4-30-1994

An Empirical Study of Small-Area Variation for ICD-9 Surgical Procedures

Paula Diehr
University of Washington, pdiehr@u.washington.edu

Kevin Cain

Zhan Ye

John Loeser

An Empirical Study of Small-Area Variation

For ICD-9 Surgical Procedures

Paula Diehr, Ph.D.(1)
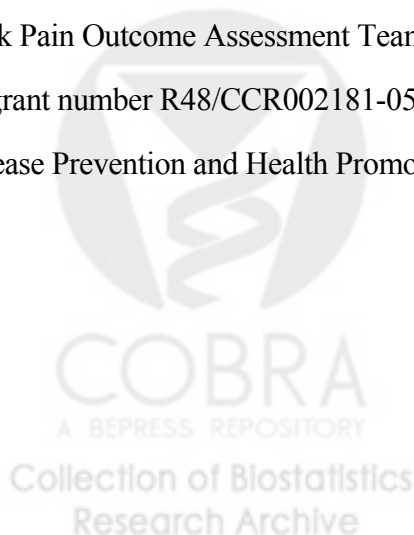
Kevin Cain, Ph.D.(1)

Zhan Ye, M.D.(2)

John Loeser, M.D. (3)

July-11-1994.

Key words: Small area analysis, prevalence, procedure, coefficient of variation.

From the departments of (1) Biostatistics, (2) Epidemiology and (3) Medicine at the University of Washington. Address comments to Dr. Diehr, Dept of Biostatistics SC-32, University of Washington, Seattle, WA 98195. (206) 543-8004.

An Empirical Study of Small-Area Variation

For ICD-9 Surgical Procedures

ABSTRACT

Objective. Several measures of variation have been used in SAVA. One study of DRGs found that the coefficient of variation from analysis of variance (CVA) had superior performance. That work is replicated here for ICD-9 surgical procedures, and extended to age/sex-standardized rates. Results are compared with those in the literature, and recommendations are made for assessing small-area variation in future studies.

Data Sources. Data were taken from Washington State's "Episode of Illness" file of hospital discharges in the State in 1987. Up to three ICD-9 surgical procedures and a unique patient identifier were available for each discharge.

Study Design. We calculated the usual small-area variation statistics for 153 different surgical procedures, among 28 counties in Washington state, with and without standardization for age and sex. We tested each variation statistic to determine whether it was correlated with the prevalence of the surgical procedure, and to see which statistic was most correlated with the true variation among counties. We used the empirical results to provide guidelines on how much of the observed variability among counties is likely to be due to chance alone.

Principal Findings. As in the previous study of DRGs, the CVA was uncorrelated with the prevalence and highly related to the true variance of a procedure, whereas the other statistics performed less well. The age/sex-standardized CVA's were usually smaller than CVA's based on the raw data. Confidence intervals for the CVA, either crude or standardized, can be calculated from the appropriate chi-square statistic. Typical CVs were similar to those published elsewhere, in the range of .3 to .5.

Conclusions. The CVA is the statistic of choice to measure small-area variation for procedures, with or without age/sex standardization. A CVA larger than .5 should be considered as "high" variation. In studies where fewer than 10 procedures are expected in the smallest area, the observed variation is likely to be primarily due to chance, and should be tested statistically before small-area analyses are conducted to determine the reason for the variation. The magnitude of the variation statistics here are similar to those published elsewhere, suggesting that these findings can be generalized to other settings.

INTRODUCTION

Small-area variation analysis (SAVA) is a popular methodology in outcomes research. In a typical study one calculates the hospitalization rate for a particular procedure or diagnosis in each of several geographic areas (we will refer to them as counties), and computes some descriptive statistic such as the standard deviation or coefficient of variation to show how much the rates vary among

counties. These descriptive statistics are calculated for many different procedures, and higher- and lower-variability procedures are identified. One hypothesis is that higher variability is associated with procedures for which there is more uncertainty about appropriate treatment. Thorough expositions on this approach and its implications have been published recently [1,2].

The measures of variation used in these studies have not always been consistent. Some studies consider only the standard deviation of the rate, calculated across counties, while others use some form of the coefficient of variation (the standard deviation among counties divided by the prevalence of the procedure). Adjusting for prevalence can have serious consequences. For example, in Washington State, DRG 373 (uncomplicated vaginal delivery) had the highest standard deviation of the 199 DRGs studied; its coefficient of variation, however, was the 18th smallest [3]. If procedures are to be labelled as "high" or "low" variation, this discrepancy needs resolution.

In previous work we proposed a model of variation at the person level, and used that model to study 197 DRGs. We found that the standard deviation and the prevalence were strongly correlated, and that the usual descriptive statistics were negatively correlated with prevalence [3]. We recommended that a new statistic, the CVA (coefficient of variation estimated from analysis of variance), be used to describe variation. The CVA was found to be independent of the prevalence of the DRG. It was also shown to be related to the chi-square statistic, meaning that it has high power to detect variation when it occurs [4].   In addition, a confidence interval can be computed so that one DRG can be compared to another.

The previous work addressed only diagnostic categories (DRGs) and considered raw rather than age/sex-standardized rates. In this paper we extend the previous work to study 153 ICD-9 surgical procedures codes, and extend the CVA to standardized rates. Some substantive results are shown and compared to others in the literature. These findings are used to identify situations in which data bases are so large that the recommended statistical procedures may safely be ignored.

METHODS

Hosted by The Berkeley Electronic Press

Data were taken from Washington State's "Episode of Illness" file which has one record for every hospital discharge in the State in 1987 [5]. Up to three ICD-9 surgical procedures are available for each discharge. Unique identifiers were available for patients, so that means and variances of the number of admissions per person could be computed, and from these the multiple admission factors (MAF's) were calculated [6].

Eleven (of the original 39) counties were excluded because they were very small (population less than 10,000) or because they were border counties, meaning that some out-of-state hospital admissions are not in the data set. We did this to make the rates more stable so that the variation could be examined independent of county size and border problems.

We studied 153 ICD-9 procedures. These include individual procedure codes that had a prevalence of at least 1 per ten thousand (so that one or more procedures were expected in each county). In addition we studied 12 combinations of procedures defined and studied by HCFA [7] and 8 combinations of procedures studied (but not defined) by Gittelsohn [8]. (See Appendix A for procedure group definitions).

For each procedure, for each county, we calculated the prevalence rate (number of procedures to people who lived in the county divided by the county population). We also estimated the "true" variance among counties; that is, the variance remaining after subtracting the variance within counties from the observed variance among counties (see Appendix B for definitions of terms). The true coefficient of variation (CV) for a particular procedure is the standard deviation (among counties) of the true rates divided by the prevalence. There are several ways of estimating the CV, which differ on whether counties are weighted by their population, and whether the sample variance or the "true" variance (subtracting variation within counties) is used to compute the standard deviation. We computed four different measures of the coefficient of variation for each procedure: the CVU (unweighted standard deviation over unweighted mean or prevalence); CVW (weighted standard deviation over weighted mean); CVA ("true" standard deviation over weighted mean); and RSCV, the square root of Wennberg and MacPherson's SCV (see Appendix B). These

statistics were then examined to determine whether they were correlated with the prevalence, or with the true variation among counties.

The relationship of the CVA to the chi-square statistic is:

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

An approximate 95% confidence interval for $(\sigma_A/\mu)^2$, based on the non-central chi-square distribution (see Appendix C), is:

$$CVA^2 \pm (1.96)*(4*\chi^2 - 2*(k-1))^{.5}/(\mu*(k-1)*n_o).\qquad\qquad \{2\}$$

This formula was used to compute confidence intervals for the CVA for each procedure. (The "k-1" term in the denominator of {2} was inadvertently omitted in an earlier paper [3]).

Most SAVA studies use age/sex-standardized rates. We calculated indirectly standardized rates for each procedure for each county. One test to determine whether standardized rates are the same in all counties is the Mantel-Haenszel chi-square test. This is similar to the unstandardized chi-square except that the expected numbers are computed separately for each stratum and then summed to get the expected number for the county, before computing the chi-square statistic. It is not clear how the variance among counties of age/sex standardized weights should be calculated. By analogy, we substituted the Mantel-Haenszel chi-square for the unstandardized chi-square in equations {1} and {2}, to provide an estimate and confidence interval for CVAS, the standardized CVA.

The analytic strategy was to compare the various estimates of the coefficient of variation, testing first whether they were independent of prevalence and whether they were highly related to the true variance among counties. We compared the raw and the standardized CVA measures to determine if standardizing made any difference. We then examined the distribution of the standardized CVA, and noted which procedures had significantly higher variation than was "typical". Some of these results were then compared to those in the literature.

FINDINGS

There were data from 28 counties, whose populations ranged in size from 10,000 to 1.4 million, and are given elsewhere [9].

Comparison of Different Measures of the Coefficient of Variation

We estimated the true standard deviation, $\hat{\sigma}_A$, and the prevalence, $\hat{\mu}$, for 153 different ICD-9 procedures or combinations of procedures (not standardized). (See Appendix B for definitions). As was true for the DRGs [3], log($\hat{\sigma}_A$) was highly correlated with log($\hat{\mu}$) (r=.84, slope=.954), suggesting that the coefficient of variation was a reasonable measure of variation. Table 1 shows descriptive statistics for the 5 estimates of the CV defined in Appendix B. The average values of the statistics range from .40 (for CVAS) to .65 (for CVU). The largest individual value is 2.57, for the CVU. The CVA is lower on average than the other measures, and CVAS is lower yet. Both the CVU and the RSCV give equal weights to every county, which may explain their high values. CVA and CVAS give somewhat more weight to the larger counties, and CVW gives very high weight to the largest counties.

[Table 1 about here]

Table 1 also shows the correlation of each CV estimate with , the prevalence of the procedure. CVU and CVW are significantly negatively correlated with the prevalence (one-tailed test). This means that they tend to be too small for the more prevalent procedures. The RSCV, CVA and CVAS are not significantly correlated with the prevalence. The second column shows the correlation between the CV estimate and $\hat{\sigma}_A$. All measures are significantly correlated, but the correlation is highest for the CVAS and is substantially lower for the RSCV. Since CVA and CVAS are not significantly correlated with the prevalence, and are highly correlated with the true county variance, we argue that the CVA had superior performance as compared to the other CV measures.

Age/sex Adjustment

One question of interest is whether age/sex adjustment is even necessary when examining county variation, since large areas such as counties are unlikely to vary greatly. We computed the ratio of CVA to CVAS for each procedure. The average of that ratio is 1.13, showing that the

standardized measure of variation is smaller on average. The ratio varied from .71 to 2.03, meaning that some procedures showed even more variation after age/sex adjustment than before adjustment. The remainder of this paper will deal only with CVAS.

Small-Area Variation for Procedures

Figure 1 is a histogram of CVAS for the 153 procedures. The CVAS has mean .402, median .313, and ranges from 0 to 2.29. Eight procedures have CVAS's that are apparent outliers. In some sense, these are the procedures with the greatest small-area variation, since their variation is much higher than that of "typical" procedures.

[Figure 1]

The most variable procedure is procedure 57.94 (insertion of indwelling catheter) which has CVAS=2.29. However, the expected number of cases in the smallest county is only 3.5, and some counties had no procedures. Since the chi-square statistic may not have a chi-square distribution when expected values are less than 5, we restricted the following analyses to procedures with at least 5 cases expected in the smallest county.

In Table 2, the most variable procedure is 75.34 (fetal monitoring NOS). The state prevalence is 88.1 per 100,000 and county rates ranged from 0 to 749 per 100,000. The CVAS (standardized CVA) is 2.01, with an approximate 95% confidence interval of 1.98 to 2.04. This procedure is "significantly" more variable than the next on the list, 75.32 (fetal EKG) whose CVAS is between 1.37 and 1.45. (The CVAS's for different procedures are calculated from some of the same people, using the same counties, and some of the procedures are repeated. They are thus not strictly independent, and these tests should only be taken as guidelines).

[Table 2 about here]

The first 11 procedures have "high" variation in the sense that the 95% confidence intervals for the CVAS do not include .313 (the median CVAS for all procedures). The following seven and procedure 39.61 are not significantly different from the median. And, the remainder are

significantly less variable than the median.   Many of the procedures in Table 2 have to do with childbirth, since it is the most frequent reason for hospitalization. We repeated some of these runs after removing the obstetrical procedures (72.00-75.99) and three diagnostic procedures (37.22, 37.23, and 3.31). The unstandardized CVA became significantly related to the mean, but the standardized CVA remained uncorrelated with the mean.

SUMMARY AND DISCUSSION

For reasonably prevalent procedures in 28 "large" (population > 10,000) non-border Washington counties, the CVA had better performance than the usual measures of the coefficient of variation, CVU, CVW, root SCV, which were highly correlated with the CVA, but tended to be larger and more variable than the CVA, and were negatively correlated with the prevalence. Age/sex-standardized means seemed to vary less than unstandardized rates, suggesting that age/sex-adjustment should be used. We thus recommend that the CVAS be used as the statistic of choice in studying procedures with "reasonably high" prevalences. We do not know whether it is a good statistic for the lower prevalence procedures.   Equations {1} and {2} can be used (with the Mantel-Haenszel chi-square) to estimate the CVAS and its confidence interval.

We compared these findings to Wennberg's data on 30 surgical procedures in 16 hospital market areas [1]. Although there was insufficient information to calculate the CVA, we could compare the CVW and the RSCV. Wennberg's average CVW was .36, versus our .45; the median CVW's were .35 and .38 respectively. Similar statistics for the RSCV are .31 versus .56 and .31 versus .47. We also compared variation statistics for DRGs [1,3]. For the CVW, the means were .43 versus .38, and the medians were .40 versus .33. For the RSCV, the means were .39 versus .47 and medians were .38 versus .41. The mean CVW thus varied from .36 to .45 and the median varied from .33 to .40. For the RSCV, means varied from .31 to .56 and medians from .30 to .47.

Thus, in two very different settings, using different groups of DRGs or procedures, all "typical" measures of the coefficient of variation were between .3 and .57, with the RSCV showing

8

more variation than the CVW. It seems likely, then, that typical values for the CVA will also be generalizable to other settings. We found the mean and median CVA for procedures to be .44 and .34, respectively. For DRGs, the mean and median CVA were .35 and .30 [1]. For the age/sex-standardized measure, CVAS, the mean and median were .31 and .26 for DRGs, and .40 and .30 for surgical procedures. This suggests that a procedure with a CVA or CVAS of .5 or more can be considered as having high variability. It will be interesting to test this speculation in other settings.

We do not take the results for particular procedures very seriously, for several reasons. The data were from a subset of counties in Washington state, with a lower bound on prevalence, and included only non-federal hospitals. Analyses based on different procedures, in larger or smaller counties, or in other settings, might provide different findings. Due to the high volume of discharges analyzed, we were unable to check the data for accuracy. A few random miscodes would not affect these results, but systematic differences in coding in the smaller counties could well be associated with the differences that were seen. Further, we used mostly single procedures, rather than meaningful groups of procedures as the unit of analysis. We made these choices to ensure common calculation methods for each procedure, but they are probably not the choices that would have been made in a study of any particular procedure.

It was interesting to note, however, that the procedures that displayed high variation in our data were generally not the same as those found elsewhere. For 18 procedures we compared the various CV measures from this paper to those presented by Wennberg [1]. There was not good agreement in the relative amount of variation. In particular, Wennberg usually finds inguinal hernia (our Procedure C) to have low variation, whereas it is significantly above the median for our data (95% CI = .38 to .49). Possible reasons are that we defined the procedure groups differently, or that the denominators that we used (entire county population) were different. It may also be that the high-variation procedures are not the same in different regions of the country.

However, it is also possible that differences in statistical methods account for some of these discrepancies. We limited our analysis to counties with at least 1 procedure expected per year, and

used the CVA rather than the RSCV or CVW to measure variation. It may be that other published literature analyzed counties with smaller expected values, and that the RSCV and CVW were measuring in part variations in population sizes or prevalences as well as the true variation among counties. Better definitions should be provided in papers about small-area variation, if their results are to be compared or replicated.

We recommend that the CVA or CVAS be used as a measure of variability when either DRGs or procedures are compared and the expected number of admissions in the smallest county is at least 1. We further speculate that a CVAS above .5 can be considered as "high".

Implications for Small-Area Variation Analysis

One appeal of small-area variation analysis is that it is easy to perform, and the results seem obvious. Unfortunately, high-appearing variation can be caused by small numerators, small denominators, poor choice of descriptive statistics (notably highest rate divided by lowest rate, which still seems to be the informal measure of choice), and by multiple admissions per person, even if there is no true underlying variation [9]. In addition, an area's observed rate will vary somewhat from year to year, especially if that area is small and the procedure is infrequently performed. This "random" component of variance is easy to confuse with the systematic variation among counties which is of interest. Many people have an interest in performing SAVA on low prevalence procedures, for areas with small populations, or where multiple admissions are possible. Since it is possible to obtain large-appearing variation by chance alone, it is good practice to test formally whether there is more than chance variation. Although the null hypothesis (that $\sigma_{At} = 0$) is probably never strictly true, it is not productive to try to "explain" observed variation that can not be distinguished from chance.

Investigators with large data bases often find that their small-area variation is enormously statistically significant, and wonder if such analyses are necessary. For example, in our study, all but three of the 153 chi-square statistics were significant, meaning that all but three of the CVAS's were significantly different from zero. When is a formal test necessary? We used some of the empirical

findings to address this question.

McPherson [11] developed a method to separate the observed coefficient of variation (CVUW: the unweighted variation about the weighted mean [4]) among areas into a systematic component of variance (SCV) and a random component (RCV). The systematic component (which is the squared "true" coefficient of variation [3]) is estimated by McPherson as SCV = CVUW^2 - RCV , or

$$\text{SCV} = (1/k)\, 3((O_i\text{-}E_i)^2)/E_i{}^2 - (1/k)3(1/E_i) \qquad \{3\}$$

where E_i and O_i are the usual expected and observed number of admissions in area i, and there are k areas. Equation {3} illustrates the importance of area population and prevalence (whose product is the expected number of admissions), and also the effect of multiple admissions. Note that the first term on the right side of equation {3} (CVUW^2) is sensitive to low expected values, since it gives all counties equal weight. The SCV sometimes has rather erratic behavior [3,4,9], probably for that reason.

The second term on the right of {3} is an estimate of the RCV, under (essentially) the assumption that there are not multiple admissions. A more general estimate is RCV = $(\text{MAF}/k)3(1/E_i)$, where MAF is the multiple admission factor [6,12]. Considering the two situations in which all counties have the same expected value, or where one county has a very small expected value, the following inequality can be shown to hold:

$$\text{MAF}/(k\, E_{min}) < \text{RCV} < \text{MAF}/(E_{min}), \qquad \{4\}$$

where E_{min} is the number of admissions expected in the smallest of k areas.

We used inequality {4} to determine some bounds on the random component of variation. Letting k = 28, we can calculate the percent of observed variation "typically" due to chance (in 28 counties in Washington) as RCV/(SCV+RCV)*100%. We calculated this value letting the true CV = .313 (the median) and .402 (the mean CVAS); letting MAF = 1 and 4 (the largest value tabled [6]), and letting E_{min} vary from 1 to 1000. We let the RCV vary from the low to the high limits in inequality {4}.
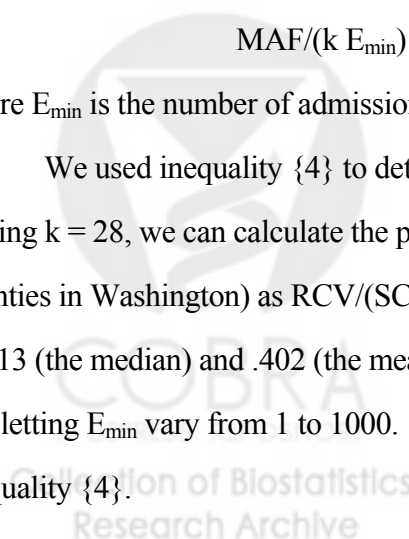
11

Table 3 shows the percent of the observed variability that is "typically" due to random variability as a function of the true CV, the MAF, E_{min} (the number of procedures expected in the smallest county), and MAF. For example, in column 1, where CVAS = .313, MAF=1 (no readmissions), and the number of admissions expected in the smallest county (E_{min}) = 1.0, then 26.72 to 91.07% of the observed variability is due to random variation. Clearly, hypothesis testing would be crucial in such situations. For E_{min} = 1000, however, less that 1 percent of the observed variation is likely due to random variation, and formal statistical testing may not be necessary. One might even feel safe with E_{min} = 25, since then only about a quarter of the observed variation would be due to chance.

For MAF = 4.0, and for E_{min} = 1, 59.31 to 97.60% of the observed variation will be due to random variation; even for E_{min} = 100, as much as 29% of the observed variation might be due to chance. Again, with E_{min} = 1000, the amount of chance variation is minimal. It is thus clear that the random component of the variance and the multiple admission factor may be very important if E_{min} is low. Results are similar but less extreme if the "true" SCV is taken as $.402^2$ (the mean for the 153 procedures) rather than the median, $.313^2$.

[Table 3 about here]

The values in Table 3 were computed assuming a typical coefficient of variation for the procedures we have been studying. This number may be different for a particular procedure or in other settings. Since the average CVA for DRGs (rather than procedures) was .35 [3], similar to that for procedures, this rule of thumb may well hold both for DRGs and for procedures.

The CVA is a simple measure of variation that can be used to assess the relative amount of variation for various DRGs and procedures. More complex analyses must be used [13, 14, 15] to examine possible explanations of the small-area variation.

In summary, we recommend that the CVAS (and its confidence interval) be used when small-area variation is being compared for several procedures or DRGs, as long as the minimum expected number of procedures is 1 or higher. It appears that a CVA or CVAS greater than about .5

can be considered as "large". We also recommend hypothesis testing to see if there is more variation than expected by chance alone when the minimum expected number of procedures is 10 or lower.

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; pp. 177-206. DHHS Publication No. (PHS) 90-3454, U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. 1990.

2.Roos N, Wennberg J, McPherson K. Using diagnosis-related groups for studying variations in hospital admissions. Health Care Financing Review 9(4):53-62. 1988.

3.Diehr P, Cain K, Ye Zhan, Abdul-Salam F. Small-area variation analysis: methods for comparing several DRGs. Medical Care 31:YS45-YS53. Supplement. 1993.

4.Diehr P, Cain K, Kreuter W, Rosenkranz S. Can small-area analysis detect variation in surgery rates? The power of small-area variation analysis.   Medical Care 1992; 30:484-502.

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

6.Cain KC, Diehr P, Ye Z. The multiple admission factor (MAF) in small area variation analysis. Department of Biostatistics, University of Washington Technical Report No. 116, 1992.

7.U.S. Department of Health and Human Services. Special report. Hospital data by geographic area for aged Medicare beneficiaries: Volume 2.   Selected procedures, 1986. Health Care Financing Administration. HCFA Pub. 03301. June 1990.

8.Gittelsohn A, Halpern J, Sanchez R. Income, race, and surgery in Maryland. American Journal of Public Health 81:1435-1441. 1991.

9.Diehr P, Cain K, Connell F, Volinn E. What is too much variation? the null hypothesis in small area analysis. Health Serv Res 1990; 24:741-771.

10.Chambers, Cleveland, Kleiner, Tukey. Graphical methods for data analysis. Duxbury press; North Scituate, Massachusetts. 1983.

11.McPherson K, Wennberg J, Hovind O, Clifford P: Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. N Engl J Med 1982; 307:1310-1314.

12.Cain K, Diehr P. Testing the null hypothesis in small area analysis. Health Services Research 1992; 27:267-294.

13   Wolfe R, Petroni G, McLaughlin C, et al. Empirical evaluation of statistical models for counts or rates. Stat Med 1991; 10:1405.
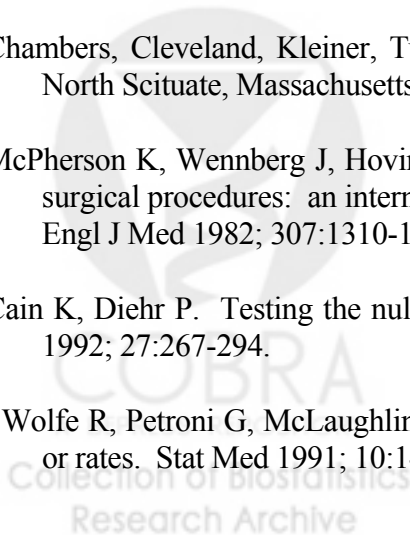
14 Gatsonis C, Normand SL, Liu Chuanhai, Morris C. Geographic variation of procedure utilization: a hierarchical model approach. Medical care 1993;31:YS54-YS59.

15 Shwartz M, Ash AS, Anderson J, Iezzoni L, Payne S, Restuccia J. Small area variations in hospitalization rates: how much you see depends on how you look. Medical Care 1994;32:189-201.

Definitions of Combined Procedure Categories

```
PROC  PNAME                         ICD-9 CODES

A     HCFA PTCA                     36.00-36.09
B     HCFA CARDIAC CATH             37.21-37.23
C     HCFA INGUINAL HERN REPAIR     53.00-53.17
D     HCFA PROSTATECTOMY            60.20-60.69
E     HCFA HYSTERECTOMY             68.40-68.70
F     HCFA REDUCT FX OF FEMUR       79.05, 79.15, 79.25, 79.35
G     CAROTID ENDARTERECTOMY        38.12
H     PART EXCIS LARGE INTESTINE    45.70-45.79
I     CHOLECYSTECTOMY               51.20-51.22
J     TOTAL KNEE REPLACEMENT        81.41
K     TOTAL HIP REPLACEMENT         81.50, 81.51, 81.59
L     OTHER ARTHROPLASTY OF HIP     81.61-81.69
M     SINUS PROCEDURES              22.00-22.99
N     TONSILLECTOMY                 28.2-28.39
O     HEMORRHOIDECTOMY              49.40-49.49
P     APPENDECTOMY                  47.0
Q     GASTRECTOMY                   43.50-43.99
R     AMPUTATION OF LIMB            84.00-84.19
S     MASTECTOMY                    85.41-85.48
T     THYROIDECTOMY                  6.20- 6.79
```

Parameters

μ =   grand mean (prevalence)=# of admissions/population

σ^2_A true variance among counties

Sample Statistics

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

$Y_{i.}$ Sample Mean for County i

Y_u Unweighted Sample Mean = $\Sigma Y_{i.}/k$

$Y_{..}$ Grand Mean = Weighted Sample Mean = $\Sigma n_i Y_{i.}/\Sigma n_i$

S^2_u Unweighted Sample Variance = $\Sigma(Y_{i.}-Y_u)^2/(k-1)$

S^2_w Weighted Sample Variance = $\Sigma[n_i(Y_{i.}-Y_{..})^2]/(\Sigma n_i-1)$

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,n and s^2 are the mean and variance of the number of people per area, and $n_o=n-s^2_n/(k*n)$.

$\hat{\mu}$ Weighted Sample Mean = $\Sigma\Sigma Y_{ij}/\Sigma n_i = \Sigma n_{iY_{i.}} / \Sigma n_i$

CVA   Anova Estimate of "true" coefficient of variation = $\hat{\sigma}_A / \hat{\mu}$

CVU   Unweighted Coefficient of Variation = S_u/Y_u

CVW   Weighted Coefficient of Variation = $S_w/ \hat{\mu}$

RSCV Square root of the Systematic Component of Variance [1,11] or
    $[(1/k) [3((O_i-E_i)^2)/E_i{}^2 -3(1/E_i)]^{.5}$

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

χ^2_{MH} Same as χ^2, but expected values are calculated separately for each stratum and summed for each county.

CVAS CVA for age/sex-standardized prevalences, using χ^2_{MH} in Equations 1 and 2.

Appendix C

Relationship of Chi-square to CVA

A non-central chi-square with N df and non-centrality parameter L has mean M = N+L and variance V = 2(N+2L).

The normalized χ^2 statistic $(\chi^2-M)/V^{**}.5$, approaches normality as L approaches infinity for N fixed. (This is probably the case, since we are interested in "large" chi-squares".)

We can then estimate:
N = k-1   (number of counties - 1).
M = χ^2 (let observed χ^2 = estimate of mean).
Then
L = M-N = χ^2 -k+1.
and
$$V \quad = 2(N+2L) = 2(N+2(M-N)) = 2(N+2m-2N) = 2(2M-N)$$
$$= 2(2\chi^2 - k+1)$$
$$= 4\chi^2 -2(k-1).$$

From equation {1} in the main text,

$$CVA^2 = a + b\,\chi^2$$

where

$$a \; = \; -(k-1)*MAF/[\mu*(k-1)*n_o]$$
and
$$b = \; 1/[\mu*(k-1)*n_o] \; (b \text{ is positive})$$

The variance of CVA^2 is $b^2\,V$

$$=\{\,4\chi^2 -2(k-1)\}/[\mu*(k-1)*n_o]^2$$

so the 95% confidence interval is

$$CVA^2 \pm 1.96\{\,4\chi^2 -2(k-1)\}^{.5}/[\mu*(k-1)*n_o].$$
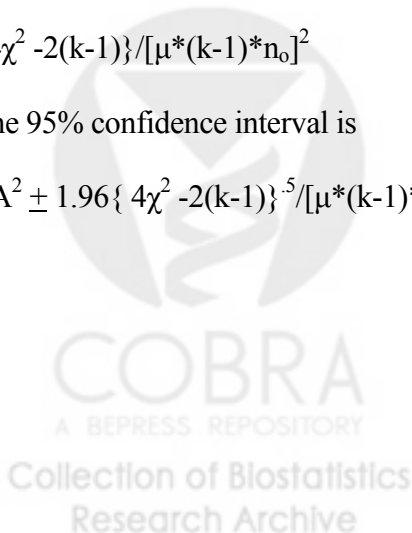
Table 1

Descriptive Statistics for Different CV Estimates

```
Number of Valid Observations (Listwise) =        153.00

Variable        Mean        Std Dev   Minimum   Maximum      N  Label

CVU             .65         .34        .28       2.57      153  unweighted CV
CVW             .45         .26        .18       1.91      153  weighted CV
RSCV            .56         .35        .00       2.53      153  root SCV
CVA             .44         .29        .06       2.03      153  CVA
CVAS            .40         .31        .00       2.29      153  standardized CVA
```

Correlation with: $\hat{\mu}$ $\hat{\sigma}_A$

```
  CVU            -.1762        .3452
                 P= .029    P= .000


  CVW            -.1392        .3821
                 P= .086    P= .000


  RSCV           -.0948        .2927
                 P= .244    P= .000


  CVA            -.0714        .4291
                 P= .381    P= .000


  CVAS           -.0292        .4944
                 P= .721    P= .000
```

Table 2

**Prevalence and Standardized CVA, With Confidence Intervals
For ICD-9 Surgical Procedures with Minimum Expected Number > 5**

PROCEDURE	$\hat{\mu}$	MIN	MAX	LOWER	CVAS	UPPER
7534 FETAL MONITORING NOS	88.1	.0	749.4	1.98	2.01	2.04
7532 FETAL EKG	58.7	.0	275.4	1.37	1.41	1.45
3723 RT/LEFT HEART CARD CATH	73.8	6.1	235.9	.83	.87	.91
7309 ARTIF RUPT MEMBRANES NEC	233.0	7.1	705.9	.73	.75	.77
7359 MANUAL ASSIST DELIV NEC	157.5	5.0	501.8	.73	.75	.78
736 EPISIOTOMY	334.4	8.9	842.1	.55	.57	.59
6632 BILAT TUBAL DIVISION NEC	62.0	8.3	187.6	.54	.59	.63
721 LOW FORCEPS W EPISIOTOMY	87.7	6.6	182.4	.50	.54	.57
3722 LEFT HEART CARDIAC CATH	139.6	13.4	278.9	.44	.47	.49
734 MEDICAL INDUCTION LABOR	90.1	18.7	162.5	.44	.47	.50
4524 LG BOWEL ENDOSCOPY NEC	60.8	16.6	170.3	.34	.38	.42
5732 CYSTOSCOPY NEC	124.5	30.4	296.7	.29	.33	.35
640 CIRCUMCISION	249.3	45.8	452.2	.29	.31	.33
3601 PTCA-NO THROMBOLYSIS	56.2	2.4	109.1	.28	.33	.37
685 VAGINAL HYSTERECTOMY	61.6	11.4	138.5	.28	.32	.36
8051 IV DISC EXCISION	95.7	28.2	178.0	.27	.30	.33
A HCFA PTCA	63.7	2.4	120.0	.26	.31	.35
3961 EXTRACORPOREAL CIRCULAT	55.9	3.7	108.1	.25	.30	.34
6561 REMOVE BOTH TUBES OVAR	108.3	26.7	197.3	.25	.28	.31
B HCFA CARDIAC CATH	226.3	23.1	384.1	.25	.27	.29
7569 REPAIR OB LACERATION NEC	221.9	24.8	353.6	.24	.26	.28
4513 SM BOWEL ENDOSCOPY NEC	136.5	11.2	276.4	.23	.26	.29
545 PERITONEAL ADHESIOLYSIS	82.1	11.4	144.1	.23	.27	.30
0331 SPINAL TAP	82.7	24.8	193.3	.22	.26	.29
5122 TOTAL CHOLECYSTECTOMY (I)	162.2	57.2	287.9	.21	.24	.26
741 LOW CERVICAL C-SECTION	281.7	38.2	493.7	.20	.22	.24
E HCFA HYSTERECTOMY	230.3	45.8	363.9	.19	.21	.23
684 TOTAL ABD HYSTERECTOMY	167.3	34.3	266.9	.19	.21	.24
470 APPENDECTOMY (P)	96.3	15.3	170.9	.18	.22	.25
8622 WOUND DEBRIDEMENT	73.5	22.9	130.8	.16	.20	.24
D HCFA PROSTATECTOMY	155.2	42.0	402.1	.15	.17	.20
602 TRANSURETHRAL PROSTATECT	136.9	34.3	377.3	.13	.16	.19
3893 VENOUS CATHETER NEC	56.0	19.9	88.0	.12	.18	.22
H PART EXCIS LARGE INTESTINE	65.4	15.3	199.3	.09	.14	.18
K TOTAL HIP REPLACEMENT	58.9	19.1	139.5	.09	.15	.19
7935 OPEN REDUC-INT FIX FEMUR	51.4	11.4	91.9	.08	.14	.19
F HCFA REDUCT FX OF FEMUR	63.5	15.3	109.6	.05	.12	.16

Table 3

Percent of Observed Variation Typically Due to Random Variation*
As a function of the True CV, MAF, and E_{min}

16 True CV	.313	.313	.402	.402
MAF	1	4	1	4
E_{min}				
1	26.72-91.07%	59.31-97.60%	18.09-86.08%	46.92-96.11%
5	6.80-67.12	22.57-89.09	4.23-55.31	15.02-83.19
10	3.51-50.51	12.72-80.32	2.16-38.22	8.12-71.22
25	1.44-28.99	5.51-62.02	.88-19.84	3.42-49.75
50	0.72-16.95	2.83-44.95	.44-11.01	1.73-33.11
100	.36- 9.26	1.43-28.99	.22- 5.82	.87-19.84
1000	.03- 1.01	.14- 3.92	.02- .61	.08- 2.41

* Percent of observed variation due to "sampling error" for a procedure with a typical value of the CVAS.

Lower limit = ((MAF/(28E_{min})/(True CV^2 + MAF/(28E_{min}))*100%
Upper limit = ((MAF / E_{min})/(True CV^2 + MAF/E_{min}))*100%

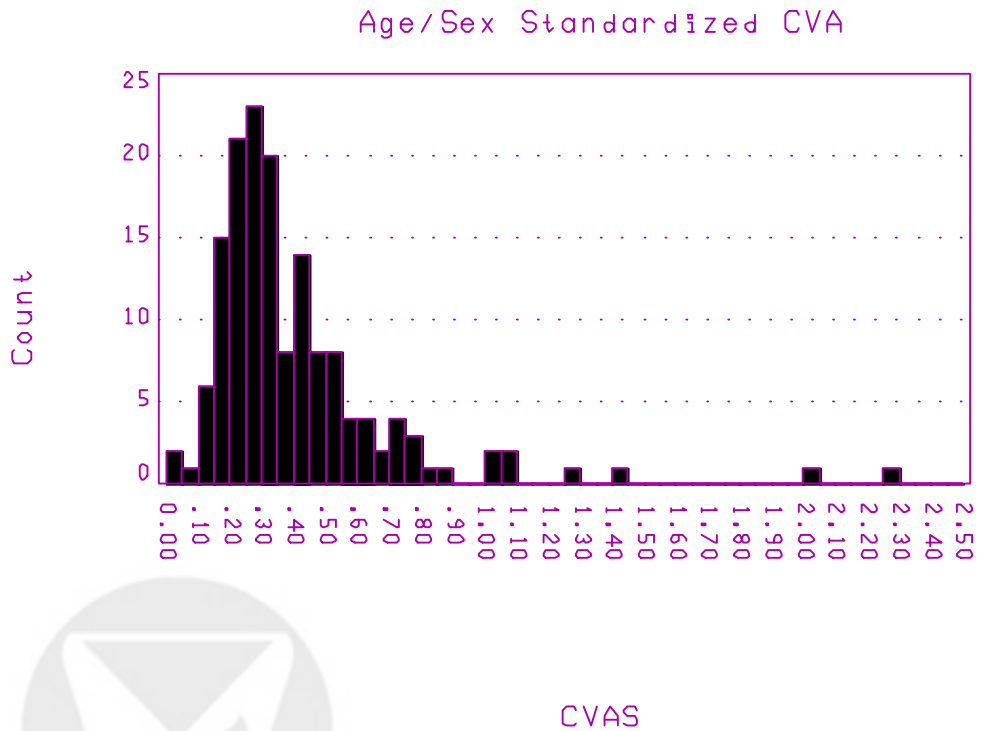**Figure 1** Distribution of CVAS for 153 Surgical Procedures

Age/Sex Standardized CVA

CVAS

23