A Small Area Simulation Approach to Determining Excess Variation in Dental Procedure Rates

PAULA DIEHR, PHD, AND DAVID GREMBOWSKI, PHD

Abstract: All small area analyses need to compare the observed variability in rates to that expected by chance alone, but the expected variability is usually not known. This paper uses patient-level data for five dental procedures to simulate the distributions of the summary statistics that are usually generated in such studies. These statistics are found to vary greatly even under the "null hypothesis" that all dentists are using procedures at the same rates. The simulated dentist rates are compared to observed rates obtained in a different study. These findings illustrate problems that can occur in small area analysis studies, and emphasize the importance of using statistical techniques that are appropriate for the data that are to be analyzed. Investigators should make every effort to obtain patient-level data, or at least to understand the underlying distribution of the number of procedures per patient, to avoid mistaking significant deviations from an incorrect model as evidence for significant variation among small areas. (*Am J Public Health* 1990; 80:1343–1348.)

Introduction

Small area analysis is a popular methodology in health services research. A typical study might calculate the utilization rate for a service in several small areas, compare the largest rate to the smallest, note that the difference is large, and attempt to explain the high variability as a function of service availability, physician uncertainty, and other variables of interest. The statistical methods used in such studies have been questioned by an author of this paper.[1]

A recent paper using computer simulation to examine the statistical properties of the most frequently used descriptive statistics[2] demonstrated that all of the usual descriptive statistics could be deceptive, showing large apparent variation when there was no more variation than would be expected by chance alone. A simple 2 × k chi-square test (classifying people in each of k communities into two cells, by whether they had or did not have the procedure) was an appropriate test for excess variability as long as each person could have the procedure at most once. However, the chi-square test could be deceptive; results would be "statistically significant" too often if an individual could have the procedure more than once. Simulation studies were recommended for these situations; this requires that the distribution of the number of procedures per person be known. Finally, it was pointed out that studies very similar to small area analyses were being conducted in other situations: e.g., a "small area" might be a hospital or a dental practice. This paper applies the small area simulation approach to a set of dental data, to determine if there is more variation in procedure rates among dentists than would be expected by chance alone.

Grembowski, *et al*,[3] calculated procedure rates for 200 dentists in general practice in four urban counties of Washington State from 1984 and 1985 dental claims of members of the Washington Education Association (WEA) and their dependents. The number of patients seen at least once in the two-year period ranged from 75 to 300 for the 200 dentists, as shown in Table 1.

Procedure rates for each dentist were calculated by dividing the total number of procedures performed in a two-year period by the number of people seen at least once by the dentist during that period. Results were age-adjusted. Table 2 shows the substantial variability that was found among dentists: rates of oral examinations ranged from 10 per 1,000 patients to 3,543 per 1,000 patients. The coefficient of variation (the standard deviation divided by the mean) was 0.289. This seems like substantial variation. However, a good deal of variation can be expected by chance alone, even if the underlying rate of examinations per person is the same in each practice.[2] If "exams" were a binary variable, which a person could have once or never, it would be appropriate to use a 2 × 200 chi-square test to examine the null hypothesis that the probability of having an oral exam was the same in all practices.[2] However, it is obvious that many people have more than one examination in two years, which means that the chi-square test and the usual small area statistics are either inappropriate or have unknown distributions under the null hypothesis.

If the number of oral exams for each person were known, techniques could be developed to study whether there was excess variation. The data could be recoded so that each

TABLE 1—Number of Washington Education Association Patients Seen by each Dentist in Two Years

N of Patients Seen	N of Dentists
75–99	94
100–124	54
125–149	19
150–199	24
200–300	9

TABLE 2—Observed Procedure Rates in 200 Dental Practices (number of procedures per 1,000 patients)

Category	Mean	SD	Min	Max	EQ	CV
Oral Exams	1617	467	10*	3543*	354.3*	.289*
Fillings	889	373	188*	2379	12.65*	.420*
Extractions	119	83	0	529	—	.698*
Root Canals	44	40	0	228	—	.909*
Scalings	23	47	0	345*	—	2.044*

EQ = Maximum/Minimum.
CV = coefficient of variation.
*Minimum lower than 5th percentile of null distribution, or maximum higher than 95th percentile (see Tables 3 and 4).

Address reprint requests to Paula Diehr, PhD, Professor, Department of Biostatistics, SC-32, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195. Dr. Grembowski is Assistant Professor of Health Services and Dental Public Health Sciences at the University of Washington. This paper, submitted to the Journal September 12, 1989, was revised and accepted for publication March 8, 1990.

person was coded 1 if he had one or more exams, and 0 otherwise, and the chi-square statistic could be used; or, one could assume that the number of exams per person was approximately normally distributed, and perform an analysis of variance to study whether the average number of exams per person were different among the dentists. Unfortunately, the only data available for the Washington study were the rates per dentist of the various procedures. No patient level data were available. For these reasons, we attempted a simulation approach, discussed elsewhere in detail.[2] The goal is to determine what the distribution of procedures among the 200 practices would have been under the null hypothesis that no dentists were "high users" but, on the contrary, that all dentists performed the same number of procedures per person, on average. The simulation method attempts to determine how much variation among practices would be expected by chance alone. This is then used as a benchmark to determine whether the observed variation in Table 2 is important.

Methods

Data

A different data source was used to obtain information about the distribution of procedures at the patient level. Annual dental utilization data were obtained for 12,897 insureds from a random sample of adults and children covered by Pennsylvania Blue Shield dental insurance in 1980. These data have been analyzed elsewhere.[4,5] The 5,912 people with no dental visits in 1980 were eliminated from the data set. Appendix 1 contains frequency distributions for five dental procedures—oral exams, fillings, simple extractions, root canals, and full-mouth periodontal scalings. (For example, 2,764 patients had no oral exams, 2,717 had one exam, and six people had 12 oral exams in one year.)

Fitting the Pennsylvania Data

We first attempted to fit the Pennsylvania data to the Poisson and negative binomial distributions.[6] If a known distribution was appropriate, we could use it to generate random data for the simulation. Clearly, none of the data are normally distributed because of the long tails and the high number of observations clustered at zero. The Poisson distribution is not likely to fit the data because a Poisson process assumes that procedures occur independently, which is probably not the case for dental procedures. The negative binomial distribution assumes that each person uses procedures with a Poisson distribution, but that each person has a different Poisson parameter, distributed according to a gamma distribution.[6] This distribution was considered because it has longer tails than the Poisson distribution.

We estimated the parameters of the Poisson and negative binomial distributions using the maximum likelihood procedure.[6] This yielded expected numbers which were compared to the observed numbers with the chi-square goodness of fit test.[7] Neither distribution provided an acceptable fit for the number of oral exams or for fillings. The negative binomial distribution did provide an adequate model for the number of extractions, root canals, and scalings.

The following simulation work did not assume any mathematical distribution, because there was none that worked for all the data. Rather, data were generated directly from the empirical distributions, as described below. Future work in simulating procedure rates for extractions, root canals, and scalings might use the negative binomial distribution to generate random data, however.

Simulation of Utilization Rates

Our goal was to develop a simulation model to assess the amount of variability that would have occurred among the dental practices by chance alone if the null hypothesis (that all dentists use procedures at the same underlying rate) is true. We generated 100 different simulated sets of procedure rates, in which the null hypothesis was true.

For each of the 200 dentists, a uniform random number between 0 and 1 was generated for each patient in that practice (e.g., for a dentist with 75 patients, 75 random numbers were generated). Each of these numbers was compared to the cumulative distribution in Appendix 1. For example, 2764/6985 or 39.5 percent of the people had no oral exam; 78.46 percent had 0 or 1 exam; 97.7 percent had 0, 1, or 2 exams, etc. For each random number, if it was less than .395, the patient was assigned 0 exams; if between .395 and .7846, the patient was assigned one exam, etc. Thus, on average, the number of exams per patient match the distribution of Appendix 1, but the distributions were somewhat different for each dentist.

After the number of exams were generated for all 75 patients in the first practice, the exams were summed, multiplied by 2 (see discussion section) and divided by 75 to provide a simulated two-year rate of oral exams for that dentist. This was then continued for all 200 dentists, which generated 200 different simulated examination rates. Next, the minimum rate, the maximum rate, the extremal quotient (EQ, ratio of the maximum rate to the minimum), the coefficient of variation, and other statistics explained in the results section were calculated. This process was repeated 100 times to give, for example, the "average" maximum rate, or the 95th percentile of the maximum rate, under the null hypothesis that each practice has the underlying distribution of examinations shown in Appendix 1.

Results

Mean Procedure Rates

Table 2 shows the observed results, and Table 3 shows the simulated results. The first column in Tables 2 and 3 is the estimated mean number of procedures per 1,000 patients. Recalling that the two sets of numbers arise from different data sets, there is fairly good correspondence on the rate of oral exams (1,725 versus 1,617), root canals (55 versus 44), and full mouth scaling (34 versus 23). There is very poor agreement on fillings (2,123 versus 889) and for simple extractions (211 versus 119). This may be due to differences in billing practices between Pennsylvania and Washington Dental Service—the carrier for Washington Educational Association. There are probably also socioeconomic differences between the Washington and the Pennsylvania patients. These differences are considered further below.

TABLE 3—Simulated Procedure Rates (number of procedures per 1,000 patients)

	Rate Mean	Minimum Mean	Minimum 5%	Maximum Mean	Maximum 95%	EQ Mean	EQ 95%
Oral Exams	1725	1260	1136	2304	2537	1.8	2.1
Fillings	2123	1195	974	3422	3897	2.9	3.4
Extractions	211	30	0	529	667	17.5	30.1
Root Canals	55	0	0	259	342	—	—
Scalings	34	0	0	227	316	—	—

EQ = Maximum/Minimum
Based on 100 iterations per procedure.

Maximum and Minimum

The simulated minimum and maximum are shown in Table 3. For example, on average, the lowest rate of oral exams was 1,260 per 1,000 patients, and the highest rate was 2,304 per 1,000, even though the underlying rate for all dentists was 1,725. Thus, an observed minimum rate near 1,260 or a maximum rate near 2,300 would not be an indication of excess variability among practices. Table 3 also shows the simulated 5th percentile of the minimum and 95th percentile of the maximum. That is, 95 percent of all minimum values are above 1,136, and 95 percent of all maximum values are below 2,537.

The observed minimum rates for oral exams, shown in Table 2, can be thought of as extremely low, since 10 is substantially below 1,136. The same is true for fillings. The minimum for the other procedures is 0 for both the observed and the simulated data. Comparing the maximum values of Table 2 to the 95th percentiles in Table 3, the maximum rates for oral exams and for full mouth scaling are higher than would be expected by chance alone, but the other maxima are not particularly high.

Extremal Quotient (EQ)

Table 3 shows the simulated average value and the 95th percentile of the EQ (maximum divided by minimum). For oral exams, one would expect an EQ of about 1.8, and values above 2.1 (the 95th percentile) would be considered particularly high. For fillings, a ratio of 2.9 is expected, and 3.4 would be considered high. For extractions the minimum value was 0 in 10 percent of the simulations, and the EQ could not be computed. For the remaining 90 percent, the expected EQ was 17.5 and the 95th percentile of the EQ was 30.7. The EQ is infinite for root canals and scalings, as the minimum is always zero. In Table 2, the observed EQ could be computed only for oral examinations and fillings. Both of the observed EQs are much larger than the 95th percentile of the EQ distribution, indicating that the null hypothesis that all dentists had the same rates is probably not true. The variability in the rates is due in part to some dentists having much *lower* use of procedures than would be expected.

Coefficient of Variation (CV)

Table 4 shows that the 95th percentile of the simulated CV (standard deviation over mean) for oral exams is 0.115, as compared to an observed value of 0.289 shown in Table 2. The observed variation is thus higher than would have been expected by chance alone. All of the observed CVs are considerably above what would have been expected. Therefore, the null hypothesis probably is not true, and there are some dentists who use procedures at different rates from the others.

TABLE 4—Simulation Results—Other Summary Statistics*

	95th percentiles (per 1,000 patients)				
	CV	SCV	199 df Chi sq	1 df Chi sq	F
Oral Exams	.115	8	−441	7.28	1.22
Fillings	.194	35	−1116	27.31	1.13
Extractions	.442	148	1070	13.15	1.17
Root Canals	.914	674	1035	14.35	1.19
Scalings	1.188	1468	1208	18.97	1.20
Degrees of Freedom			199	1	199, inf
Tabled 95th Percentile			233	3.84	1.17

*CV is the coefficient of variation and SCV is an estimate of the systematic component of the variation, defined in the text. Chi-square with 199 degrees of freedom is 2 × k chi-square assuming that each procedure is a person. Chi-square with 1 df is the test for a particular practice (see text). F is the F statistic that arises from a 1-way analysis of variance. Tabled 95th percentiles are taken from tables of the chi-square and F distributions.

Based on 100 iterations per procedure.

Systematic Component of Variation (SCV)

Table 4 shows some other statistics commonly used in small area analysis, which were not calculated for the Washington data. They are included for completeness. The SCV has been proposed by McPherson and Wennberg as a test for excess variability among small areas.[8] The formula is:

$$SCV = (1/k)\ [\Sigma((O_i - E_i)^2)/E_i^2 - \Sigma(1/E_i)]^*\ 1000.$$

There is an F-test associated with this statistic, but it is not often used. The SCV was derived under the assumption of Poisson rates, which is not appropriate here. The data in Table 4 illustrate further what has been shown elsewhere[2]; that is, that the SCV can take on very large values, even when there is no underlying variation among the small areas. In addition, it becomes larger as the mean rate becomes smaller. Thus, it would be unwise to use the SCV alone as a descriptive statistic in this situation.

Chi-Square

Two chi-square tests have been used in small area analyses. One test is a 2 × 200 chi-square test, which assumes that each person received a procedure at most once, calculates the number not receiving the procedure as "population size minus the number of procedures," and creates a 2 × 200 table. This is clearly inappropriate in our situation, since many people had more than one procedure. In fact, the 95th percentile of this chi-square statistic is negative for oral exams and fillings, because the expected numbers of people who did not have the procedure are negative for some dentists (for example, there are more fillings than patients). A chi-square value greater than 1,000 would be needed to demonstrate "significance" for the last three procedures, as compared to the tabled 95th percentile of the chi-square with 199 degrees of freedom, which is 233.

Another test which is proposed in these situations is a 1 degree of freedom chi-square test of the hypothesis that a particular dentist has a rate significantly different from what would have been expected. The expected number of procedures would be the grand rate (e.g., 1,725 oral exams per 1,000) times the number of patients (e.g., 75, for an expected rate of 129 exams). If the observed rate were, e.g., 155, the chi-square statistic would be $X = (O - E)^2/E = (155 - 129)^2/129 = 5.2$. This is much higher than the usual critical value of 3.84 for the chi-square distribution with 1 degree of freedom. However, Table 4 shows that the 95th percentile of chi-square statistics in this situation is 7.28, meaning that 5.2 is not a particularly large value. The 95th percentiles for the other dental procedures are even larger, ranging from 13 to 27. Thus, the simple chi-square test, to look for outliers, could be extremely deceptive in this situation, indicating significant variation when a practice was operating, on average, like all other practices. Or, put another way, 16 percent of the chi-square statistics would have been greater than 3.84 for oral examinations, giving a Type I error of 16 percent rather than the commonly accepted 5 percent level. The other Type I errors are 47 percent for fillings, 30 percent for extractions, 28 percent for root canals, and 19 percent for scalings. If these data were being used to identify outlying dentists, far more than 5 percent would be incorrectly labeled as over-providers. This discrepancy occurs because the data violate the essential property required for the chi-square test:

that the counts be independent (or, that the distribution of counts per person follow the Poisson distribution).

There is another problem with this chi-square test, which is that it ignores multiple comparisons. If a chi-square test, even using the correct 95th percentile, were performed for all 200 dentists, about 5 percent (10 practices) would be judged to have excessive utilization rates by chance alone. The probability that no dentist was classified incorrectly is $.95^{200}$ = .00004. Thus, the experiment-wise Type I error is almost 1.0. This might be adjusted for by using an alpha level of .05/200 = .00025 instead of alpha = .05. This would require comparing the observed chi-square values to the 99.975th percentile of the distribution (about 17.17 instead of 3.84 in this situation).

Clearly, the combination of these two problems (that the chi-square statistic does not follow the chi-square distribution, and that there are multiple comparisons) will tend to yield extremely high observed chi-square values even in the null case. Users who are not aware of this problem risk labeling small areas as outliers that are not extreme at all. We have seen examples in which a large percent, or even *all* of the small areas being considered, were labeled as outliers based on this test![9]

Analysis of Variance (F-test)

The 95th percentile of the F distribution is given in Table 4. This statistic would have resulted if we had performed an analysis of variance to detect a difference among the 200 dentists in the mean number of procedures per person. Analysis of variance requires having patient-level data, so that the variation among people can be computed. This was not available in the Washington data, but could be computed from the Pennsylvania distributions. The 95th percentile for an F distribution with 199 and infinite degrees of freedom is 1.17 if the null hypothesis is true. This is close to the simulated F values for the dental procedures (1.22, 1.13, 1.17, 1.19, 1.20), showing that analysis of variance would have been appropriate if patient-level data had been available, even though the data in Appendix 1 are far from normally distributed. (We used only 100 iterations, so the percentiles might be even closer than is shown.) If the data are not available at the individual level, to permit analysis of variance, it will be necessary to perform a simulation study to determine whether there is more variability than would be expected by chance alone.

Adjusting the Pennsylvania Rates

It is unfortunate that some of the average rates were so different in the Washington and Pennsylvania data sets. This suggests that some of the "significant" variation detected could have been caused not by excess variation but because the Washington data were being compared to the wrong underlying distribution. To examine this possibility we standardized the results, so that the average simulated rate would be the same as the average observed rate. Specifically, we multiplied the simulated rates by a constant factor (e.g., for oral exams 1617/1725 = 0.937 so that the simulated mean would be 1617). This has the effect of multiplying most of the values in Tables 3 and 4 by a factor less than one (0.937 for oral exams). (The EQ, CV and F values would be unchanged). The observed data would then be as extreme or even more extreme, if compared to these adjusted tables. All of the observed maxima would be larger than the adjusted 95th percentile except for root canals.

Discussion

The intent of this paper was to demonstrate that small area analysis issues were similar to those in detecting excess variation in other settings, and in particular to illustrate the simulation approach. Both the findings and the methods merit some discussion.

If we assume that the underlying distribution of procedures is about the same in the Washington and Pennsylvania data, the simulation method has demonstrated that there was excess variability for oral examinations and full mouth scaling. The other procedures also showed excess variability, although not as unequivocally.

There are some shortcomings to these results, and also to the methods that we used to obtain them. The Washington population was a fairly homogeneous group of middle-class patients. The Pennsylvania Blue Shield data were from a random sample of all insureds, who varied by social class. The differences between the two study populations may explain why the mean rates for the rehabilitative services in Table 3 were higher than those in Table 2. Also, Pennsylvania had better coverage for fillings than did the Washington contract, which may account for the discrepancy in filling rates between Tables 2 and 3.

The estimated rates from one year of Pennsylvania data were multiplied by 2 to give two-year rates for comparison with the Washington data. This is not quite appropriate, as it assumes that non-users in the first year would not have used any services in the second year. The correct multiplier is between 1.08 and 2.0 (details from authors). The "2" was chosen as an upper bound, to provide as much variability as possible. If 1.08 had been used, all of the observed maxima would have been above the 95th percentile.

No effort was made in either data set to remove people who were not covered for the entire period of time. If a person was insured for only one month of the two-year period, the rate was calculated as though the person had been at risk for the entire 24 months. This would tend to lower the rates, and would lower them more for the Washington data than for the Pennsylvania data, since the time period was longer for Washington. This may explain some of the differences in the two data sets.

We did not model one aspect of the variability—that a different number of patients might have been seen in different two-year periods. Age and sex differences were not incorporated in the simulation, as the Washington data had already been age standardized, and sex is not an important determinant of use in insured populations.[4,5] We showed elsewhere[2] that this was a reasonable way to proceed. Finally, the estimated means and percentiles are not completely accurate, as there were only 100 simulations per procedure.

Once excess variability has been established, it is appropriate to examine reasons for this variability. Grembowski, *et al.*,[10] have shown that structural features of the practice, such as its age and size, explain some of the variation in the rates.

Although not completely satisfactory, the simulation approach has provided support for the finding of excess variability among dentists. Utilization review programs operated by dental insurers are often implemented to identify dentists who have high procedure rates. The insurer may, for example, intensify review of the dentist's claims as a means of reducing the dentist's rate. Our results indicate that insurers need to determine whether excess variability exists for a given procedure before a dentist is labeled as providing "too many" services. Of equal importance, if excess variability is detected, dentists with the lowest rates may be underutilizing appropriate services.[3] Therefore, insurers should target utilization review at both ends of the rate distribution.

The methodology of this study is applicable in other situations. We used the simulation method to examine the characteristics of small area analysis statistics applied to dental practices instead of geographic areas. Dental practices have smaller numbers of patients in each "small area" and some of the procedures are applied multiple times to the same person. The distributions of root canals or scalings, which are rare, are more similar to the distribution of the number of hospital admissions for a particular diagnosis, which is often considered in small area analyses. The main results of this study have to do with the underlying distribution of services, the properties of several descriptive statistics, and various hypothesis testing procedures.

None of the distributions of services in Appendix 1 had a Poisson distribution. This is important because several approaches are now being used in the small area literature which assume that the underlying distribution is Poisson: the 1 df chi-square test and the SCV were derived under these assumptions[8] and some regression approaches model deviations from the Poisson distribution.[11,12] In all of these methods, statistically significant departures from the Poisson distribution are taken as evidence that there is unexplained variability among the small areas. However, as shown here and elsewhere,[2] the underlying distribution at the person level may not have a Poisson distribution, and finding "significant" variation is as likely to be caused by this fact as to represent excess variation among the small areas. Unless there is evidence that the underlying distribution is Poisson (e.g., that a person cannot have more than one of the procedures of interest) this is an inappropriate inference.

Tables 3 and 4 demonstrate that a considerable amount of variability can be expected among small areas by chance alone, and that it is dangerous to "eyeball" descriptive statistics as a test for excess variability. The minimum and maximum rates, and the extremal quotient, could vary considerably under the null hypothesis. The coefficient of variation and the systematic coefficient of variation also vary substantially, and tend to be larger for the less prevalent procedures (i.e., higher for root canals and scaling than for oral exams and fillings). It is common to compare CVs or SCVs across several procedures, and to claim that those with the largest coefficients represent "practitioner uncertainty." If we had done this with the data of Table 4, we would have "found" that there is less certainty about root canals and scaling than about oral exams and fillings, even though under the simulation model all dentists used procedures at exactly the same underlying rates! It is inappropriate to infer differences in "certainty" based only on CVs or SCVs if procedures have different mean rates.

Several tests of the hypothesis that there is underlying variation among the small areas have been developed. One is a 2 × k chi-square test which is appropriate if a person can have the procedure of interest at most once (e.g., death, removal of an organ). This procedure failed for the data of this study, even for procedures such as root canals, where the assumption that few people had more than one per year would have seemed "reasonable." Thus, the assumptions of this test must be examined with care.

The second chi-square test, which purports to test whether an individual dentist or small area is an "outlier" (observed significantly different from expected) was also shown to be inappropriate first, because the data do not meet the underlying assumption of a Poisson distribution, and second, because the problems of multiple comparisons are ignored. It is very important that this test be used correctly, since it singles out particular areas or dentists as outliers, and may have deleterious effects.

A third test which is sometimes proposed is a simple analysis of variance. The simulation results of Table 4 show that it would have been appropriate for these data, if data had been available at the patient level. (Such data may be difficult to obtain, as there must be a unique patient identifier to permit ascribing multiple procedures to a person.) If investigation of patient level data showed that they were inappropriate for analysis of variance, logarithmic or other transformations could be attempted.

The simulation method we have proposed has so far been used only in the null situation. The power of the various procedures to detect true underlying variation has not been studied; it should be studied. It is possible that tests based on the CV, for example, would be more powerful than other types of tests. More methods research is needed in this area.

The major lesson of this paper is that it is important in small area analysis to understand the underlying distribution of services, at the patient level, to avoid mistaking misspecification of the underlying model for statistically significant excess variability among small areas.

DISCLAIMER

Interpretations of the data are the authors' own and do not necessarily represent those of the National Center for Health Services Research and Health Care Technology Assessment, Washington Dental Service, or the Washington Education Association.

APPENDIX 1

Number of Procedures per Person in Pennsylvania Data*

Number of Procedures	Procedure: Exams	Fillings	Root Canals	Full Mouth Scaling	Extractions
0.	2764	4020	6839	3722	6446
1.	2717	1239	120	43	417
2.	1349	775	16	5	88
3.	92	382	5		15
4.	39	206	3		6
5.	4	117			3
6.	14	93	1	2	10
7.		49			
8.		28	1		
9.		20			
10.		18			
11.		4			
12.	6	15			
13.		2			
14.		4			
16.		4			
18.		2			
19.		1			
21.		1			
22.		1			
24.		1			
28.		1			
29.		1			
36.		1			

*This is a sample of 12,897 people enrolled in Pennsylvania Blue Shield in 1980. Of these, 5,912 had no dental procedures in this period. These were removed from this table to match the WEA data more closely. (WEA data included only people who used one or more procedures in two years). Thus, the distributions represent the utilization of 6,985 people who had at least one dental procedure in one year. Only 3,772 of the patients had coverage for full-mouth scalings.

REFERENCES

1. Diehr P: Small area statistics: Large statistical problems. Am J Public Health 1984; 74:313–314.
2. Diehr P, Cain K, Connell F, Volinn E: What is too much variation? The null hypothesis in small area analysis. Health Serv Res February 1990; 24:741–771.
3. Grembowski D, Milgrom P, Fiset L. Variation in dentist service rates in a homogeneous patient population. J Public Health Dent. Volume 50, Summer 1990, in press.
4. Grembowski D, Conrad D, Milgrom P: Dental care demand among children with dental insurance. Health Serv Res 1987; 21:755–775.
5. Conrad D, Grembowski D, Milgrom P: Dental care demand: Insurance effects and plan design. Health Serv Res 1987; 22:341–367.
6. Johnson N, Kotz S: Discrete Distributions. Boston: Houghton Mifflin, 1969.
7. Dixon W, Massey F: Introduction to Statistical Analysis. New York: McGraw-Hill, 1957.
8. 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.
9. Wennberg JE: 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.
10. Grembowski D, Milgrom P, Fiset L: Factors influencing variations in dentist source rates. J Public Health Dent. Volume 50, Summer 1990, in press.
11. Wolfe R, Griffith J, McMahon L, Tedeschi P, Petroni G, McLaughlin C: Patterns of surgical and nonsurgical hospital use in Michigan communities from 1980 through 1984. Health Serv Res 1989; 24:67–82.
12. McMahon L, Wolfe R, Tedeschi P: Variation in hospital admissions among small areas: A comparison of Maine and Michigan. Med care 1989; 27:623–631.