Guidelines for Using Confidence Intervals for Public Health Assessment

Washington State Department of Health

July 13, 2012

For the other data analysis guidelines, see
http://www.doh.wa.gov/DataandStatisticalReports/DataGuidelines.aspx.

1 Purpose

The Assessment Operation Group in the Washington State Department of Health is coordinating the development of guidelines related to data development and use in order to promote good professional practice among staff involved in assessment activities within the Washington State Department of Health and in Local Health Jurisdictions in Washington. While the guidelines are intended for an audience of differing levels of training related to data development and use, they assume a basic knowledge of epidemiology and biostatistics. They are not intended to recreate basic texts and other sources of information related to the topics covered by the guidelines, but rather they focus on issues commonly encountered in public health practice and where applicable, to issues unique to Washington state.

2 Scope of these guidelines

These guidelines describe what confidence intervals are, and why and when they are used. We recommend methods for calculating confidence intervals in a few special circumstances that often arise in government public health work. However, a general description of how to calculate confidence intervals and formulae for calculating confidence intervals in a wide variety of situations are beyond the scope of these guidelines.

3 Basics

Confidence intervals provide a means of assessing and reporting the precision of a point estimate, such as a mortality or hospitalization rate or a frequency of reported behaviors. Confidence intervals account for the uncertainty that arises from the natural variation inherent in the world around us. In the special case of sample surveys, confidence intervals also account for the difference between a sample from a population and the population itself. Confidence intervals do not account for several other sources of uncertainty in point estimates, including missing or incomplete data or other data errors, or bias resulting from non-response or poor data collection. When confidence intervals are used to describe health data such as incidence or mortality rates, confidence levels of 95% are generally used (although 90% or 99% confidence intervals are not uncommon). Confidence intervals are sometimes used as a test of significance (see below).

What is a confidence interval?

A confidence interval is a range of values that is normally used to describe the uncertainty around a point estimate of a quantity, for example, a mortality rate. Therefore confidence intervals are a measure of the variability in the data. Generally speaking, confidence intervals describe how much different the point estimate could have been if the underlying conditions stayed the same, but chance had led to a different set of data. Confidence intervals are calculated with a stated probability (say 95%), and we say that there is a 95% chance that the confidence interval covers the true value. Most confidence intervals are calculated as 95% confidence intervals for the same reason that most statistical tests are done at the 0.05 level—in other words, only because it's conventional. It is completely arbitrary that we consider a result that would happen only 5 out of 100 times by chance as being statistically significant, while we consider one happening 6 out of 100 times as not being statistically significant. It is good to remember that the true population value is a constant, even though its value is unknown, but a confidence interval is a random quantity whose value depends on the random sample or data from which it is calculated. Therefore we describe a 95% (say) confidence interval as having a 95% probability of covering the true value, rather than saying that there is a 95% probability that the true value falls within the confidence interval.

When should confidence intervals be used?

Confidence intervals or p-values can be used whenever there is a need to describe the uncertainty in a point estimate. This is always the case when the estimate is derived from a sample. While confidence intervals may provide a less precise measure of statistical significance than p-values do, we recommend confidence intervals because they provide a better description of the range of possible values and are less subject to misinterpretation.

There are a few in public health who believe that confidence intervals should not be used around estimates derived from 'population' statistics such as the death rate in a given population, because they believe there is no statistical uncertainty in such estimates. This belief is contrary to the statistical theory underlying confidence intervals, and the biological and random processes governing the occurrence of events such as deaths and illnesses (Brillinger, 1986).

Confidence intervals as statistical tests

In a one sample case, as for example if one is comparing the age-adjusted rate for a particular county to a standard value, confidence intervals are equivalent to statistical tests. That is, if a 95% confidence interval around the county's age-adjusted rate excludes the comparison value, then a statistical test for the difference between the two values would be significant at the 0.05 level. It is tempting to use confidence intervals as statistical tests in two sample cases, for example, to say that if the confidence intervals around the age-adjusted rates in two counties overlap, then the rates are not significantly different, or vice versa. Although this may be a good approximation to a statistical test, it is not equivalent to one. When each confidence interval is constructed, it takes into account the sample size and variance in the one sample for which it is constructed. A proper statistical test for the difference between two samples will take into account the larger pooled sample size of the two samples together, and therefore provide a different result. This error is conservative, that is, in some cases an appropriate statistical test would indicate a statistically significant difference even though the confidence intervals do overlap, falsely implying no significant difference. However, if two confidence intervals do not overlap, a comparable statistical test would always indicate a statistically significant difference.

Standardized mortality or morbidity ratios (SMRs) should never be compared by assessing overlapping confidence intervals. An SMR for a particular population, say a county, is essentially an age-adjusted rate with the age distribution in that county being used as the standard. Therefore, the SMR for a different county is an age-adjusted rate with a different population as the standard. Because of this, SMRs should only be compared to the null value of 100, and not to each other (see the Guidelines for Using and Developing Rates for Public Health Assessment).

4 Recommended standards for specific situations

It is beyond the scope of these guidelines to recommend methods for calculating confidence intervals for the vast majority of situations in which they are needed.

However, we do describe how to calculate confidence intervals in several situations that commonly arise in government public health work. These include confidence intervals for age-adjusted rates, crude rates and age-specific rates, for SMRs, and for binomial proportions.

4.1 The normal approximation method

In many simple situations, especially those involving normally-distributed data, or large samples of data from other distributions, the normal approximation may be used to calculate the confidence interval. In this method, confidence intervals are given by

$$\hat{\mu} \pm z_{\alpha/2} \sqrt{\widehat{\mathrm{var}(\hat{\mu})}} \tag{1}$$

where μ is the parameter of interest (for example, a rate), $\hat{\mu}$ is its estimated value, $\widehat{\mathrm{var}(\hat{\mu})}$ is its estimated variance, and $z_{\alpha/2}$ is the $\alpha/2$-level normal deviate (e.g. 1.96 for 95% confidence intervals).

4.2 Age-adjusted rates

For a description of how and when to calculate age-adjusted rates, see the Guidelines for Using and Developing Rates for Public Health Assessment.

We recommend that confidence intervals for age-adjusted rates be calculated with the method based on the gamma distribution (Fay and Feuer, 1997). This method produces valid confidence intervals even when the number of cases is very small. When the number of cases is large these confidence intervals are equivalent to those produced with more traditional methods, as described by Chiang (1961) and Brillinger (1986).

Although the derivation of this method is based on the gamma distribution, the relationship between the gamma and χ^2 distributions allows the formulae to be expressed in terms of quantiles of the χ^2 distribution, which may be more convenient for computation.

Notation

Say the age-adjusted rates are calculated according to the following. Multiply the age-specific rates in the target population by the age distribution of the standard population:

$$\hat{R} = \sum_{i=1}^{m} s_i(d_i/P_i) = \sum_{i=1}^{m} w_i d_i \tag{2}$$

where m is the number of age groups, d_i is the number of deaths (or other events) in age group i, P_i is the population in age group i, and s_i is the proportion of the standard population in age group i. This is a weighted sum of Poisson random variables, with the weights being (s_i/P_i).

The variance is given by

$$v = \sum_{i=1}^{m} d_i(s_i/P_i)^2 \tag{3}$$

Confidence intervals

Then the confidence intervals are calculated according to these formulae:

$$\text{Lower Limit} = \frac{v}{2y}(\chi^2)^{-1}_{\frac{2y^2}{v}}(\alpha/2) \tag{4}$$

$$\text{Upper Limit} = \frac{v + w_M^2}{2(y + w_M)}(\chi^2)^{-1}_{\frac{2(y+w_M)^2}{v+w_M^2}}(1 - \alpha/2) \tag{5}$$

where y is the age-adjusted rate, v is the variance as calculated in equation 3, w_M is the maximum of the weights s_i/P_i, $1-\alpha$ is the confidence level desired (i.e. if 95% confidence intervals are needed, use $\alpha = 0.05$), and $(\chi^2)^{-1}_x$ is the inverse of the χ^2 distribution with x degrees of freedom. A fragment of SAS code that illustrates how to implement these two equations in SAS is available here.

4.3 SMRs

For standardized mortality or morbidity ratios (SMRs) we recommend two methods—one to be used for large numbers, where there are 100 or more observed cases, and another for smaller numbers, where there are less than 100 observed cases.

Notation

Say the SMR is given by $(O/E) \cdot 100$, where O is the number of observed cases and E is the number of expected cases.

Large numbers

For large numbers, we recommend the following method (Breslow and Day, 1987, p69).

$$\text{Lower Limit} = \left[1 - \frac{1}{9 \cdot O} - \frac{z_{\alpha/2}}{3\sqrt{O}}\right]^3 \frac{O}{E} \cdot 100 \tag{6}$$

$$\text{Upper Limit} = \left[1 - \frac{1}{9 \cdot (O+1)} + \frac{z_{\alpha/2}}{3\sqrt{O+1}}\right]^3 \frac{O+1}{E} \cdot 100 \tag{7}$$

where $z_{\alpha/2}$ denotes the $(1-\alpha/2)$-level standard normal deviate (e.g. use 1.96 for 95% confidence intervals).

Small numbers

If the number of observed cases is less than 100, we recommend that the confidence interval be calculated directly from the Poisson distribution. To do this, use the Poisson distribution to calculate a confidence interval for the observed number of cases, and then plug the upper and lower limits of that confidence interval into the standard formula for the SMR to obtain the confidence interval for the SMR. E.g. if LL is the lower limit and UL the upper limit for the confidence interval around the observed number of cases, then the confidence limits for the SMR are given by

$$\text{Lower Limit} = (LL/E) \cdot 100 \tag{8}$$

$$\text{Upper Limit} = (UL/E) \cdot 100 \tag{9}$$

A SAS macro (Daly, 1992) for computing these confidence limits is available here. For 95% confidence intervals, the upper and lower limits for the observed number may be taken from Table 1.

4.4 Crude and age-specific rates

Crude and age-specific rates are assumed to follow the Poisson distribution. Just as with SMRs, we recommend that the confidence intervals be calculated

Table 1: Poisson distribution 95% confidence limits.

Observed	Lower Limit	Upper Limit	Observed	Lower Limit	Upper Limit
0	0.0000	2.9957	50	37.1110	65.919
1	0.0253	5.5716	51	37.9728	67.056
2	0.2422	7.2247	52	38.8361	68.191
3	0.6187	8.7673	53	39.7006	69.325
4	1.0899	10.2416	54	40.5665	70.458
5	1.6235	11.6683	55	41.4335	71.590
6	2.2019	13.0595	56	42.3018	72.721
7	2.8144	14.4227	57	43.1712	73.850
8	3.4538	15.7632	58	44.0418	74.978
9	4.1154	17.0848	59	44.9135	76.106
10	4.7954	18.3904	60	45.7863	77.232
11	5.4912	19.6820	61	46.6602	78.357
12	6.2006	20.9616	62	47.5350	79.481
13	6.9220	22.2304	63	48.4109	80.604
14	7.6539	23.4896	64	49.2878	81.727
15	8.3954	24.7402	65	50.1656	82.848
16	9.1454	25.9830	66	51.0444	83.968
17	9.9031	27.2186	67	51.9241	85.088
18	10.6679	28.4478	68	52.8047	86.206
19	11.4392	29.6709	69	53.6861	87.324
20	12.2165	30.8884	70	54.5684	88.441
21	12.9993	32.1007	71	55.4516	89.557
22	13.7873	33.3083	72	56.3356	90.672
23	14.5800	34.5113	73	57.2203	91.787
24	15.3773	35.7101	74	58.1059	92.900
25	16.1787	36.9049	75	58.9923	94.013
26	16.9841	38.0960	76	59.8794	95.125
27	17.7932	39.2836	77	60.7672	96.237
28	18.6058	40.4678	78	61.6558	97.348
29	19.4218	41.6488	79	62.5450	98.458
30	20.2409	42.8269	80	63.4350	99.567
31	21.0630	44.0020	81	64.3257	100.676
32	21.8880	45.1745	82	65.2170	101.784
33	22.7157	46.3443	83	66.1090	102.891
34	23.5460	47.5116	84	67.0017	103.998
35	24.3788	48.6765	85	67.8950	105.104
36	25.2140	49.8392	86	68.7889	106.209
37	26.0514	50.9996	87	69.6834	107.314
38	26.8911	52.1580	88	70.5786	108.418
39	27.7328	53.3143	89	71.4743	109.522
40	28.5766	54.4686	90	72.3706	110.625
41	29.4223	55.6211	91	73.2675	111.728
42	30.2699	56.7718	92	74.1650	112.830
43	31.1193	57.9207	93	75.0630	113.931
44	31.9705	59.0679	94	75.9616	115.032
45	32.8233	60.2135	95	76.8607	116.133
46	33.6778	61.358	96	77.7603	117.232
47	34.5338	62.500	97	78.6605	118.332
48	35.3914	63.641	98	79.5611	119.431
49	36.2505	64.781	99	80.4623	120.529

directly from the Poisson distribution when the number of observed cases is less than 100 (see Table 1). When the number of cases is 100 or more, the normal approximation may be used to calculate the confidence intervals. This is

$$d/P \pm z_{\alpha/2}\sqrt{d}/P \tag{10}$$

where d is the number of deaths, P is the population, and $z_{\alpha/2}$ is the $\alpha/2$-level normal deviate (e.g. 1.96 for 95% confidence intervals).

Although the Ury-Wiggins approximation to the Poisson is more accurate than the normal approximation, it is harder to calculate, and the difference between the two is inconsequential when the number is cases is 100 or more.

4.5 Binomial proportions

We recommend the score interval (Vollset, 1993; Brown et al, 2001), which is found by solving the quadratic equation:

$$\frac{(x - n\hat{p}_*)^2}{n\hat{p}_*\hat{q}_*} = z^2_{\alpha/2} \tag{11}$$

where n is the sample size, x is the number of successes, $z_{\alpha/2}$ is the $\alpha/2$-level normal deviate (e.g. 1.96 for 95% intervals), and $\hat{p}_*$ is the confidence limit to be estimated.

The solution of the quadratic equation is:

$$\hat{p}_* = \frac{2nx + z^2_{\alpha/2}n \pm \sqrt{(2nx + z^2_{\alpha/2}n)^2 - 4(n^2 + z^2_{\alpha/2}n)x^2}}{2(n^2 + z^2_{\alpha/2}n)} \tag{12}$$

An Excel spreadsheet for calculating these confidence limits is available here. The spreadsheet was prepared by Alicia Thompson.

In the past, many analysts, including the authors of this document, have recommended computing "exact" confidence intervals directly from the binomial distribution when the sample size is small. However, exact confidence intervals tend to be conservative (too wide). Agresti and Coull (1998) have shown that the score interval given above works better in almost all circumstances than exact intervals, even for the smallest sample sizes. Therefore, we now recommend the score interval for all sample sizes.

4.6 Multiple admissions

Sometimes we want to estimate rates and confidence intervals in situations where the assumption of independence between events does not hold. For example, we may want to measure hospital admission rates. For some conditions, such as asthma, a few people may be hospitalized many times. The multiple admissions for an individual person are not likely to be independent of each other, in the sense that a person who is once hospitalized for asthma is more likely to be hospitalized later for asthma than is a person who has not been hospitalized for asthma. Therefore, the total count of admissions may not follow a Poisson distribution. It is typical in such situations for the total count to exhibit greater variability than it would have if it were Poisson (hence the term *extra-Poisson variation* is often used). Because of this, if the methods described elsewhere in this document are applied to hospital admission rates (whether these are in the form of age-adjusted, age-specific, or crude rates, or SMRs) they may produce confidence intervals that are too narrow.

Several statistical methods are available for analyzing data that has extra-Poisson variation, including generalized estimating equations (GEE) and other quasi-likelihood models, and the bootstrap. Analysts who have the knowledge and computer software to use those methods should do so when appropriate.

Here we describe how to calculate confidence intervals for age-adjusted hospital admission rates. This method was described by Carriere and Roos (1994) and by Stukel et al. (1994). In its basic principles, this method is similar to using Multiple Admission Factors, as proposed by Cain and Diehr (1992), or the negative binomial distribution, as described by Glynn et al (1993). We wish to caution users that this method may not work well for small numbers (e.g. $n < 50$). We have no alternative method to recommend for small numbers.

4.6.1 Age-adjusted rates

Notation

The notation is similar to that used in Section 4.2 on age-adjusted rates. Say the age-adjusted rates are calculated according to the following. Multiply the age-specific rates in the target population by the age distribution of the standard population:

$$\hat{R} = \sum_{i=1}^{m} s_i(d_i/P_i) = \sum_{i=1}^{m} s_i h_i \qquad (13)$$

where m is the number of age groups, d_i is the number of hospitalizations in age group i, P_i is the population in age group i, s_i is the proportion of the standard population in age group i, and h_i is the age-specific hospitalization rate in age group i.

Variance

The variance of h_i is estimated by

$$\widehat{\mathrm{var}(h_i)} = \sum_{j=1}^{P_i} (d_{ji} - h_i)^2 / P_i(P_i - 1) \tag{14}$$

where d_{ji} is the number of hospital admissions for individual j in age group i. Some algebraic manipulation can be used to rewrite this in a form that is easier for computation:

$$\widehat{\mathrm{var}(h_i)} = \frac{(\sum_{j=1}^{P_i} d_{ji}^2) - P_i h_i^2}{P_i(P_i - 1)} \tag{15}$$

In this form the summation only needs to be performed over the people in the population who have at least one hospital admission, since $d_{ji} = 0$ for people who are not hospitalized, and they make no contribution to the sum.

Then the variance of the age-adjusted hospital admission rate is estimated by:

$$\widehat{\mathrm{var}(\hat{R})} = \sum_{i=1}^{m} s_i^2 \frac{(\sum_{j=1}^{P_i} d_{ji}^2) - P_i h_i^2}{P_i(P_i - 1)} \tag{16}$$

Confidence intervals

Finally, confidence intervals can be calculated with the usual normal approximation method as:

$$\hat{R} \pm z_{\alpha/2} \sqrt{\widehat{\mathrm{var}(\hat{R})}} \tag{17}$$

where $z_{\alpha/2}$ is the $\alpha/2$-level normal deviate (e.g. 1.96 for 95% confidence intervals).

4.6.2 Crude and age-specific rates

For crude or age-specific rates, the rate is given by

$$\hat{R} = d/P \tag{18}$$

where d is the number of hospitalizations and P is the population.

Then the variance of the rate is given by

$$\widehat{\mathrm{var}}(\hat{R}) = \frac{(\sum_{j=1}^{P} d_j^2) - d^2/P}{P(P-1)} \tag{19}$$

where d_j is the number of hospital admissions for individual j. The summation only needs to be performed over the people in the population who have at least one hospital admission, since $d_j = 0$ for people who are not hospitalized, and they make no contribution to the sum.

Then confidence limits may be obtained with the normal theory method, as in equation 17.

4.6.3 SMRs

Confidence intervals for SMRs may be derived in a way similar to that used for age-adjusted rates (see papers by Carriere and Roos (1994) and Stukel et al (1994)).

4.7 Complex survey sample designs

Surveys with complex sample designs are surveys which include one or more of these features: stratification, clustering of observations, or unequal weighting of observations. Computation of confidence intervals from complex survey samples must account for the design effect of the survey. Depending on the sampling design, this may require the use of software designed to analyze data from complex survey samples, such as SUDAAN, STATA, or the new survey analysis modules in SAS version 8.

References

Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* **52**:119-126, 1998.

Breslow NE, Day NE. Statistical Methods in Cancer Research: Volume II—The Design and Analysis of Cohort Studies. Lyon: International Agency for Research on Cancer, 1987.

Brillinger DR. The natural variability of vital rates and associated statistics [with discussion]. *Biometrics* **42**:693-734, 1986.

Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statistical Science* **16**:101-133, 2001.

Cain KC, Diehr P. Testing the null hypothesis in small area analysis. *Health Serv Res* **27**:267-294, 1992.

Carriere KC, Roos LL. Comparing standardized rates of events. *Am J Epidemiol* **140**:472-482, 1994.

Chiang CL. Standard error of the age-adjusted death rate. *Vital Statistics, Special Reports* **47**:271-285, USDHEW, 1961.

Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. *Comput Biol Med* **22**:351-361, 1992.

Fay MP, Feuer EJ. Confidence intervals for directly adjusted rates: a method based on the gamma distribution. *Stat Med* **16**:791-801, 1997

Glynn RJ, Stukel TA, Sharp SM, Bubolz TA, Freeman JL, Fisher ES. Estimating the variance of standardized rates of recurrent events, with application to hospitalizations among the elderly in New England. *Am J Epidemiol* **137**:776-786, 1993.

Stukel TA, Glynn RJ, Fisher ES, Sharp SM, Lu-Yao G, Wennberg JE. Standardized rates of recurrent outcomes. *Stat Med* **13**:1781-1791, 1994.

Vollset SE. Confidence intervals for a binomial proportion. *Stat Med* **12**:809-824, 1993