communities

$\sigma_A/\mu$ = true coefficient of variation

$\sigma_e^2$ = variance among persons, within community

$\sigma_e^2/\mu$ = multiple admission factor (MAF) = true variance/Poisson variance

### Sample Statistics

$Y_{ij}$ = number of admissions for person j in county i (k counties total)

$\bar{Y}_{i.}$ = sample mean for county i

$\bar{Y}_u$ = unweighted sample mean = $\Sigma\bar{Y}_{i.}/k$

$\bar{Y}_{..}$ = grand mean = weighted sample mean = $\Sigma n_i\bar{Y}_{i.}/\Sigma n_i$

$S_u^2 = \Sigma(\bar{Y}_{i.} - \bar{Y}_u)^2/(k-1)$

$S_w^2 = \Sigma[n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2]/(\Sigma n_i - 1)$

### Estimates

$\hat{\sigma}_A^2 = (MSA - MSW)/n_o$, where MSA and MSW are the mean square among and mean square within from the analysis of variance, $n_i$ is the number of residents in county i, $\bar{n}$ and $s_n^2$ are the mean and variance of the number of people per area, and $n_o = \bar{n} - s_n^2/(k^*\bar{n})$.

$\hat{\mu}$ = weighted sample mean = $\Sigma\Sigma Y_{ij}/\Sigma n_i = \Sigma n_i\bar{Y}_{i.}/\Sigma n_i$

CVA = analysis of variance estimate of "true" coefficient of variation = $\hat{\sigma}_A/\hat{\mu}$

CVU = $S_u/\bar{Y}_u$

CVW = $S_w/\hat{\mu}$

SCV = systematic component of variance = $(1/k)\,[\Sigma((O_i - E_i)^2)/E_i^2 - \Sigma(1/E_i)]$

$\chi^2$ = chi-square = $\Sigma(O_i - E_i)^2/E_i$ (for prevalence small), $k - 1$ degrees of freedom.