
Articles

Testing the Null Hypothesis in Small Area Analysis

Kevin C. Cain and Paula Diehr

The goal of small area analysis is often to demonstrate that hospital admission rates or procedure rates vary greatly among regions, suggesting the occurrence of unnecessary admissions or procedures in some regions. Recent articles have shown that such variation may be largely due to chance, even if no underlying differences exist among the small areas; thus, it is important to test if the observed variation is larger than expected by chance. In this article we discuss how the appropriate method for testing the null hypothesis depends on the distribution of the number of admissions at the person level. If it is not possible for an individual to have more than one admission for a given procedure, the appropriate test is a simple chi-square test. If multiple admissions are possible, a modified chi-square test can be used to account for the excess variability due to multiple admissions. Failure to make the correct modification to the chi-square test in this latter case can result in spurious results. This underscores the importance of collecting data on multiple admissions in order to estimate the distribution of the number of admissions at the individual-patient level.

Small area analysis is a popular methodology in health services research. This type of analysis is used when health services use rates are known within each of several "small areas." The goal of the analysis is usually to demonstrate that the rates differ across areas, and perhaps to explain these differences in terms of differences in physician practice styles or patient population characteristics. Many such analyses have

This work was supported in part by USPHS grant number R01-CA40644 and grant number HS06344 (the Back Pain Outcome Assessment Team) from the Agency for Health Care Policy and Research.

Address correspondence and requests for reprints to Kevin C. Cain, Ph.D., Assistant Professor, Department of Biostatistics, SC-32, University of Washington, Seattle, WA 98195. Paula Diehr, Ph.D. is Professor of Biostatistics, University of Washington. This article, submitted to *Health Services Research* on August 10, 1990, was revised and accepted for publication on June 21, 1991.

been published, and several review articles have been written in this field (*Health Affairs* 1984; Copenhagen Collaborating Center 1985; Paul-Shaheen, Clark, and Williams 1987; Wennberg 1990).

In this article we restrict attention to small area analyses in which health services use rates are calculated as the number of hospital admissions (or surgical procedures, office visits, etc.) divided by the population size. We consider only rates that are counts of events and not analyses in which health services use rates are measured as dollar cost or length of stay. For simplicity, we refer throughout this article to the "small area" as a county and the event of interest as an "admission." However, the methods discussed here are also applicable to any other health services use measure that is a count of the number of events. Other examples of events of interest might be surgical procedures, diagnostic procedures, lab test orders, office visits for a specific disorder, or dental check-up visits. These methods are also of use in outcomes research and epidemiology. Examples of relevant event types here would be incidence of surgical complications in a hospital, and incidence of cancer or mortality in a community.

We can assume that the observed rate will differ from the expected rate in a county, due to random variation. In order to make inferences about how rates vary across counties, we need to know the distribution (or at least the variance) of this random variation. In particular, we are often interested in testing the null hypothesis that the underlying admission rate is actually the same in all counties, and the observed rates differ from each other *only* due to random variation. In order to test whether the observed variation is or is not bigger than expected, we need to know the amount of variation that is expected by chance when the null hypothesis is true.

A recent article (Diehr et al. 1990) used simulations to show that the variation in rates can be quite large, even when the null hypothesis is true. In this article we take a somewhat more theoretical approach. We consider four theoretical distributions for modeling the number of admissions an individual can have, and the situations in which these models might be appropriate. We then show how these assumptions about the behavior of individuals translate into the distribution of number of admissions at the county level. Once we know the probability distribution of number of admissions (and hence rates) at the county level, we can determine the appropriate analysis method. The most important aspect of the distribution at the individual level is whether multiple admissions are possible or not. If multiple admissions (or events) in the specified time period are either impossible or highly unusual, then the analysis is simple and straightforward. Examples of

this are most "ectomies" such as hysterectomy, and some outcome measures such as mortality.

The analysis is more problematic when readmissions are possible. Examples in which multiple admissions (or events) are common are hospital admission for cancer chemotherapy, chronic obstructive pulmonary disease, manifestations of coronary artery disease, or complications of diabetes, as well as office visits or laboratory tests for conditions such as urinary tract infections. If one incorrectly performs an analysis by assuming that multiple admissions are not possible when they are in fact common, the results may well be spurious since the variance in number of admissions will in general be underestimated. We propose an analysis based on a normal approximation that can be used when the variance of number of admissions at the individual level is known and none of the counties is too small. In order to use this proposed analysis method one needs to know only the variance of the distribution; it is not necessary to know precisely which theoretical distribution best models the data.

THE MODELS

Small area analyses usually use age-sex-adjusted rates. However, for ease of presentation we will focus first on the simple case in which age-sex stratification can be ignored because the probability of an admission is the same across all strata. (Alternatively, we can think of the analysis as being done within a single stratum.) The generalization to stratified analyses is discussed later.

Let n_j denote the number of people in county j . Let y_{ij} denote the number of admissions for the i th person in county j . The total number of admissions in county j is

$$Y_j = \sum_{i=1}^{n_j} y_{ij}$$

and the rate of admissions per thousand is $1000 Y_j/n_j$. In most small area analyses, person-level data are not available. The total number of admissions Y_j (or the rate per thousand), is known, but the y_{ij} s are not. However, it is generally easier to consider hypothetical distributions and decide which ones make sense at the individual level than at the county level. We will therefore start with assumptions about the distribution of individual admissions (y_{ij}) and then show the resulting distribution at the county level (Y_j).

We begin with the following assumptions, which hold for all the distributions we consider:

Assumption 1. The random variables $y_{1j}, y_{2j}, \dots, y_{n,j}$ are independent.

Assumption 2. The random variables $y_{1j}, y_{2j}, \dots, y_{n,j}$ are identically distributed, and all have the same expected value (i.e., mean) m_j .

Assumption 1 implies that the probability that person 1 has an admission does not depend on whether person 2 has an admission. This seems reasonable for most types of admissions or procedures. However, there are some situations for which this assumption clearly does not hold. Examples are infectious diseases and trauma due to large-scale disasters such as airplane crashes and hurricanes. Statistical analysis methods for use when assumption 1 is violated are beyond the scope of this article and are not discussed here.

Assumption 2 implies that everyone in county j has the same expected number of visits. At first this may seem inappropriate since individuals clearly differ in their level of health. For example, an individual with a chronic disease is more likely to be hospitalized than one without. However, assumption 2 means not that all persons have the same level of health, but rather that we do not have information about the health status of individuals. Person 1 is just as likely to be healthy or sickly as person 2; therefore, the probability of having one admission is the same for person 1 as for person 2 if we have no individual information about health status. Statistical analysis methods are available for use with data in which individual level information is known (Mauritsen 1984; Wong and Mason 1985), but again, these are beyond the scope of this article.

The expected value (M_j) and variance of the total number of admissions in county j can be calculated based on assumptions 1 and 2:

$$M_j = E(Y_j) = \sum_{i=1}^{n_j} E(y_{ij}) = \sum_{i=1}^{n_j} m_j = n_j m_j,$$

$$\text{Var}(Y_j) = \sum_{i=1}^{n_j} \text{Var}(y_{ij}) = n_j \text{Var}(y_{ij}).$$

We now present four possible models for the distribution of the y_{ij} s, and show how the distribution and variance of Y_j depends on the model for generating the y_{ij} s. These four models were chosen because

they represent simple and understandable mechanisms for generating the number of admissions at the individual level. The purpose of this section is to gain insight into sources of variability in Y_j , not to present an exhaustive class of probability distributions among which one should choose when doing an analysis. In fact, the data presented at the end of this section are too skewed for any of these four distributions to fit well.

BERNOULLI (BINOMIAL) DISTRIBUTION

Suppose that multiple admissions are not possible, so that a person can have at most one admission. This is true for some procedures such as hysterectomy. There are other types of admission for which multiple admissions theoretically are possible but are relatively rare so that it is a reasonable approximation to assume that they do not occur.

The Bernoulli distribution is the appropriate distribution to use when multiple admissions are not possible. (In fact, the Bernoulli is the *only* distribution applicable when individuals can have only zero or one admission. This is discussed further in the section titled "Chi-square Tests.") Under a Bernoulli distribution with expected value m_j , the random variable y_{ij} is equal to one with probability m_j and is equal to zero with probability $(1-m_j)$. The variance of y_{ij} is $\text{Var}(y_{ij}) = m_j(1-m_j)$, and the total number of admissions for the county, Y_j , has a binomial distribution with $p = m_j$ and $n = n_j$, so that:

$$\text{Var}(Y_j) = n_j m_j (1 - m_j) = M_j (1 - m_j). \quad (1)$$

The probability density functions for the Bernoulli and binomial distributions are given in Appendix A. Figure 1a shows the distribution of y_{ij} when m_j is 0.7. This is admittedly a rather artificial situation (with 70 percent of persons having one admission, and no one having more than one), but m_j equal to 0.7 was chosen to accentuate the differences among the four distributions presented here. Figure 2a shows the distribution of Y_j in a county with 5,000 people and an expected admission rate of 1/1000 (i.e., $m_j = 0.001$ and $n_j = 5000$).

POISSON DISTRIBUTION

We next consider the Poisson distribution (Johnson and Kotz 1969). The probability density function for a Poisson distribution with mean m_j is shown in Appendix A. The variance of a Poisson random variable is equal to its mean, so that $\text{Var}(y_{ij}) = m_j$. A sum of Poisson random variables also has a Poisson distribution. Thus Y_j is Poisson with mean M_j and:

Figure 1: Plots of the Distribution of the Number of Admissions for One Individual (y_{ij}), under Four Different Assumed Distributions; in all cases, $E(y_{ij}) = m_j = 0.7$; (a) Bernoulli; (b) Poisson; (c) Bernoulli Mixture of Poissons, with $b_j = 2$; (d) negative binomial with $k_j = 0.54$

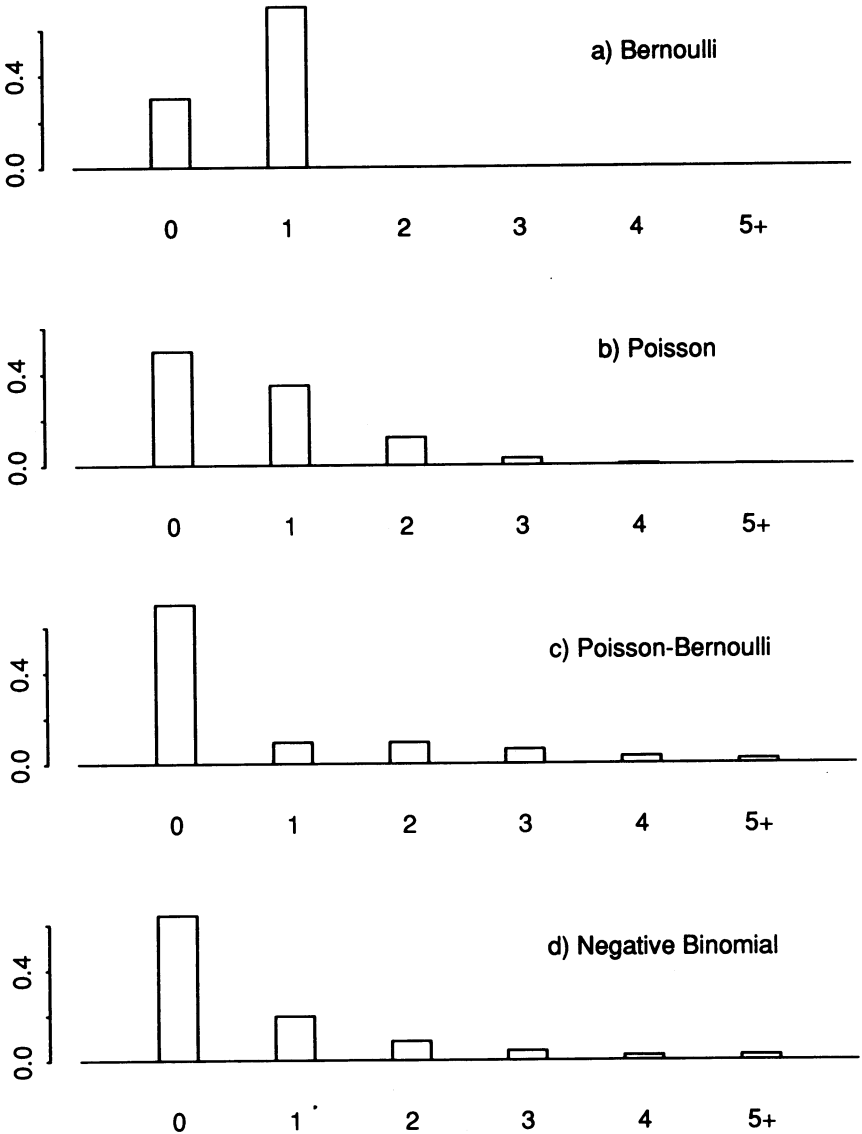
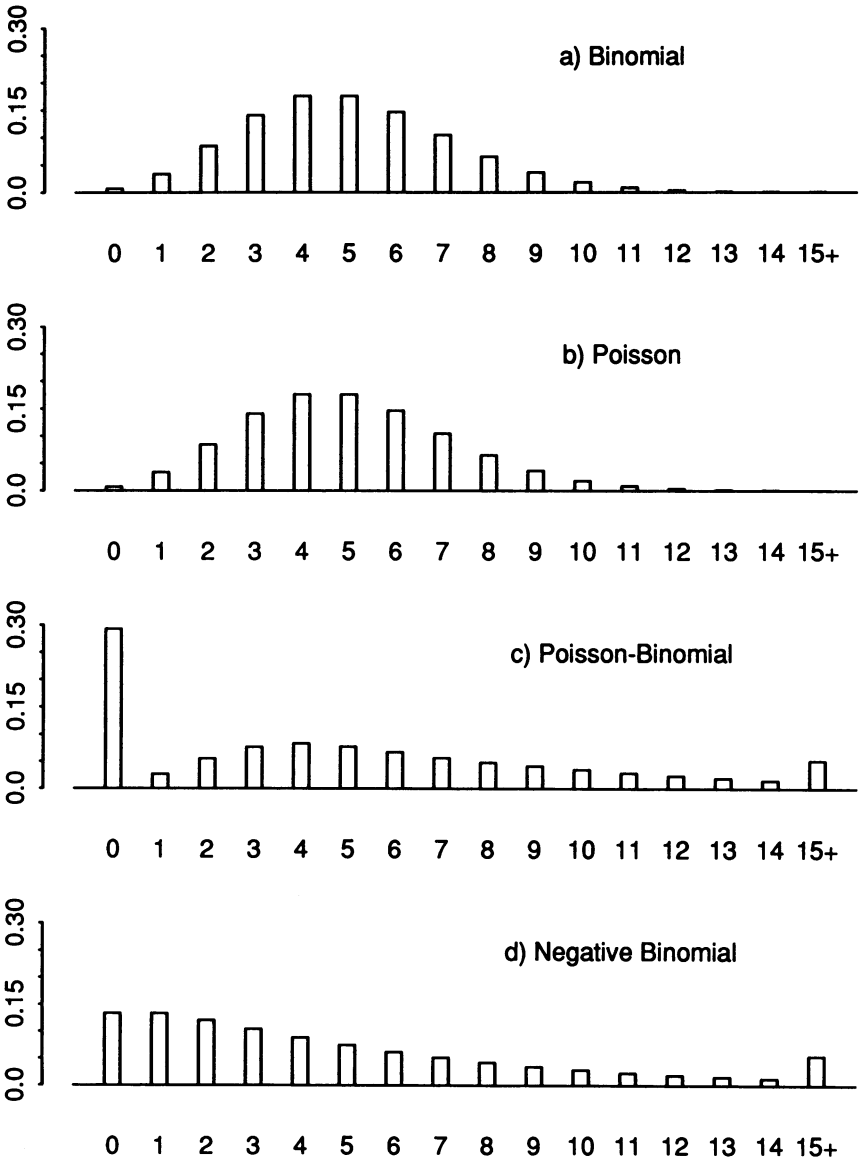


Figure 2: Plots of the Distribution of the Number of Admissions for 5,000 Individuals (Y_j), under Four Different Assumed Distributions; in all cases, $E(y_{ij}) = 0.001$ and $E(Y_j) = M_j = 5.0$; the distribution of y_{ij} for each Individual is (a) Bernoulli; (b) Poisson; (c) Bernoulli Mixture of Poissons, with $b_j = 4$; (d) negative binomial with $k_j = 0.00025$



$$\text{Var}(Y_j) = n_j m_j = M_j. \quad (2)$$

A comparison between Equation 1 and Equation 2 shows that the variance of Y_j is somewhat smaller under the Bernoulli assumption than under the Poisson assumption, but if m_j is close to zero the two variances will be about the same.

Figure 1b shows a Poisson distribution with $m_j = 0.7$. This distribution looks very different from the Bernoulli distribution shown in Figure 1a, which has the same mean. However, if m_j is close to zero the two distributions will look very similar since the probability of multiple admissions will be very small. Figure 2b shows the distribution of Y_j when y_{ij} is from a Poisson distribution with mean 0.001 and the population size of the county is 5,000. A comparison of figures 2a and 2b shows very close agreement between the two distributions.

The primary way in which y_{ij} can have a Poisson distribution is if admissions follow a Poisson process. This means that the times between successive admissions are independent and follow an exponential distribution. This implies, for example; that the expected number of admissions a person has in February is independent of the number of admissions he or she had in January. This is not in general a reasonable assumption. For most types of admission, a person with one admission in January is more likely to have an admission in February than is a person with no admission in January. For some procedures, such as "ectomies," the converse is true.

Even though the Poisson distribution is usually not appropriate when multiple admissions are possible, it can be useful as an approximation to the binomial distribution when m_j is small and multiple admissions are not possible. The Poisson distribution is also useful as a standard against which to compare other distributions. Since the variance of the Poisson distribution is so simple (Equation 2), we will describe other distributions in terms of how much bigger (or smaller) their variance is compared to the Poisson variance.

We now turn to a class of distributions known as mixture distributions. The next two subsections describe two members of this class, the Poisson-Bernoulli distribution and the negative binomial distribution.

POISSON-BERNOULLI (POISSON-BINOMIAL) DISTRIBUTION

In the previous subsection the distribution of number of admissions followed a Poisson distribution with mean m_j for all persons in the county. We now generalize the Poisson distribution to a heterogeneous population.

Assume that among persons with a given health status the number of admissions follows a Poisson distribution. However, health status is not the same for everyone in the county, which means that the mean of the Poisson distribution is different for different subgroups. Persons with good health will have a lower mean than those with poorer health.

Since the health status of individuals is not known, we will think of health status as a random variable that varies across individuals in the population. There are now two sources of variability in y_{ij} : the random health status of person i , and the random number of admissions conditional on health status. The distribution of y_{ij} in this situation is referred to as a mixture distribution.

In this subsection we describe a simple mixture distribution in which only two levels of health are possible: either a person does or does not have a particular chronic disease. Let us use diabetes as an example, and assume that the events of interest are hospital admissions for diabetes, so that "admission" means hospitalization for diabetes.

Let p_j be the probability that a person has diabetes. Let b_j denote the mean number of admissions among persons with diabetes, and assume that among persons with diabetes the number of admissions follows a Poisson distribution. Persons without diabetes have no admissions for diabetes (this can be thought of as a degenerate Poisson distribution, with mean zero). For a person whose disease status is unknown, we can think of y_{ij} as being generated by a two-step process: first, a health status is randomly generated for the person (diabetes or no diabetes); then, y_{ij} is randomly generated from a Poisson distribution with mean dependent on health status (b_j or 0). The expected value of Y_{ij} under this two-step process is

$$\begin{aligned} m_j &= E(y_{ij}) = E(y_{ij}|\text{diabetes}) \Pr(\text{diabetes}) \\ &\quad + E(y_{ij}|\text{no diabetes}) \Pr(\text{no diabetes}) \\ &= b_j p_j + 0(1 - p_j) = b_j p_j. \end{aligned}$$

A somewhat more complicated derivation gives the variance of y_{ij} under the two-step process as

$$\text{Var}(y_{ij}) = m_j [1 + b_j(1 - p_j)].$$

The total number of admissions (for diabetes) in the county, Y_j , has mean $M_j = n_j p_j b_j$ and variance

$$\text{Var}(Y_j) = M_j [1 + b_j(1 - p_j)]. \quad (3)$$

The distribution of Y_{ij} is referred to as a Bernoulli mixture of Poissons, and the distribution of Y_j is a binomial mixture of Poissons or

a Poisson-binomial distribution (Johnson and Kotz 1969). A comparison of Equation 3 with Equation 2 shows that the variance of the Poisson-binomial distribution is larger than the variance of a Poisson distribution with the same mean, and can be much bigger than the Poisson variance if b_j is large and p_j is small.

Figure 1c shows a plot of this mixture distribution for one individual with $m_j = 0.7$ and $b_j = 2$ (and hence $p_j = m_j/b_j = .35$). Compared to the Poisson probability density function with the same mean, the Poisson-Bernoulli density has a higher probability of zero admission, a much lower probability of only one admission, and a much higher probability of having three or more admissions. A comparison of equations 2 and 3 shows that when $m_j = 0.7$ and $b_j = 2$ the variance of the Poisson-Bernoulli distribution is 2.3 times that of a Poisson with the same mean. For rare diseases (i.e., p close to zero) the variance is approximately $(1 + b_j)$ times the variance of a Poisson distribution.

Figure 2c shows the distribution of the total number of admissions for a county of 5,000 people, each of whom has a Poisson-Bernoulli distribution with $m_j = 0.001$, $b_j = 4$, and $p = .00025$. The probability of 15 or more admissions is 5 percent, compared to virtually zero under the Poisson distribution. The probability of zero admissions in the county is about 30 percent, compared to 1 percent for the Poisson distribution. Even though the expected number of admissions is five, the expected number of individuals with the disease is only $p_j n_j = 1.25$.

The variance of Y_j in Figure 2c is 4.99 times what it would be under a Poisson distribution. The heterogeneity in the population allowed by this model (some persons have diabetes, some do not) has greatly increased the variance over what it would be in a homogeneous population where everyone has the same disease status.

NEGATIVE BINOMIAL DISTRIBUTION

In the previous subsection we assumed that there were only two possible levels of health. Letting m_{ij} denote the expected number of admissions for person i , we saw that under the Poisson-Bernoulli model m_{ij} is either b_j if the person has diabetes or 0 if he does not. Now suppose that the level of health actually varies continuously from perfect health ($m_{ij} = 0$) to terrible health (m_{ij} very large). Let us further suppose that level of health (as indexed by m_{ij}) is a random variable distributed as a gamma distribution with mean m_j and variance m_j/k_j (Johnson and Kotz 1970). The parameter k_j is called the shape parameter of the gamma distribution: if k is small the distribution is highly skewed to the

right, while large k corresponds to a distribution that is approximately normal. If kj is large relative to m_j , the variance of m_{ij} is small so that level of health does not vary much across individuals. On the other hand, if kj is small relative to m_j , the variance of m_{ij} is large so that health varies greatly across individuals.

Assume now that if m_{ij} is known, the number of admissions for person i is distributed as a Poisson random variable with mean m_{ij} . For a person whose disease status is unknown, we can think of y_{ij} as being generated by the two-step process of first generating m_{ij} from a gamma distribution with mean m_j and variance m_j/k_j and then generating y_{ij} from a Poisson distribution with mean m_{ij} . It can then be shown that when m_{ij} is unknown, $E(y_{ij}) = m_j$ and $\text{Var}(y_{ij}) = m_j(1 + m_j/k_j)$, and hence

$$\text{Var}(Y_j) = M_j(1 + m_j/k_j). \quad (4)$$

The distribution of y_{ij} can be referred to as a gamma mixture of Poissons, but it is more commonly called a negative binomial distribution (Johnson and Kotz 1969). Its probability density function is shown in Appendix A.

If m_j/k_j is small, the negative binomial is almost the same as the Poisson distribution. If m_j/k_j is large, the negative binomial is more skewed to the right and has a much bigger variance than the Poisson. It can be shown that a sum of negative binomial variables with a common shape parameter will also have a negative binomial distribution. That is, Y_j will be negative binomial with mean $M_j = n_j m_j$ and shape parameter $n_j k_j$.

Figure 1d shows a negative binomial distribution for one individual with mean 0.7 and shape parameter 0.54. The shape of this probability density function is intermediate between the Poisson and the Poisson-Bernoulli. A comparison of equations 2 and 4 shows that the variance of this negative binomial is 2.3 times that of the Poisson with the same mean.

Figure 2d shows the distribution of Y_j when the distribution of y_{ij} for each of the 5,000 individuals is negative binomial with mean 0.001 and shape parameter 0.00025. This means that Y_j is negative binomial with mean 5.0 and shape parameter 1.25, so that the variance is 5.0 times that of the Poisson. The probabilities of observing zero, one, and two admissions in the negative binomial are very different from such probabilities in the Poisson-binomial distribution, but the two distributions are similar for more than three admissions.

Example: Hospitalizations for Nonelective Procedures in the Elderly

We now apply the Poisson, Bernoulli-Poisson, and negative binomial distributions to an actual data set in order to see how well these distributions approximate the distribution of y_{ij} . The first column of Table 1 shows the distribution of number of hospital discharges at the individual level, based on data from the Washington State Commission Hospital Abstracting System (CHARS) for the calendar year 1987. This table only includes hospitalizations of persons aged 65 and over in which one of the nonelective surgeries listed by Pasley, Vernon, Gibson, et al. (1987) was performed. Note that 11,815 of the 531,155 elderly residents of Washington State (2.2 percent) had at least one admission, and that three of them had seven admissions each. The mean number of hospitalizations per person is 0.0235 and the variance is .0263. This variance is 11 percent higher than expected under a Poisson distribution.

The second column of Table 1 shows the expected frequencies if the data are from a Poisson distribution with mean 0.0235. Notice that the number of persons with two or more admissions is severely underestimated. This underestimation carries over to the distribution of Y_j , the total number of admissions in a county. Consider a county of size 1000, with expected number of admissions $M_j = 1000 m_j = 23.5$. If 1000 values of y_{ij} are sampled (with replacement) from the population

Table 1: Observed and Expected Frequencies of Number of Admissions for Certain Nonelective Procedures among the Elderly in Washington State

	<i>Expected Frequency under Theoretical Probability Distributions*</i>			
	<i>Observed Frequency</i>	<i>Poisson</i>	<i>Bernoulli-Poisson</i>	<i>Negative Binomial</i>
0	519,340	518,831.3	519,548.8	519,517.6
1	11,312	12,179.7	10,784.2	10,872.7
2	415	143.0	782.7	703.5
3	50	1.1	37.9	55.8
4	22	0.0	1.4	4.8
5	10	0.0	0.0	0.4
6	3	0.0	0.0	0.0
7	3	0.0	0.0	0.0
>7	0	0.0	0.0	0.0
Total	531,155	531,155.0*	531,155.0*	531,155.0*

Source: Washington State Commission Hospital Abstracting System.

*Expected frequencies do not sum to 531,155 due to rounding.

shown in the first column of Table 1 and then added up, the probability that the total is greater than or equal to 40 is .0025. However, if observations are sampled from the Poisson distribution shown in the second column, the probability that the total is greater than or equal to 40 is only .0010. Thus, the Poisson distribution provides an anticonservative approximation when testing whether a county with a high admission rate is an outlier. It should be noted that in this example the variance is only slightly larger (11 percent) than the Poisson. If the variance were 50 percent larger, the discrepancy between the actual data and the Poisson approximation would be much worse.

We next fit a Poisson-Bernoulli distribution to the data by finding the parameters p_j and b_j , which give the same mean and variance as observed in the data. Letting m_j and v_j denote the observed mean and variance, we get

$$b_j = m_j + v_j/m_j - 1 = 0.0235 + 0.0263/0.0235 - 1 = 0.143,$$

$$p_j = m_j/b_j = 0.0235/0.143 = 0.164.$$

The third column of Table 1 shows the expected frequencies if the data are from a Poisson-Bernoulli distribution with these parameters. The fit of observed to expected is better than with the Poisson distribution, but the number of persons with four or more admissions is still underestimated.

The fourth column of Table 1 shows the expected frequencies if the data are from a negative binomial distribution with parameters $m_j = 0.0235$ and $k_j = m_j^2/(v_j - m_j) = 0.197$. The fit is only slightly better than the Poisson-Bernoulli. In order to substantially improve the fit in the upper tail (for number of admissions greater than three), we would need to use a more complicated distribution, such as a Bernoulli mixture of negative binomials.

The two mixture distributions give reasonably good fits to the distribution of Y_j , the total number of admissions for a county of size 1000. The probability of observing 40 or more admissions is .0022 under the Poisson-Binomial model and .0026 under the negative binomial model, compared to .0025 in the observed data.

As the size of the county (n_j) goes to infinity, the distribution of the total admissions in the county Y_j approaches a normal distribution with mean $n_j m_j$ and variance $v_j m_j$. If we approximate the actual distribution by any theoretical distribution with mean m_j and variance v_j , then the distribution of Y_j under this theoretical distribution will also approach a normal distribution with mean $n_j m_j$ and variance $v_j m_j$. Thus, any theoretical distribution will give a good approximation to the distribu-

tion of Y_j for large counties, as long as the mean and variance are correct. The problem with using the Poisson distribution as an approximation is that the variance will usually be wrong.

In this section we have seen how the distribution (and in particular the variance) of the number of admissions at the county level Y_j is quite dependent on the model that generates the data at the person level. Heterogeneity of health status in the population can lead to the variance of Y_j being much larger than it would be under the Poisson model. However, the Poisson distribution is frequently used as the basis for inference in small area analyses due to its simplicity, even when it is not appropriate. In the next section we discuss methods for testing the null hypothesis, and propose an adjusted chi-square test for use when the variance is bigger than the Poisson variance. The researcher does not need to know which distribution fits the data best, only its variance. Methods that do require knowledge of the distribution (likelihood-based and simulation-based methods) are discussed briefly, but a full discussion of these methods is beyond the scope of this article.

METHODS FOR TESTING THE NULL HYPOTHESIS

We wish to test the null hypothesis that the expected admission rate is the same in all counties, and that the differences in observed rates are no bigger than that expected by chance. Formally, we have

$$H_0: m_1 = m_2 = \dots m_J,$$

where J is the number of counties. We now describe methods for testing H_0 . We begin with a class of chi-square tests which are appropriate if the expected number of admissions is "large" in all counties. We then describe two alternative methods which may be preferred when this condition is not met, although further research is needed to determine rules for deciding how large is "large."

CHI-SQUARE TESTS

If there are no multiple admissions, the Bernoulli distribution is applicable. In this case a simple chi-square test can be used to test H_0 . Let \hat{m} denote the observed admission rate for all counties taken together:

$$\hat{m} = \frac{\sum_{j=1}^J Y_j}{\sum_{j=1}^J n_j}.$$

Under the null hypothesis the expected number of admissions in county j is estimated to be $\hat{m} n_j$. The familiar formula for the chi-square test statistic is written in terms of the observed (O) and expected (E) numbers in the cells of the 2 by J table. In our notation this becomes

$$\begin{aligned} X_B^2 &= \Sigma(O - E)^2 / E \\ &= \sum_{j=1}^J \left\{ (Y_j - n_j \hat{m})^2 / n_j \hat{m} + [(n_j - Y_j) - (n_j - n_j \hat{m})]^2 / (n_j - n_j \hat{m}) \right\} \\ &= \sum_{j=1}^J (Y_j - n_j \hat{m})^2 / \hat{m}(1 - \hat{m})n_j, \end{aligned} \quad (5)$$

where the subscript B in X_B^2 indicates that this formula is appropriate for data with a Bernoulli distribution. The value calculated from this formula can then be compared to a chi-square distribution with $J-1$ degrees of freedom in order to test H_0 . The expected number of admissions should be at least five in every county in order for the chi-square approximation to be reasonable.

This simple chi-square test is appropriate if multiple admissions are not possible so that the data have a Bernoulli distribution at the individual level. This is true even if the probability of admission is different for different individuals, as long as we do not know what the individual probabilities are. The recent article by Kazandjian, Durance, and Schork (1990) assumes that multiple admissions are not possible and then claims to model variation among individuals using a binomial-beta distribution (Johnson and Kotz 1969). However, a beta mixture of Bernoullis is still a Bernoulli (a proof of this is given in Appendix B). What Kazandjian, Durance, and Schork are in fact doing is modeling Y_j , the total number of admissions for the county, as a binomial-beta, which means they assume that m_j varies randomly across counties. This in effect assumes that the null hypothesis is not true, which is not appropriate if the goal is to test the null hypothesis.

Pasley, Vernon, Gibson, et al. (1987) present data on rates of elective surgery among the elderly in the counties of New York State. We will use some of the data from this article to demonstrate the calculation of X_B^2 . (Note: the rates given by Pasley et al. are age-sex adjusted, but for the purpose of this example we ignore this fact, and compute the number of surgeries by multiplying the rate by the population size.) The first two columns of Table 2 show the elderly population size and number of elective surgeries for the seven smallest counties in New York. The overall rate for these counties is $\hat{m} = 0.024$

Table 2: Illustration of the Calculation of the Chi-square Statistics, Using Data on the Number of Nonelective Procedures among the Elderly Population in the Seven Smallest Counties in New York State

j	n_j	Y_j	$n_j\hat{m}$	$(Y_j - n_j\hat{m})^2$	$(Y_j - n_j\hat{m})^2$
				$n_j\hat{m}(1 - \hat{m})$	$n_j\hat{m}$
1	837	15	18.30	0.61	0.59
2	2,196	48	48.00	0.00	0.00
3	2,913	64	63.70	0.00	0.00
4	3,266	72	71.40	0.00	0.00
5	3,872	62	84.60	6.19	6.05
6	4,424	87	96.7	1.00	0.97
7	4,543	134	99.3	12.39	12.12
Total	22,051	482	482.00	20.19	19.75*

Source: Pasley, Vernon, Gibson, et al. 1987.

*The last column does not sum to 19.75 due to rounding.

(i.e., 24 per thousand). The third column shows the expected number of surgeries under the null hypothesis. Since it is clearly possible for a person to have more than one elective surgery in a year it is not appropriate to use Equation 5 for these data. However, we will do so as an illustration of the computations. The fourth column of Table 2 shows the terms for the sum in Equation 5, with the resulting sum shown at the bottom of this column. This sum, 20.19, can be compared to a table of the chi-square distribution with 6 degrees of freedom, giving a significance level of $p = .0026$.

The chi-square test easily generalizes to the situation in which the data are age-sex stratified. Within each stratum a separate 2 by J table of observed and expected is calculated as above. These tables are then combined into one by summing over all strata the observed and expected numbers within each cell. The X^2_B statistic is then calculated from this combined table using Equation 5, and compared to a chi-square distribution with $J-1$ degrees of freedom. This test is referred to as the Mantel-Haenszel test (Mantel and Haenszel 1959; Fleiss, 1981).

In addition to testing the overall null hypothesis, one may wish to determine which counties have rates significantly higher or lower than the other counties. This can be done as follows. For each county j , test whether the rate in county j is significantly different from that in the other counties taken together by doing a chi-square test on the 2 by 2 table where the first row gives the numbers of persons with and without

an admission in county j , and the second row gives the numbers of persons with and without an admission in all of the other counties combined. The Yates continuity correction or Fisher's exact test can be used if the expected number of admissions in county j is small. The Bonferroni adjustment for multiple comparisons should be used when testing individual counties. This means that the p -value required before a test is declared significant is $.05/J$ rather than $.05$.

As an example, consider again the data shown in Table 2. There are 134 surgeries and a population of 4,543 in county 7, and 348 surgeries and a population of 17,508 in the six other counties combined. The resulting 2 by 2 table gives $X^2_B = 15.64$. Comparing this to a chi-square distribution with 1 degree of freedom gives a p -value of $.00008$. Since this is less than $.05/7 = .007$, we conclude that the surgery rate in county 5 is significantly different from that of the other counties.

An equally simple chi-square test is appropriate when the data are from a Poisson distribution (Brown and Hollander 1977, 196). Under the null hypothesis, the expected number of admissions in county j is $n_j \hat{m}$. Comparing the observed to the expected number of admissions in each county and summing over counties gives

$$\begin{aligned}
 X^2_p &= \sum (O - E)^2 / E \\
 &= \sum_{j=1}^J (Y_j - n_j \hat{m})^2 / (n_j \hat{m}).
 \end{aligned}
 \tag{6}$$

As above, this can be compared to a chi-square distribution with $J-1$ degrees of freedom. Notice that Equation 5 differs from Equation 6 only by a factor of $(1-\hat{m})$ in the denominator of Equation 5. If \hat{m} is close to zero, the value of X^2_p calculated from Equation 5 will be almost the same as X^2_B calculated from Equation 6. This is illustrated in the example shown in Table 2. The fifth column shows the terms of the sum in Equation 6. The resulting chi-square statistic is 19.75 ($p = .0031$), compared to 20.19 ($p = .0026$) calculated from Equation 5.

Notice that in both equations 5 and 6, the denominator is $\text{Var}(Y_j) = n_j \text{Var}(y_{ij})$. This suggests a generalization of the definition of the chi-square test statistic. First, define

$$\text{MAF} = \text{Var}(y_{ij}) / \hat{m}.
 \tag{7}$$

MAF stands for Multiple Admission Factor and is the ratio of the actual variance of y_{ij} to what the variance would be if the data were from a Poisson distribution. Now define

$$\begin{aligned}
 X_{\text{MAF}}^2 &= \sum_{j=1}^J (Y_j - n_j \hat{m})^2 / [n_j \text{Var}(y_{ij})] \\
 &= \sum_{j=1}^J (Y_j - n_j \hat{m})^2 / (n_j \hat{m} \text{MAF}) \\
 &= X_p^2 / \text{MAF}.
 \end{aligned}
 \tag{8}$$

Suppose that some data are available on the distribution of admissions at the individual level (y_{ij}), so that it is possible to estimate m_j and $\text{Var}(y_{ij})$ and hence MAF. These data could be a sample of persons from the counties under consideration, or could be from studies done by other researchers on similar populations. By analogy with the Bernoulli and Poisson tests described earlier, we propose the following procedure for testing the null hypothesis when MAF is (at least approximately) known.

Calculate X_p^2 from Equation 6, then calculate $X_{\text{MAF}}^2 = X_p^2 / \text{MAF}$. Finally, compare X_{MAF}^2 to a chi-square distribution with $J-1$ degrees of freedom.

It can be shown that regardless of the distribution of y_{ij} , the distribution of X_{MAF}^2 when the null hypothesis is true will be approximately chi-square with $J-1$ degrees of freedom (Appendix C). The approximation will only be good if n_j is large enough in all counties to ensure that the distribution of Y_j is approximately normal.

How large do the n_j s have to be for this to be a reasonable approximation? The rule of thumb for the usual chi-square test is that the expected number of admissions should be greater than five in every county. In Figure 2a the expected number of admissions for the county is five and the Poisson distribution looks fairly normal, with only a slight skewness. However, in Figure 2c the expected number of admissions is also five, but the distribution is decidedly non-normal, being very skewed to the right. The probability of observing 18 or more admissions is .023, but the normal approximation gives a probability of only .005. Recall that under the Binomial-Poisson distribution shown in Figure 2c the expected number of people with the disease in the county is only 1.25. This suggests some possible alternative rules of thumb: (i) the expected number of *individuals* with one or more admissions should be greater than five in every county; (ii) the expected number of admissions in each county should be greater than five times the mean number of admissions among persons who have at least one admission; and (iii) the expected number of admissions in each county should be greater than five times MAF. An alternative strategy would

be to use an adjustment similar to the Yates continuity correction for 2 by 2 tables. For example, one could subtract $MAF/2$ from the absolute value of the numerator before squaring (or set to zero if the result is negative). Further work is needed to determine whether any of these ad hoc rules work well in practice under a wide range of distributions.

Consider again the example shown in Table 2. In order to use the test based on Equation 8 we need an estimate of MAF. The first column of Table 1 shows data from Washington State on hospitalizations of persons aged 65 and over in which one of the nonelective surgeries listed by Pasley, Vernon, Gibson et al. (1987) was performed. The mean number of hospitalizations per person is 0.0235 and the variance is 0.0263, which gives $MAF = 0.0263/0.0235 = 1.11$. Combining this with the value of X_p^2 from Table 2 gives $X_{MAF}^2 = 19.75/1.11 = 17.79$. Comparing this to a chi-square distribution with 6 degrees of freedom gives $p = .007$, still highly significant. The value of MAF in this example (1.11) is actually not much larger than 1. We have observed considerably larger values (1.5 to 4) for other types of admission.

LIKELIHOOD-BASED TESTS

Suppose that the data used to estimate the distribution of Y_{ij} are approximated fairly well by a Poisson-Bernoulli or a negative binomial distribution. In this case we could consider using a score test or a likelihood ratio test to test H_0 (Cox and Hinkley 1974). If the data do in fact have the specified probability density, then these likelihood-based tests will be more powerful than the test based on X_{MAF}^2 . The small sample performance should be better as well, although these tests are also based on asymptotic approximations and will behave poorly if the counties are too small.

Further research is needed to determine whether the improvement in power and small-sample performance is large enough or not to justify the added complexity of the likelihood-based methods. Another issue to be explored is how these tests behave if the distribution of the data is in fact somewhat different from that assumed (i.e., how robust they are to model misspecification).

TESTS BASED ON SIMULATIONS

Another approach to testing H_0 would use simulations to estimate the null distribution of a test statistic such as X_p^2 (Diehr et al. 1990; Diehr and Grembowski 1990). The simulations can use random numbers generated from a parametric distribution such as negative binomial, if

such a distribution appears to fit the data well. Alternatively, the simulations can be based on a nonparametric density estimate, the simplest being the unsmoothed empirical density function. We are currently exploring computation methods, such as the fast Fourier transform, which we hope will make the nonparametric approach feasible in practice.

OTHER APPROACHES

Another approach has been suggested for testing the null hypothesis. This consists of correlating the variation in rates across counties with some county population characteristics, such as per capita income or number of surgeons per capita. If the variation in rates is totally due to chance, then there should be no correlation with any population characteristics. This approach is somewhat reasonable, but one should be aware of the potential pitfalls described in Diehr et al. (1990). The small number of counties typical of small area analyses, the unequal variance of the estimated rates, the possibility of spurious correlation, and non-normality and influential outliers all potentially can lead to anticonservative tests. These problems might be alleviated somewhat by first taking z -scores of the rates [let $Z_i = (Y_i - \hat{m}_j) / \sqrt{\hat{m}_j}$] and then using Spearman rank correlation rather than Pearson correlation.

When the variation in rates across counties is larger than expected under the Poisson model, a random effects model can be used to incorporate this extra-Poisson variation into a regression analysis (Breslow 1984; Pocock, Cook, and Beresford 1981; Tsutakawa 1988; Wolf et al. 1989). The approach of these articles is rather different from the approach we have chosen. We have assumed that the extra-Poisson variation is due to the individual y_{ij} s having larger variance than under the Poisson (or Bernoulli) model. Thus, extra-Poisson variation can be present even if the null hypothesis is true.

Analyses based on random effects models, on the other hand, assume that the Poisson (or Bernoulli) model is appropriate for individuals, and the extra-Poisson variation is due to m_j being different in different counties (m_j varies randomly across counties). Such an analysis therefore assumes a priori that the null hypothesis is not true, and then incorporates this extra-Poisson variation in order to test the significance of regression coefficients for some variables of interest like physicians per capita. This is an entirely appropriate thing to do if multiple admissions are not possible and this type of extra-Poisson variation is present. However, this approach is not amenable to testing whether the null hypothesis is true if multiple admissions *are* possible.

McPherson et al. (1982) propose the systematic component of variation (SCV) as a measure of the component of variance in rates among counties that is *not* explained by random variation. However, their calculation is based on the assumption that the random variation follows a Poisson distribution. The authors state that "since surgery is a relatively rare event, we concluded that the distribution of O_i is approximately Poisson." This reasoning is not correct. Instead, what is required for the Poisson assumption to be approximately correct is that *multiple* surgeries (or admissions) be extremely rare. When this condition does not hold, the SCV should be interpreted as a measure of how much bigger the variance is than the Poisson variance, *not* how much true variance there is among the small areas.

SUMMARY AND DISCUSSION

In this article we show that when multiple admissions are possible, the variance of Y_j will probably be larger than it would be under a Poisson distribution, and hence tests based on the unadjusted chi-square, Equation 6, will not be valid. However, there are numerous examples of published works that use an unadjusted chi-square test to test H_0 when multiple admissions are possible. For example, Pasley, Vernon, Gibson, et al. (1987) use a chi-square test to compare elderly surgical discharge rates across counties in New York. Connell, Day, and LoGerfo (1981) use a chi-square test to compare rates of hospital admission for ear, nose, and throat surgery; gastroenteritis; and upper and lower respiratory infection in children across regions in Washington State. Chassin, Brook, Park, et al. (1986) use a chi-square test to compare rates for several procedures in Medicare beneficiaries across 13 sites around the United States. Some of these procedures, such as appendectomy and cholecystectomy, can only be done once and thus a chi-square test is appropriate. For some other procedures it is quite possible to have more than one procedure done in a year. Examples here are destruction of benign skin lesions and coronary angiography.

We offer the following recommendations for testing the null hypothesis:

1. *If multiple admissions are not possible*, use a chi-square test based on the binomial distribution (Equation 5). For rare admissions (admission rate less than 1 per 100), the chi-square test based on the Poisson distribution (Equation 6) will provide a good approximation. Small counties in which the expected number

of admissions is less than five should be excluded from the analysis.

2. *If multiple admissions are possible*, then obtain an estimate of the Multiple Admission Factor, MAF, and perform an adjusted chi-square test based on Equation 8. Small counties in which the expected number of admissions is less than five times MAF should be excluded from the analysis. Alternatively, one of the other rules of thumb suggested earlier could be used.
3. *If many of the counties are small*, then consider using a likelihood-based or simulation-based hypothesis test. These two approaches have not been described in detail here, in part because we believe that it is premature to recommend their use until completion of further research into the relative merits of the three approaches in various situations. In any case, a statistician should be consulted if one wishes to use a likelihood-based or simulation-based analysis.

Ideally, the data used to estimate MAF are a random sample of the population of the small areas being analyzed. Alternatively, data or published values of MAF may be available from another area and/or time period, as in our example in which data from Washington State for 1987 were used to estimate MAF for an analysis of data from New York for 1981. A third possibility is that no estimates of MAF are available for the specific type of admission of interest, but values of MAF are available for some other types of admissions similar in some respects to the admission type of interest. In all cases, and especially in this last situation, one should perform a sensitivity analysis to see if the result of the hypothesis test changes as MAF is varied over its range of plausible values. If so, the results must be interpreted with caution.

Sometimes data are available on the distribution of number of admissions among persons with at least one admission, but the number of persons who had no admissions is not accurately known. This is often the case when the small area is a hospital catchment area, and the size of the population served by the hospital is not accurately known due to overlap with other areas. In this case, it is still possible to calculate an approximation to MAF. Let m' and v' denote the mean and variance, respectively, of the number of admissions among persons with at least one admission. It can be shown that $MAF = m'(1-p) + v'/m'$, where p is the fraction of the population who had at least one admission. If p is small (which is usually the case), then an approximation to MAF can be obtained by setting p equal to 0 in this formula.

This approximation will lead to somewhat conservative hypothesis tests.

FURTHER RESEARCH

We conclude with some issues that need to be resolved by further work in this area:

1. The use of X_{MAF}^2 for testing H_0 is appropriate when all the counties are large. Is there a rule of thumb that works well in practice for deciding how small is too small? Under what circumstances do the likelihood-based, or the simulation-based approaches work best? Simulation studies can address these issues.
2. What values for the variance multiplier MAF are reasonable for procedures or admission types of interest? This can be addressed by analyzing individual-level data, where available. We plan to perform such an analysis on data from Washington State.
3. Any estimate of MAF will have some error associated with it. How does one incorporate this uncertainty into inference about the null hypothesis? This requires some more theoretical work.
4. What implications do multiple admissions hold for longitudinal data analysis? In particular, if a certain county has a consistently high admission rate over several years, can this be due to a few very sick individuals? Individual-level data over several years are needed to address these issues.

APPENDIX A

Probability Density Functions Mentioned in the Text

Bernoulli distribution with $p = m_j$:

$$f(y_{ij}) = m_j^{y_{ij}}(1 - m_j)^{(1-y_{ij})}, \text{ for } y_{ij} = 0, 1$$

Binomial distribution with $N = n_j$, $p = m_j$:

$$f(Y_j) = \binom{n_j}{Y_j} m_j^{Y_j} (1-m_j)^{(n_j - Y_j)}, \text{ for } Y_j = 0, 1, \dots, n_j$$

Poisson distribution with mean m_j :

$$f(y_{ij}) = \frac{e^{-m_j} m_j^{y_{ij}}}{y_{ij}!}, \text{ for } y_{ij} = 0, 1, \dots$$

Binomial-Poisson distribution of Y_j for a county of population n_j where each person has probability p_j of having the disease, and b_j is the expected number of admissions for people who have the disease (note: $m_j = p_j b_j$):

$$f(Y_j) = b_j^{Y_j} / Y_j! \sum_{t=0}^{n_j} e^{-b_j} t^{Y_j} \binom{n_j}{t} p_j^t (1-p_j)^{n_j-t}, \text{ for } Y_j = 0, 1, \dots$$

Gamma distribution with mean m_j and shape parameter k_j :

$$g(m_{ij}) = \frac{(k_j / m_j)^{k_j}}{\Gamma(k_j)} y_{ij}^{k_j-1} e^{-k_j y_{ij} / m_j}$$

Negative binomial distribution with mean m_j and shape parameter k_j :

$$f(y_{ij}) = \frac{\Gamma(k_j + y_{ij})}{\Gamma(k_j) y_{ij}!} \left(\frac{k_j}{m_j + k_j} \right)^{k_j} \left(\frac{m_j}{m_j + k_j} \right)^{y_{ij}}, \text{ for } y_{ij} = 0, 1, \dots$$

APPENDIX B:

Proof That a Mixture of Bernoulli Distributions is Still a Bernoulli Distribution

Suppose that the probability of admission for individual i in county j , m_{ij} , is a random variable with density $g(m_{ij})$. Define m_j to be $E(m_{ij})$. Conditional on m_{ij} , y_{ij} has a Bernoulli distribution with density

$$h(y_{ij} | m_{ij}) = m_{ij}^{y_{ij}} (1-m_{ij})^{(1-y_{ij})}$$

The marginal distribution of y_{ij} is

$$f(y_{ij}) = \int_0^{\infty} m_{ij}^{y_{ij}} (1-m_{ij})^{(1-y_{ij})} g(m_{ij}) dm_{ij}$$

If $y_{ij} = 1$, this is equal to

$$f(1) = \int_0^{\infty} m_{ij} g(m_{ij}) dm_{ij} = E(m_{ij}) = m_j.$$

If $y_{ij} = 0$, this is equal to

$$f(0) = \int_0^{\infty} (1-m_{ij}) g(m_{ij}) dm_{ij} = 1 - E(m_{ij}) = 1 - m_j.$$

Thus, in the absence of knowledge of m_{ij} , y_{ij} has a Bernoulli distribution with probability of admission m_j , regardless of what distribution $g(m_{ij})$ is.

APPENDIX C

Testing H_0 Using ANOVA

Suppose for a moment that y_{ij} is known for every individual. Then H_0 can be tested using a one-way analysis of variance (ANOVA) where county is the factor. Even though the y_{ij} s are not normally distributed, the F -test from the ANOVA will provide a valid test of H_0 as long as all of the Y_j s are approximately normally distributed. By the central limit theorem, the distribution of Y_j will approach a normal distribution as n_j goes to infinity. The F -test statistic is calculated as

$$F = MSC/MSE \tag{C1}$$

Where MSC stands for mean square due to counties and MSE stands for mean square due to error. MSC is defined as

$$MSC = \frac{J}{\sum_{j=1}^J} n_j (\hat{m}_j - \hat{m})^2 / (J - 1) \tag{C2}$$

where $\hat{m}_j = Y_j/n_j$ is the admission rate in county j . MSE is defined as

$$MSE = \frac{J}{\sum_{j=1}^J} \frac{n_j}{\sum_{i=1}^{n_j}} (y_{ij} - \hat{m}_j)^2 / (N - J) \tag{C3}$$

where $N = \sum_{j=1}^J n_j$ and is the total population size.

MSE is an estimate of the within-county variance of y_{ij} . The *F*-statistic calculated from Equation C1 can be compared to an *F*-distribution with $J - 1$ and $N - J$ degrees of freedom.

Now suppose that individual level data are not available so that it is not possible to use Equation C3 to calculate *MSE*. Suppose, however, we do know what *MAF* is. Then we can use *MAF* \hat{m} as an estimate of within-county variance.

Substituting this into the denominator of Equation C1 gives

$$\begin{aligned} F = MSC/MAF \hat{m} &= \frac{\sum_{j=1}^J n_j (\hat{m}_j - \hat{m})^2 / (J - 1)}{MAF \hat{m}} \\ &= \frac{\sum_{j=1}^J (n_j \hat{m}_j - n_j \hat{m})^2 / n_j \hat{m} (J - 1)}{MAF} \\ &= \frac{\sum_{j=1}^J (Y_j - n_j \hat{m})^2 / n_j \hat{m} (J - 1)}{MAF} \\ &= X_p^2 / MAF (J - 1), \end{aligned}$$

where X_p^2 is as defined in Equation 6.

From the properties of the chi-square and *F* distributions, we know that if X_p^2 / MAF has a chi-square distribution with $J - 1$ degrees of freedom, then $X_p^2 / MAF (J - 1)$ has an *F* distribution with $(J - 1)$ and infinity degrees of freedom. If $N - J$ is large enough to be considered infinity, (i.e., more than 100), then the *F*-test using *MAF* \hat{m} in place of *MSE* is virtually identical to the chi-square test adjusted for *MAF*.

Statistical analysis programs that allow specification of case-weights (e.g., SPSS, STATA) can be used to calculate *MSC*. This is done by specifying \hat{m}_j as the outcome variable and n_j as the caseweight so that the analysis is done as if there were n_j observations with identical outcome. The extension to a stratified analysis is simple in this context also—one merely uses a two-way ANOVA with stratum and county as factors (with no interaction term). The *MSE* estimated from such a model is meaningless and, as before, one uses *MAF* \hat{m} for *MSE*.

REFERENCES

- Breslow, N. "Extra-Poisson Variation in Loglinear Models." *Applied Statistics* 33, no. 1 (1984): 38-44.
- Brown, B. W., and M. Hollander. *Statistics: A Biomedical Introduction*. New York: John Wiley & Sons, Inc., 1977.
- Chassin, M. R., R. H. Brook, R. E. Park, J. Keeseey, A. Fink, J. Kosecoff, K. Kahn, N. Merrick, and D. H. Solomon. "Variations in the Use of Medical and Surgical Services by the Medicare Population." *New England Journal of Medicine* 314, no. 5 (30 January 1986): 285-90.
- Connell, F. A., R. W. Day, and J. P. LoGerfo. "Hospitalization of Medicaid Children: Analysis of Small Area Variations in Admission Rates." *American Journal of Public Health* 71, no. 6 (June 1981): 606-13.
- Copenhagen Collaborating Center. *CCC Bibliography on Regional Variations in Health Care*. Copenhagen: Vedbaek, Tekst og Tryk A/S, 1985.
- Cox, D. R., and D. V. Hinkley. *Theoretical Statistics*. London, England: Chapman and Hall, 1974.
- Diehr, P., K. Cain, F. Connell, and E. Volinn. "What Is Too Much Variation? The Null Hypothesis in Small-Area Analysis." *Health Services Research* 24, no. 6 (February 1990): 741-71.
- Diehr, P., and D. Grembowski. "A Small-Area Simulation Approach to Determining Excess Variation in Dental Procedure Rates." *American Journal of Public Health* 80, no. 11 (November 1990): 1343-48.
- Fleiss, J. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, Inc., 1981.
- Health Affairs*. "Special Issue on Medical Practice Variations." Vol. 3, no. 2 (1984).
- Johnson, N. L., and S. Kotz. *Discrete Distributions*. Boston: Houghton Mifflin, 1969.
- . *Continuous Univariate Distributions*. New York: John Wiley & Sons, Inc., 1970.
- Kazandjian, V., P. Durance, and M. Schork. "The Extremal Quotient in Small Area Variation Analysis." *Health Services Research* 24, no. 5 (1989): 665-84.
- Mantel, N., and W. Haenszel. "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease." *Journal of the National Cancer Institute* 22, no. 4 (April 1959): 719-48.
- Mauritsen, R. "Logistic Regression with Random Effects." Ph.D. diss., Department of Biostatistics, University of Washington, 1984.
- McPherson, K., J. Wennberg, O. Hovind, and P. Clifford. "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, no. 21 (18 November 1982): 1310-14.
- Pasley, B., P. Vernon, G. Gibson, M. McCauley, and J. Andoh. "Geographic Variations in Elderly Hospital and Surgical Discharge Rates, New York State." *American Journal of Public Health* 77, no. 6 (June 1987): 679-84.
- Paul-Shaheen, P., J. D. Clark, and D. Williams. "Small-Area Analysis: A Review and Analysis of the North American Literature." *Journal of Health Politics, Policy and Law* 12, no. 4 (1987): 741-809.

- Pocock, S. J., D. G. Cook, and S. A. A. Beresford. "Regression of Area Mortality Rates on Explanatory Variables: What Weighting Is Appropriate?" *Applied Statistics* 30, no. 3 (1981): 286-95.
- Tsutakawa, R. K. "Mixed Model for Analyzing Geographic Variability in Mortality Rates." *Journal of the American Statistical Association* 83, no. 401 (1988): 37-42.
- Wennberg, J. "Small Area Analysis: The Medical Care Outcome Problem." In AHCPR Conference Proceedings (Tucson, AZ, April 8-10, 1987). *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. Edited by L. Sechrest, E. Perrin, and J. Bunker. DHHS Publication no. (PHS) 90-3454. Washington, DC: U.S. Department of Health and Human Services, Public Health Service Agency for Health Care Policy and Research, 1990.
- Wolfe, R. A., J. R. Griffith, L. F. McMahon, P. S. Tedeschi, G. R. Petroni, and C. G. McLaughlin. "Patterns in Surgical and Non-surgical Hospital Use in Michigan Communities from 1980 through 1984." *Health Services Research* 24, no. 1 (April 1989): 67-82.
- Wong, G. Y., and W. M. Mason. "The Hierarchical Logistic Regression Model for Multilevel Analysis." *Journal of the American Statistical Association* 80, no. 391 (1985): 513-24.