The Extremal Quotient in Small-Area Variation Analysis

Vahé A. Kazandjian, Paul W. Durance, and M. Anthony Schork

This article reviews the current small-area variation analysis (SAVA) approach to population-based rates of surgery, and describes a new method for ascertaining variance based on the beta-binomial probability distribution of small-area rates. The critical review of the current SAVA approach focuses (1) on how incidence rates are calculated, and (2) on how the significance of the observed magnitude between the largest and smallest rates (i.e., the extremal quotient) is ascertained. While reducing the problems of calculating rates by considering only certain operative procedures, the new method addresses the current inadequacies of ascertaining significant differences among small areas. Not only does it correctly assess likelihood of an extremal quotient, it also can determine the particular area's rate, producing an unlikely extremal quotient. The method evaluates the probability that the observed magnitude of the extremal quotient is due solely to chance and study design effects, and tables of these probabilities are available for the method's application. A mathematical model, based on a combination of the binomial and beta distributions, uses (1) the sample size, (2) the average of the areas' rates, (3) the variance among the rates, and (4) a specific quotient level to determine the probability of observing the quotient by chance. After computerizing this calculation, probability tables for reasonable values of these four parameters are generated. In addition to looking at just one quotient for each sample, the probability tables facilitate the easy examination of intermediate quotients when the extremal quotient is unlikely due to chance. By alternatively ignoring the highest and lowest rates, two new quotients can be produced and tested. Given that one of these two quotients

---

Address correspondence and requests for reprints to Vahé A. Kazandjian, M.P.H., Director of Research, The Maryland Hospital Association, 1301 York Road, Suite 800, Lutherville, MD 21093-6087. He is a doctoral candidate, Medical Care Management and Policy Department, School of Public Health, University of Michigan, Ann Arbor. Paul W. Durance, Ph.D. is Adjunct Assistant Professor, Department of Health Services Management and Policy, University of Michigan, Ann Arbor. M. Anthony Schork, Ph.D. is Professor, Department of Biostatistics, The University of Michigan, Ann Arbor.

is likely due to chance, the excluded rate (i.e., producing the unlikely extremal quotient) can be classified as an outlier, and the associated small area should be the focus of more detailed investigation. The probability tables reveal that the extremal quotient is not the appropriate statistic to be applied in studies where many small areas are to be included. The probability of seeing even a "large" extremal quotient simply by chance rapidly approaches one as the sample size increases. However, an extremal quotient modeled from a beta-binomial distribution can be useful for studies with small sample sizes (e.g., six counties). The use of this beta-binomial model for small-area rates provides a new method of designing and evaluating small-area studies where costs or domain limit the number of areas under consideration. The availability and correct evaluation of small-area studies are critical to policy decision making and practice pattern investigation within government and the health care industry. By making possible SAVA of small sample size, this new method promotes a fuller understanding of health care delivery.

Variations in hospital utilization rates between geographic areas have been consistently reported in the literature. These variations have been documented internationally (CaGeorge and Roos 1984; Kazandjian et al. 1989; Lewis 1969; Roos and Roos 1983; Roos et al. 1978; Roos 1984; Roos and Roos 1977, 1982; Roos and Gilbert 1975; Roos et al. 1977; Stockwell 1977); nationally (Caper 1986; Lembcke 1952; Wennberg 1984; Wennberg et al. 1977; Wennberg and Gittelsohn 1982; Wennberg, Bunker, and Barnes 1980; Wilson and Tedeschi 1984; Wilson et al. 1985; Griffith et al. 1985; Chassin et al. 1986; Gornick 1977; Knickman and Foltz 1984; McCarthy and Finkel 1980; Kuder 1985; Brook et al. 1984; Connell et al. 1984); and between the United States and other countries (McPherson et al. 1981, 1982). Virtually every investigator who has published analyses of areal variations in population-based hospital admission and utilization rates for selected surgical procedures has reported "substantial" differences in these use rates among communities.

These findings were interpreted as evidence of a reducible component in the cost of health care, at least insofar as the communities with high utilization rates are concerned. That is, it has been argued that if some communities are able to restrict their rates of hospitalization without adversely affecting the health of their population, then other communities should be able to do the same, assuming hazards to health are comparable across populations (Wennberg 1984, 1985).

The majority of small-area studies have primarily focused on determining and explaining the variation in surgical rates. The reason is twofold: (1) on the average, the per capita cost of surgery is higher

than that of nonsurgical treatments and (2) the occurrence of a surgical procedure is relatively easier to discern than a medical treatment. If a substantial proportion of these surgeries is "unnecessary" or could be done on an outpatient basis, cost savings could be achieved presumably without affecting the quality of care. Individual hospitals, as well as statewide agencies (e.g., Peer Review Organizations), are manifesting increasing interest in assessing and understanding the areal variations in the use of surgery (McCracken and Bognanni 1986).

Major planning and utilization control decisions have already been made in a number of states based on the findings of small-area analysis by McCracken and Bognanni (1986). Thus, studies of small-area variation in hospital use rates have policy implications for individual hospitals as well as hospital service areas regulated by some external review agency. There are indications, however, that the findings of these studies may not always have been appropriately tested for statistical significance—presenting the danger of making policy decisions based on potentially inappropriate research findings.

PURPOSE

This article critically examines the most commonly used method of statistical significance testing used in small-area studies; presents a statistical method of significance testing that affects not only the interpretation of the findings, but the evaluation of the study design; and discusses the policy implications of the proposed method.

THE METHODOLOGY OF SMALL-AREA VARIATION ANALYSIS (SAVA)

Primarily descriptive, the current SAVA technique generally focuses on an area's population-based surgical hospital utilization rate. The numerator of the rate is the number of residents of that area who received care from one or more hospitals in or out of their area of residence during a certain period of time, and the denominator of the rate is the population of the geographic area under study. The use rates for a number of geopolitical areas (e.g., counties) or service areas (e.g., hospital service areas) are calculated and compared. Differences in the magnitude of those rates are contrasted within the study sample and, often, to an external standard (e.g., a regional average).

The outcome of this exercise is to identify the rates that are higher

than the average, or to assess the magnitude of the difference between the highest and the lowest rates in the study sample. For example, consider a SAVA to ascertain and describe the differences in tonsillectomy rates across ten counties of state X. The tonsillectomy rate of each county is calculated first. The numerator of this rate is the number of tonsillectomies performed for the population of a county regardless of the hospital location (i.e., in or out of the county of residence). The denominator of the rate is the size of the population in that county. For some procedures the denominator could be assessed more precisely than for others. In the case of tonsillectomy, for example, a reasonable denominator is the number of persons in that county who are in the cohort ages 1 to 18 years and who still have tonsils. Similar adjustments by sex are made for some procedures (e.g., the prostatectomy rate has only the number of males in the 45+ years age group as the denominator). Such age and sex adjustments are necessary for reliable comparisons of population-based rates across and between different small geographic areas.

The comparison of these rates is usually completed in one of two ways—either to an average or to the magnitude of the difference in extreme rates. First, the rates could be compared to the regional average, such as the state average rate of tonsillectomy. The distribution of the ten county rates around the state average provides a quick visual description of the magnitude of the dispersion and allows the identification of "high"-rate counties—those above the state average.

The second commonly used method of comparing the small-area rates is to describe the magnitude of the difference in these rates between the highest and the lowest. For example, a fivefold variation in tonsillectomy rates will be reported if the lowest rate was 1 per 1,000 population in county A and the highest was 5 per 1,000 population in county B. This method seems to capture the attention of the reader, policymaker, and researcher more readily than when a distribution around an average is reported.

The reasons for the observed variations are frequently explored from administrative and policy perspectives (e.g., how the availability of resources influences the rate of their use), although an increasing number of researchers are proposing a more clinical and epidemiological interpretation for the variations in health services utilization rates (e.g., differences in the severity of illness or differences in physicians' diagnostic and therapeutic management preferences).

Finally, based on the actual or sometimes speculative reasons for these variations, policy decisions have been effected to ameliorate the process and, maybe, the outcome of the health care system (e.g., to

bring the "high"-use areas' rates to a predetermined standard, to identify physicians who may be using some hospital services "unnecessarily," to review and monitor the adoption and use of resources and technology through state regulatory and statutory changes, or to enhance and maintain the quality of care within a well-managed budget).

Given the increasing popularity of the SAVA method and the substantial implications its results may have for cost containment, strategic planning, resource allocation, and quality of care, identifying and assessing potential limitations of the current SAVA methodology seem appropriate.

LIMITATIONS OF THE SAVA METHODOLOGY

There are a number of methodological problems that generally have not received adequate attention in studies of variation in utilization rates using small-area analysis. These problems can be summarized as (1) the data problem and (2) the methodological problem.

THE DATA PROBLEM

The data problem stems from the source of data used to compute the hospitalization rate, especially of nonsurgical cases in a given population. Since small-area analyses often use large retrospective discharge abstract data bases, both the numerator and the denominator used to calculate the incidence rate could be misleading. For example, if a patient has been admitted to a hospital more than once for the same complaint, and if the hospital episodes constitute the numerator of the rate, the frequency of that treatment within the area population will be overestimated. One way to control the overestimation problem is to study only procedures that are usually performed once on a patient (e.g., hysterectomy).

The denominator may present another problem. The incidence rate, as used in epidemiology, is calculated by dividing the frequency of an event (e.g., a disease) by the estimated population at risk. In small-area analysis the rate is expressed as the frequency of an event (e.g., a specific surgical procedure) to the size of the population in that area. Population data are usually obtained from national or regional censuses that are often not for the same year for which the incidence is measured. Census data usually lag behind, from two to ten years—

hence the need for using demographic projections. More importantly, population enumerations include all people in that area, not just those at risk for a specific surgery. This problem can be reduced by limiting the population to age/sex-specific groups appropriate to the specific site under study.

In addressing the data problem, difficulties associated with the numerator will tend to be minimized when the latter consists only of cases for whom that procedure is performed once. For the remainder of this work, it is assumed that only this type of procedure would be understudied.

THE METHODOLOGICAL PROBLEM

The methodological problem arises in the use of the ratio of two extreme rates to calculate the magnitude of the variation. Small-area studies have reported that the per capita utilization for certain surgical procedures shows a wide variation across geographic areas. Usually, this dispersion is presented in the form of a ratio obtained by dividing the highest rate by the lowest rate. Frequently, a manifold variation in the rate of surgery is demonstrated by using only two of many observations within a sample.

These ratios of extremes, called extremal quotients, have statistical properties that need to be considered prior to their use as indicators of variation in surgical rates:

1. Extremal quotients are highly skewed and unstable statistics.
2. The size of extremal quotients is a function of the number of observation units (areas) from which they are calculated.

Consequently, it is important to determine if the magnitude of the extremal quotient is unlikely to have been a chance occurrence (i.e., the observed magnitude of dispersion between the highest and lowest rate is considered significant). At least two statisticians have raised this issue (Diehr 1984; Willemain 1982).

THE PROPOSED METHOD OF USING EXTREMAL QUOTIENTS

Willemain (1982) suggested that if extremal quotients are to be used to describe the magnitude of the difference in the rate of surgery between the high and low utilization areas, the effects of chance on the size of

the variation should be known to the researchers. He specifically points out that "ratios of extremes will accentuate the chance effects and thereby create an exaggerated impression of the underlying diversity" (p. 92). The many factors that affect the magnitude of the difference are classified herein into three categories:

1. *Real effects*. Interpatient and intersurgeon differences in the perception of the need for care by type and amount
2. *Chance effects*. Differences in rates expected through random variation alone
3. *Sample delineation effects*. The number and size of population for the small areas in the sample.

The proposed method seeks to model chance and delineation effects in such a way as to allow the assessment of the probability of real effects (i.e., real differences in area utilization) being present in a given sample; if the probability of realizing a particular quotient level due to chance and delineation effects alone is low (e.g., $< .05$), then real effects can be assumed to be present. Significance in this method is expressed, therefore, as the probability of chance and delineation effects resulting in a particular quotient.

While effects are those things that actually produce a particular realization, parameters are used to model the effects. In the proposed method, chance and delineation effects are modeled by four parameters: (1) the rate of utilization, (2) the variability in utilization rates within a sample of several areas, (3) the number of areas in the sample, and (4) the particular extremal quotient of interest (e.g., realized quotient). Although it would be entirely appropriate to rely on normative or historical variability, currently only the sample variability is available. This is analogous to the use of the t-statistic for a normal distribution when the variance is derived from a sample. The use of the sample variance might increase the coefficient of variation, making it more difficult to detect a situation containing a real effect. When a real effect is present, the sample variability generally overstates the "true" (i.e., chance-effect) variability. Therefore, any real effect that is detected would likely also have been detected using the "true" variability.

For the remainder of this work, the average rate of utilization provides the point estimator for the random variable P, the per capita probability of having surgery during a year. A binomial probability model when using this point estimator assumes that every individual has the same probability of surgery, that is, no variability in P among individuals. To account for individual differences, Willemain (1982)

suggests augmenting the binomial model with a beta distribution for the probability of any individual undergoing the surgical procedure (i.e., the random variable P). Since the shape of the beta distribution is quite flexible, it has adequately represented all distributions of P seen thus far. With a beta-binomial model, differences in the likelihood of surgery among an area's individuals can be more realistically considered when modeling area rates.

The mathematical derivations from the model are shown in the appendix, while assumptions for the model's application are listed here:

1. N_i for all areas is large, thereby markedly reducing the significance of the binomial variation in the rate; that is, $(p(1 - p)/N_i)$ becomes quite small.
2. A patient can only have the procedure once during the year or the original binomial model does not hold and dependency would result in naturally larger extremal quotients.
3. All rates must be less than one since rates are ultimately used as probabilities in the beta distribution (e.g., the rate 50/1,000 is the probability of .05).

Assumption 1 suggests that a small-area population of at least 1,000 would produce a binomial variance in the rate of surgery one million times smaller, that is, $(1/N_i)^2$, than the variance in the number of procedures in an area, that is, $N_i\, p(1 - p)$. This reduction in binomial variance to the sixth order of magnitude essentially leaves only beta variance in the surgical rate. Assumption 3 is usually met in light of assumption 2.

The beta-binomial model has been automated to calculate chance-effect probability from a closed-form equation (as opposed to a simulation). The program accepts estimates of P from .0005 to .05; variance in P, expressed as coefficients of variation (CV), of .150 to .900; number of areas (A) ranging from 2 to 40; and the quotient (Q) itself, which can range from 1.5 to 30.

RESULTS

Outcomes from this program for P of .005 and .05, along with selected values of the other three parameters, are found in Tables 1 and 2, respectively. A more extensive tabulation of the probabilities for a wide variety of common values of all four parameters has been generated by

the authors and is available from Schork (see Kazandjian, Durance, and Schork 1987). As one would expect, the probability of exceeding a given quotient level increases as the variability in the area rates (i.e., coefficient of variation) increases. Also, as the number of areas in the cohort increases, so does the probability of exceeding a given quotient level. On the other hand, the increase (one hundredfold, from 5 to 500 per 10,000) in the estimator of P makes very little difference in the probability of chance effects (see the difference between Tables 1 and 2). As a result, it is the variability and the number of areas that essentially determine the probability.

These results are not linear across the entire selected ranges, but they are monotonic. The effects on the probability due to changes in variability (CV) or number of small areas is greatest when the chance-effect probability is close to .5. Since significance is detected only at low probabilities (e.g., .01 to .05) where there is relatively little effect from CV and the number of areas, the presence of regions affected by these parameters (CV and number of areas) does not have an operational effect on most applications of the tables. In other words, linear interpolation is reasonable for determining the probability associated with parameter values not explicitly found in the tables when the chance-effect probability is close to .05. For example, if P is .005, CV is .25, and the number of areas is 11, using Table 1 to test the significance of a quotient of 3.0 produces a probability of approximately .082 [(.070 + .094)/2].

Use of the beta-binomial probability model to predict the probability due to the chance and sample-delineation effects is quite interesting. First, the greater the variability among the use rates in a sample is, the larger the ratio of extremes must be to become significant. Basically, this means that if the values in a number of observations (e.g., use rate sample) are comparable, a smaller quotient among them has a higher likelihood of being due to a real effect than the same quotient applied to observations that are quite dissimilar (e.g., larger variation between the rates of the sample).

Second, the probability becomes larger for a value of the ratio of the extremes as the number of small areas increases (i.e., sample-delineation effect). This means that by increasing the number of small areas, the likelihood of including an area with a very different utilization rate (high or low) is also increased. The result may be a large extremal quotient that could be due to chance alone if the number of small areas is also large.

Table 1: Beta-Binomial Chance-Effect Probabilities for Utilization Rate of 5 per 1,000 (P = .005)

*Extremal Quotient Level**	*Number of Areas* 2	4	6	10	12	14	16	18	20	22	24	26	30	35	40
						Coefficient of Variation = .175									
1.5	.094	.339	.554	.818	.888	.931	.958	.974	.984	.990	.993	.995	.997	.997	.997
2.0	.006	.030	.065	.152	.199	.246	.294	.340	.384	.427	.468	.507	.578	.655	.720
2.5	.000	.001	.004	.011	.015	.020	.026	.031	.038	.044	.051	.059	.074	.094	.115
3.0	.000	.000	.000	.000	.001	.001	.001	.002	.002	.003	.003	.004	.005	.007	.009
						Coefficient of Variation = .200									
1.5	.146	.470	.701	.915	.956	.977	.987	.993	.995	.997	.997	.998	.998	.998	.997
2.0	.015	.072	.145	.305	.382	.453	.519	.579	.632	.680	.722	.759	.820	.876	.915
2.5	.001	.007	.017	.045	.061	.079	.097	.117	.137	.158	.179	.200	.242	.295	.346
3.0	.000	.000	.002	.005	.007	.010	.013	.016	.019	.022	.026	.030	.038	.049	.061
4.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.001
						Coefficient of Variation = .250									
2.0	.053	.211	.376	.638	.730	.800	.853	.893	.922	.943	.959	.970	.984	.992	.995
2.5	.011	.053	.109	.238	.303	.366	.425	.481	.533	.581	.625	.665	.734	.801	.853
3.0	.002	.013	.028	.070	.094	.119	.146	.173	.201	.229	.256	.284	.338	.402	.463
4.0	.000	.000	.002	.005	.008	.010	.013	.016	.019	.023	.026	.030	.038	.049	.061
5.0	.000	.000	.000	.000	.000	.001	.001	.001	.002	.002	.002	.003	.004	.005	.007

						Coefficient of Variation = .350									
2.5	.075	.277	.468	.734	.817	.874	.915	.942	.961	.973	.982	.987	.993	.996	.997
3.0	.034	.142	.265	.492	.585	.663	.729	.782	.826	.862	.890	.913	.945	.969	.982
4.0	.008	.039	.080	.179	.231	.282	.333	.382	.428	.473	.515	.554	.625	.699	.761
5.0	.002	.012	.026	.063	.084	.106	.130	.154	.178	.203	.228	.253	.302	.361	.418
6.0	.000	.004	.009	.023	.032	.041	.051	.061	.072	.084	.095	.107	.132	.164	.196
8.0	.000	.000	.001	.004	.005	.007	.009	.011	.013	.016	.018	.021	.026	.034	.042
						Coefficient of Variation = .450									
3.0	.097	.335	.542	.801	.872	.918	.948	.967	.978	.986	.990	.993	.996	.997	.997
4.0	.038	.155	.285	.516	.609	.687	.750	.802	.843	.877	.903	.924	.953	.974	.985
5.0	.017	.076	.149	.304	.379	.448	.512	.570	.623	.670	.711	.748	.809	.866	.906
6.0	.008	.040	.081	.178	.229	.279	.328	.376	.421	.465	.506	.544	.614	.689	.750
8.0	.002	.013	.028	.065	.087	.109	.133	.157	.181	.205	.230	.254	.302	.360	.415
10.0	.001	.005	.011	.027	.037	.047	.058	.069	.081	.093	.105	.118	.144	.177	.211
15.0	.000	.000	.002	.004	.006	.008	.010	.013	.015	.018	.020	.023	.029	.036	.044
						Coefficient of Variation = .500									
4.0	.066	.243	.417	.679	.767	.832	.879	.914	.938	.956	.968	.977	.987	.993	.995
5.0	.034	.139	.258	.476	.567	.645	.710	.765	.809	.846	.876	.900	.935	.962	.977
6.0	.020	.084	.162	.326	.402	.473	.538	.596	.648	.694	.735	.770	.829	.882	.918
8.0	.007	.035	.071	.156	.200	.245	.290	.333	.375	.416	.455	.492	.560	.634	.697
10.0	.003	.017	.035	.080	.105	.131	.158	.186	.213	.241	.268	.295	.347	.410	.468
15.0	.000	.004	.008	.021	.028	.035	.043	.052	.061	.070	.079	.089	.109	.134	.160
20.0	.000	.001	.003	.007	.010	.013	.016	.019	.022	.026	.029	.033	.041	.052	.063

*The ratio of the highest rate to the lowest rate.

Table 2: Beta-Binomial Chance-Effect Probabilities for Utilization Rate of 50 per 1,000 (P = .05)

*Extremal Quotient Level**	*Number of Areas*														
	2	*4*	*6*	*10*	*12*	*14*	*16*	*18*	*20*	*22*	*24*	*26*	*30*	*35*	*40*
							Coefficient of Variation = .175								
1.5	.091	.332	.545	.811	.881	.926	.955	.972	.982	.989	.992	.995	.997	.997	.997
2.0	.005	.029	.062	.145	.191	.237	.282	.327	.371	.413	.453	.492	.562	.639	.705
2.5	.000	.001	.003	.010	.014	.019	.024	.029	.035	.041	.048	.054	.069	.088	.107
3.0	.000	.000	.000	.000	.000	.001	.001	.001	.002	.002	.003	.003	.005	.006	.008
							Coefficient of Variation = .200								
1.5	.144	.466	.697	.913	.955	.976	.987	.993	.995	.997	.997	.998	.998	.998	.998
2.0	.015	.071	.143	.301	.378	.449	.514	.574	.627	.675	.717	.755	.816	.873	.912
2.5	.001	.007	.017	.044	.060	.077	.096	.115	.135	.155	.176	.197	.238	.290	.341
3.0	.000	.000	.002	.005	.007	.010	.012	.015	.019	.022	.026	.029	.037	.048	.060
4.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.001
							Coefficient of Variation = .250								
2.0	.056	.220	.388	.651	.743	.812	.863	.901	.928	.948	.963	.973	.986	.993	.996
2.5	.012	.057	.117	.252	.319	.384	.445	.502	.554	.602	.646	.685	.753	.818	.867
3.0	.002	.014	.031	.077	.103	.131	.159	.188	.218	.247	.276	.305	.361	.428	.490
4.0	.000	.001	.002	.006	.009	.012	.015	.019	.023	.027	.031	.036	.045	.058	.071
5.0	.000	.000	.000	.000	.001	.001	.001	.002	.002	.003	.003	.004	.005	.007	.008
6.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.001

						Coefficient of Variation = .350									
2.5	.069	.258	.440	.707	.792	.854	.898	.929	.951	.966	.976	.983	.991	.996	.997
3.0	.030	.128	.241	.457	.548	.626	.693	.749	.796	.834	.866	.891	.929	.959	.975
4.0	.007	.033	.069	.156	.202	.249	.295	.340	.384	.426	.466	.504	.573	.649	.713
5.0	.002	.010	.021	.052	.070	.089	.109	.129	.150	.172	.193	.215	.258	.311	.363
6.0	.000	.003	.007	.019	.025	.033	.041	.049	.058	.068	.077	.087	.107	.134	.161
8.0	.000	.000	.001	.003	.004	.005	.007	.008	.010	.012	.014	.016	.020	.026	.032
						Coefficient of Variation = .450									
3.0	.089	.315	.515	.777	.852	.903	.937	.959	.973	.982	.988	.992	.995	.997	.997
4.0	.034	.141	.260	.481	.573	.651	.716	.770	.815	.851	.880	.904	.939	.965	.979
5.0	.015	.067	.132	.274	.343	.409	.470	.527	.578	.625	.668	.706	.771	.833	.879
6.0	.007	.034	.071	.156	.201	.246	.291	.335	.377	.418	.457	.494	.562	.637	.700
8.0	.002	.011	.023	.055	.073	.092	.112	.133	.153	.175	.196	.217	.259	.311	.362
10.0	.000	.004	.009	.022	.030	.038	.047	.057	.066	.076	.087	.097	.119	.147	.175
15.0	.000	.000	.001	.003	.005	.006	.008	.010	.012	.014	.016	.018	.022	.028	.035
						Coefficient of Variation = .500									
4.0	.060	.226	.391	.649	.738	.807	.858	.896	.924	.945	.960	.970	.984	.991	.995
5.0	.031	.127	.236	.442	.531	.608	.674	.730	.777	.816	.849	.876	.916	.948	.967
6.0	.017	.075	.146	.295	.367	.434	.497	.554	.605	.652	.693	.730	.792	.851	.893
8.0	.007	.030	.062	.136	.176	.216	.256	.296	.335	.372	.408	.443	.509	.581	.645
10.0	.003	.014	.030	.068	.090	.113	.136	.160	.184	.208	.232	.256	.303	.360	.414
15.0	.000	.003	.007	.017	.023	.029	.036	.043	.050	.058	.066	.074	.090	.111	.133
20.0	.000	.001	.002	.006	.008	.010	.013	.015	.018	.021	.024	.027	.033	.042	.050

*The ratio of the highest rate to the lowest rate.

TAKING THE USE OF BETA-BINOMIAL METHOD ONE STEP FURTHER

Willemain (1982) concludes that a "large" value of the ratio of extremes may not necessarily indicate significantly different values for the "high" and "low" rates of surgery. However, given the nature of the three effects on which the probability model was built—underlying diversity, chance effects, and the number of small areas—the beta-binomial model can be used as a study design tool in addition to an evaluative tool. The implications of such flexibility are numerous.

First, when the probabilities associated with the different combinations of the four parameters are calculated and produced in probability tables, researchers can use these tables to assist in formulating the study design. For example, if a researcher has calculated the population-based rates of surgery for 20 small areas, and found a mean rate of 6.2 operations per 1,000 population and a coefficient of variation of .220, the magnitude of the extremal quotient needed to achieve selected statistical significance can be ascertained. A table presenting these probabilities requires values for the mean rate of surgery, the CV associated with these rates, and the number of small areas in the sample.

Second, and perhaps even more challenging, is the effort to design a SAVA study by minimizing the potential for artifactual findings. If the purpose of a study is exclusively exploratory and descriptive (e.g., to document the population-based rates of surgery in each geographic small area), there may be no need for any analytic methods. As mentioned at the outset, some administrative and policy decisions may have been taken based on statistically nonsignificant areal differences in rates. Therefore, if the purpose of a study is to explain the reasons why variation exists, it is of paramount importance to first assess correctly the significance of the value of the extremal differences in rates. That is, it is inappropriate to explain the differences in the rates of surgery when those rates are in fact not significantly different.

Third, a SAVA study need not be limited to just one ratio of extremes in deciding on the magnitude of the variation, but a SAVA study may also be helpful in assessing the difference in rates found between extreme values of any set or subset of rates. Although the entire set of rates is used to calculate the CV, the general approach yields the probability associated with only the two extreme values, not the distribution probability of all rates. In addition to assessing the probability for the two extreme values of the rates, the characteristics of

the distribution of all rates (i.e., including those intermediate rates) in the study sample may be analyzed for outliers by means of a proposed "Pair-Wise Iterative Analysis" (PWIA). The original set of rates can be investigated to assess that subset that possesses the maximum number of areas having a quotient most likely due to chance effects alone (i.e., no real effects).

Consider a sample of ten areas, where the age- and sex-adjusted rates/1,000 population for surgery X are:

(1) 2.91	(6) 5.00
(2) 4.74	(7) 5.02
(3) 4.85	(8) 5.20
(4) 4.89	(9) 5.26
(5) 4.98	(10) 7.15.

The average rate (i.e., probability of surgery) is 5/1,000 with a sample coefficient of variation of .20. The PWIA steps are:

Step 1. Calculate the mean rate (as a probability of surgery X) and the CV for the rates in this sample.

Step 2. Calculate the first quotient level as 7.15/2.91 = 2.5-fold difference.

Step 3. Assess the probability that this quotient level is significant from the beta-binomial probability tables (i.e., not due to chance) based on delineation characteristics. If it is significant, continue with step 4. If not, stop (i.e., no real effects in this set (or subset) are suspected).

Step 4a. Take the "next to the highest" rate together with the lowest rate, reduce the number of areas by one, and repeat steps 2 and 3 (e.g., 5.26/2.91 = 1.8-fold difference; find the probability of occurrence of a 1.8, given the other characteristics of the sample);

And

Step 4b. Take the "next to the lowest" rate along with the highest rate, reduce the number of areas by one, and repeat steps 2 and 3 (e.g., 7.15/4.74 = 1.5-fold difference). Then if either or both of steps 4a and 4b show significance, continue with step 4c.

Step 4c. Take the "next to the highest" *and* the "next to the lowest" rates, reduce the number of areas by two, and repeat steps 2 and 3 (e.g., 5.26/4.74 = 1.1-fold difference).

The advantage of the PWIA over the exclusive use of the two extreme values in a set of observations (rates) lies in consideration of the distributional characteristics of the rates. In addition to the highest and lowest observations, the intermediate rates are contrasted, and the magnitude of these additional differences is compared for real effect. If by dropping a particular rate(s) [i.e., area(s)] the probability of real effects is reduced to an insignificant level, the area(s) for further focused investigation has (have) been identified. In other words, if the full set of areas possesses a quotient attributable to chance alone, no areas are designated as worthy of inferential investigation. However, when a subset of areas possesses a quotient due to chance, an area or areas worthy of further investigation are identified by exclusion from this subset.

DISCUSSION

This article presents (1) one of the frequently used statistical measures, and use of extremal quotients in small-area variation analysis (SAVA), (2) a discussion of some of the limitations of the statistic, and (3) based on a previous work by Willemain, an extension of the extremal-quotient method—the Pair-Wise Iterative (Steps) Analysis (PWIA). The purpose of this extended method is to make the use of extremal-quotient statistics more acceptable in SAVA.

Among the applications of SAVA results is their use in policy decisions for an individual hospital, a multihospital system, a state regulatory agency, or a federal agency. Adequate interpretation of the results is critical. In an environment of regulation, competition, and other market mechanisms, partial and inadequate interpretation and use of SAVA results may have serious effects on care-providing institutions and providers. For administrators and other decision makers, the appeal of using extremal quotients in SAVA lies in their potential for documenting and distinguishing significant differences in utilization rates.

The beta-binomial chance-effect probability methodology and sample output tables presented provide investigators with an enhanced method of assessing the significance of the observed extremal-quotient magnitude. Also, the evaluators and researchers studying geographic differences in use rates will find this statistical approach to be valuable during the designing phase of their studies. The tables illustrate particularly the utility of assessing the effects of alternative sample sizes.

Finally, the PWIA method presents an interesting possibility for

more extensive and appropriate use of the extremal quotient in SAVA. PWIA is advantageous in analyzing more than just two observations out of a distribution (a statistically more acceptable approach). Further, by identifying what might be called the "significantly high rate areas," attention and effort are focused not only on the area with the highest rate, but also on a number of areas. The potential inclusion of several high-rate areas as candidates for further inquiry has important ramifications for policy decisions.

APPENDIX

Binomial Distribution:

For the rate expressed as a positive fraction (incidence/population in area i)

$$p(R_i) = \begin{bmatrix} N_i \\ k \end{bmatrix} * P^k * (1 - P)^{N_i - k}$$

for $k = 0,1,2,\ldots, N_i$,
$0 < P < 1$, and
N_i is a positive integer

Note: $p(R_i = k/N_i) = p(x = k)$
$E(R_i) = P$
$Var(R_i) = P * (1 - P)/N_i$

Beta Distribution

$$f(p)\ (S + F + 1)!/(S! * F!) * p^S * (1 - p)^F$$
for $0 < p < 1$
$S > -1$
$F > -1$
$E(P) = (S + 1)/(S + F + 2)$
$Var(P) = [(S + 1) * (F + 1)]/[(S + F + 2)^2 * (S + F + 3)]$

While in the general form of the distribution S and F may take on some negative values, in this work positive values in both are necessary to produce desirable unimodal distributions of reasonable mean and variability. As can be seen from the distribution, when S and F are zero, for example, a uniform distribution results. S and F fully determine the shape of the distribution of both beta and beta-binomial when N is large. It is assumed that the variance observed in area rates is

based on the underlying variability in the probability of P (i.e., the probability of one person undergoing the procedure). The mean of the study sample's rates is taken as the estimate of P, and the sample variance is taken as the estimate of the $Var(P)$. These estimates allow the calculation of the beta distribution parameters S and F. The fit of this model to empirical data from 12 published studies has been shown to be excellent (Willemain 1982).

Once S and F (i.e., the distribution of P) have been set, the model requires two additional parameters to determine the probability of a quotient level: (1) the value of a desired or observed quotient (Q) itself, and (2) the number of areas (A) in the cohort.

A *PL/1* program that inputs the mean rate and the variability of a cohort, the quotient level, and the number of areas in the cohort was written to numerically integrate the joint distribution formed by the ratio of two beta-binomials. Conceptually, a particular quotient would be realized only when the rate of one area (R_i) (i.e., the largest) is at least Q times larger than the smallest (R_j) (i.e., $Q = R_i/R_j$). The program accumulates the probability of one area having at most a rate of R_j while another simultaneously has a rate from at least R_i up to $1/Q$. Assessment of this total requires the double integration of the distribution from 0 to $1/Q$ and from R_i to 1. Under no circumstances can a rate exceed 1. As a practical matter, most rates are closer to .01.

REFERENCES

Brook, R., et al. "Geographic Variations in the Use of Services: Do They Have Any Clinical Significance?" *Health Affairs* 3 (Summer 1984): 63-73.

CaGeorge, S. M., and L. L. Roos, Jr. "When Surgical Rates Change: Workload and Turnover in Manitoba, 1974-1978." *Medical Care* 22 (October 1984):890-900.

Caper, P., et al. "Tracking Physician Practice Patterns for Quality Care." *Business and Health* 3 (September 1986):7-9.

Chassin, M. R., et al. "Variations in the Use of Medical and Surgical Services by the Medicare Population." *New England Journal of Medicine* 314 (January 1986):285-90.

Connell, A. F., et al. "Clinical Correlates of Small Area Variations in Population-Based Admission Rate for Diabetes." *Medical Care* 22 (October 1984):939-49.

Diehr, P. "Small Area Statistics: Large Statistical Problems." *American Journal of Public Health* 74 (April 1984):313-14.

Gornick, M. "Medicare Patients: Geographic Differences in Hospital Discharge Rates and Multiple Stays." *Social Security Bulletin* 40 (June 1977): 22-41.

Griffith, J. R., et al. "Clinical Profiles of Hospital Discharge Rates in Local Communities." *Health Services Research* 20 (June 1985):131-52.

Kazandjian, V. A., P. Durance, and M. A. Schork. "Determining Significance of the Extremal Quotient in Small Area Variation Analysis with a Beta-Binomial Probability Model." Technical Report. Department of Biostatistics, University of Michigan, 1987.

Kazandjian, V. A., P. E. Dans, and L. Scherlis. "What Physicians Should Know About Small Area Variation Analysis." *Maryland Medical Journal* 38 (June 1989):477-81.

Knickman, J. R., and A. M. Foltz. "Regional Differences in Hospital Utilization." *Medical Care* 22 (November 1984):971-86.

Kuder, J. M. "Small Area Variation in Iowa Hospital Utilization." *Iowa Medicine* 75 (May 1985):213-17.

Lembcke, P. A. "Measuring the Quality of Medical Care Through Vital Statistics Based on Hospital Service Areas." *American Journal of Public Health* 42 (March 1952):276-86.

Lewis, C. "Variation in the Incidence of Surgery." *New England Journal of Medicine* 281 (16 October 1969):880-84.

McCarthy, E. C., and M. L. Finkel. "Surgical Utilization in the U.S.A." *Medical Care* 118 (September 1980):883-92.

McCracken, G., and S. Bognanni. "Iowa Employers Use Small Area Analysis in Benefits Reform." *Business and Health* 3 (September 1986):14-16.

McPherson, K., et al. "Regional Variations in the Use of Common Surgical Procedures: Within and Between England and Wales, Canada and the United States of America." *Social Science and Medicine* 14 (May 1981): 273-88.

McPherson, K., et al. "Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England and Norway." *New England Journal of Medicine* 307 (18 November 1982): 1310-14.

Roos, L. L., Jr., and N. P. Roos. "Assessing Existing Technologies: The Manitoba Study of Common Surgical Procedures." *Medical Care* 21 (April 1983):454-62.

Roos, L. L., Jr., N. P. Roos, and P. D. Henteleff. "Assessing the Impact of Tonsillectomies." *Medical Care* 16 (June 1978):502-18.

Roos, N. P. "Hysterectomy: Variations in Rates Across Small Areas and Across Physician's Practices." *American Journal of Public Health* 74 (April 1984):327-35.

Roos, N. P., and L. L. Roos, Jr. "A New Audit Procedure Applied to an Old Question: Is the Frequency of T and A Justified?" *Medical Care* 15 (January 1977):1-18.

_____. "Surgical Rate Variations: Do They Reflect the Health or Socio-Economic Characteristics of the Population?" *Medical Care* 20 (September 1982):945-58.

Roos, N. P., and P. Gilbert. "Tonsillectomy in Manitoba: Who Are the Patients? the Surgeons? the Hospitals?" *Journal of Community Health* 5 (Winter 1975):101-12.

Roos, N. P., L. L. Roos, and P. D. Henteleff. "Elective Surgical Rates: Do High Rates Mean Lower Standards?" *New England Journal of Medicine* 297 (18 August 1977):360-65.

Stockwell, H., and E. Vayda. "Variations in Surgery in Ontario." *Medical Care* 17 (April 1977):390-96.

Wennberg, J. E. "Dealing with Medical Practice Variation: A Proposal for Action." *Health Affairs* 3 (Summer 1984):7-31.

_____. "On Patient Need, Equity, Supplier-Induced Demand, and the Need to Assess the Outcome of Common Medical Practices." *Medical Care* 23 (May 1985):512-20.

Wennberg, J. E., and A. Gittelsohn. "Variations in Medical Care Among Small Areas." *Scientific American* 246 (April 1982):120-33.

Wennberg, J. E., L. Blowers, R. Parker, et al. "Changes in Tonsillectomy Rates Associated with Feedback and Review." *Pediatrics* 59 (June 1977):821-26.

Wennberg, J. E., J. P. Bunker, and B. Barnes. "The Need for Assessing the Outcome of Common Medical Practices." *Annual Review of Public Health* 1 (1980):277-95.

Willemain, T. R. "On the Comparison of Highest and Lowest Surgery Rates in Small Area Studies." In *Regional Variations in Hospital Use: Geographic and Temporal Patterns of Care in the United States*. Edited by David L. Rothberg. Lexington, MA: Lexington Books, 1982.

Wilson, P., and P. Tedeschi. "Community Correlates of Hospital Use." *Health Services Research* 19 (August 1984):333-55.

Wilson, P., et al. "Does Race Affect Hospital Use?" *American Journal of Public Health* 75 (March 1985):263-69.