Much Ado About Almost Nothing: Methods for Dealing With Limited Data

Stephen W. Looney
Dept. of Biostatistics, Georgia Health Sciences University, Augusta, GA 30912

Courtney E. McCracken
Dept. of Pediatrics, Emory University, Atlanta, GA 30329

Abstract

Applied statisticians are often confronted with statistical inference problems dealing with situations in which there appear to be no data, or data of only limited usefulness. For example, when attempting to find a confidence interval for a binomial proportion, the sample may contain no successes. Such a scenario could be encountered when attempting to estimate the incidence of an extremely rare side effect associated with the administration of a newly developed drug. Other statistical inference situations in which there may be no or only limited data include estimating an odds ratio when one of the cells in a 2x2 table is empty, estimating a risk ratio when one of the groups experiences none of the outcome of interest, and incorporating observations below the limit of detection into a statistical analysis. In this presentation, we illustrate each of these scenarios with real data and describe the preferred methods for handling them.

Keywords: relative risk, exact methods, limit of detection, Spearman correlation, binomial distribution, odds ratio

1. Introduction

First of all, we must apologize to Dennis Helsel for "stealing" part of the title of our presentation from him. He is the author of *Nondetects and Data Analysis* (Helsel 2005), which we consider to be the definitive text on the analysis of data that include observations below the limit of detection (LOD) of a measuring device. (We will refer to such observations as "nondetects," or ND's.) In 2010, he published a commentary entitled "Much Ado About Next to Nothing: Incorporating Nondetects in Science" (Helsel 2010), which addresses many of the issues that we will discuss today. (Of course, we also owe a debt of gratitude to the Bard of Avon for our title.)

In today's presentation, we address an issue that applied statisticians are often confronted with; namely, statistical inference problems in which there appear to be no data, or data of only limited usefulness. We will illustrate the following scenarios with real data and describe methods for handling them: (1) Finding a confidence interval (C.I.) for a binomial proportion when the sample contains no successes (or failures), (2) Finding a C.I. for an odds ratio (OR) when one of the cells in the 2x2 table is empty, (3) Finding a C.I. for a risk ratio (RR) when one of the groups experiences none of the outcome of interest, and (4) Incorporating ND's into a correlation analysis.

2. Binomial with No Successes

This scenario could occur, for example, when we wish to estimate the incidence of an extremely rare side effect associated with the administration of a drug or use of a device. For example, one of the authors (SL) was involved in a clinical study in which a dentist

wanted to examine the effects of electromagnetic interference on a neurostimulator during the operation of three dental devices (electric pulp tester, apex locator, electrocautery unit). After n = 70 independent trials, there were no failures of the neurostimulators. The researcher wanted to know: "What is a reasonable upper bound for the probability of failure of the neurostimulator?" If one formulates this in the context of the binomial distribution, letting π = the probability of failure, then the problem becomes one of finding the upper limit of a 1-sided 95% C.I. for a binomial proportion: $(0, \pi_u)$. The most commonly used method, the Wald interval, yields uninformative results: $\hat{\pi} = 0,\ ASE(\hat{\pi}) = 0,\ \pi_u = 0$.

We evaluated 8 methods that can be used to find confidence limits for a binomial proportion: (1) the Mid-P interval (Agresti 2007, p. 16), (2) the Wilson (score) interval (Wilson 1927), (3) the SAIFS interval (Borkowf 2006), (4) the Bayes-Laplace HPD (Tuyl, Gerlach, and Mengerson 2008), (5) the Clopper-Pearson (C-P) exact interval (Clopper and Pearson 1934), (6) the Poisson interval (Leemis and Trivedi 1996), (7) the continuity-corrected Wilson interval (Casella 2001), and (8) the Agresti-Coull interval (Agresti and Coull 1998).

The upper 95% C.I. for the observed sample of n = 70, x = 0, yielded the following results for each of these methods:

Method	**π_u = Upper Limit of 95% C.I. (π)**
Mid-P	0.032
Wilson (Score)	0.037
SAIFS	0.038
Bayes-Laplace HPD	0.041
C-P Exact	0.042
Poisson	0.043
Continuity-Corrected Wilson	0.050
Agresti-Coull	0.058

Compared to the other methods we evaluated, C.I.'s based on the Clopper-Pearson method are preferred when x = 0 for the following reasons: (1) they are moderate in length, (2) the required sample size is comparable to other methods, (3) they are easy to compute, (4) there are closed-form expressions for π_u and the sample size required to yield 1-sided C.I. of any length, (5) they are equivalent to a Bayesian prediction interval based on the Jeffreys prior, and (6) they always have p-confidence (Vos and Hudson

2005) equal to the nominal confidence coefficient. See McCracken and Looney (2011) for full details.

3. Odds Ratio with Empty Cells

If any of the cells of a 2x2 table are 0, then either $\widehat{OR} = 0$ or $\widehat{OR} = \infty$, depending on whether the 0 cell occurs in the main diagonal or the off-diagonal, respectively. Furthermore, $\text{ASE}[\log(\widehat{OR})] = \infty$ if any of the cells in the 2x2 table are zero, where "ASE(·)" will denote "approximate standard error" throughout the remainder of this manuscript. For a 2x2 table with a zero in one of the cells, Agresti (2007, p. 31) recommends the use of the *slightly amended* estimator of the OR:

$$\widehat{OR} = \frac{(n_{11} + .5)(n_{22} + .5)}{(n_{12} + .5)(n_{21} + .5)}.$$

The $\text{ASE}[\log(\widehat{OR})]$ is then calculated "in the usual way" after adding .5 to each cell count. In a study of the association between positive toxicology screens and clinical outcome among patients hospitalized with traumatic injuries (Blondell et al. 2005), the following 2x2 table was obtained.

Table 1. 2x2 Table Showing Association Between Positive Screen for Cocaine and Risk of Death Following Traumatic Injury

		Death	No Death
Cocaine	Positive	0	110
	Negative	30	739

For the data in Table 1, $\widehat{OR} = 0$ and $\text{ASE}[\log(\widehat{OR})] = \infty$. The "Agresti method" is based on the following amended table:

Table 2. Amended Table Corresponding to Table 1

		Death	No Death
Cocaine	Positive	0.5	110.5
	Negative	30.5	739.5

Using the data in Table 2, we obtain: $\widehat{OR} = 0.11$, with an approximate 95% CI(*OR*) of (0.01, 1.81). Exact methods are also available for estimating the true *OR* when there are empty cells. For the data in Table 1, StatXact yields $\widehat{OR} = 0$, with an exact 95% CI(*OR*) of (0.00, 0.72). Note that the conclusions differ for the 2 intervals in terms of the test of H_0: *OR* = 1, where *OR* denotes the true value of the odds ratio.

4. Risk Ratio with No Events

In a study of the effectiveness of a behavioral intervention for depression in nursing homes (Meeks et al. 2008), the following 2x2 table was obtained:

Table 3. Table Showing Association Between Risk of Fall and Intervention for Depression

	1 or More Falls	No Falls	Total
Intervention	5	24	29
Control	0	24	24

For the data in Table 3, $\widehat{RR} = \infty$ and $\text{ASE}[\log(\widehat{RR})] = \infty$. The method of Agresti & Caffo (2000) that can be used for statistical inference for the risk difference when zero events are observed in one of the groups can be adapted to inference for the *RR*. This method is based on applying the "usual" formula for calculating $\text{ASE}[\log(\widehat{RR})]$ to the data in the following "amended" table, obtained by adding 1 to each cell:

Table 4. Amended Table Corresponding to Table 3

	1 or More Falls	No Falls	Total
Intervention	6	25	31
Control	1	25	26

Using the data in Table 4, we obtain: $\widehat{RR} = 5.3$, with an approximate 95% CI(*RR*) of (0.65, 39.16). Using StatXact, the following results are obtained: $\widehat{RR} = \infty$, with an exact 95% CI(*RR*) of (1.14, ∞). In a recent article, Carter et al. (2010) proposed using the ratio of the median unbiased estimates (MUE) of the proportions in each treatment group to obtain the point estimate of the risk ratio and the deterministic bootstrap to find an approximate C.I.(*RR*). For the data in Table 4, $\widehat{RR} = 12.38$, with an approx. 95% CI(*RR*) of (1.52, 23.44). Note that the conclusions differ for the 3 intervals in terms of the test of H_0: *RR* = 1, where *RR* denotes the true value of the risk ratio:

Method	**Point Estimate**	**95% C.I. (*RR*)**	**Conclusion**
Amended	5.3	(0.65 – 39.16)	Fail to reject
Exact	∞	(1.14, ∞)	Reject
Carter et al.	12.38	(1.52, 23.44)	Reject

5. Non-Detects

Looney and Hagan (2005) considered the correlation between hippuric acid (Y) and *ortho*-Cresol (X) concentrations in urine samples of 54 individuals exposed to toluene using data taken from Amorim and Alvarez-Leite (1997). Both X and Y were subject to ND's; out of the n = 54 pairs of (X,Y) observations, there were 15 subjects with $X \geq \text{LOD}_x$ and $Y \geq \text{LOD}_y$, 35 subjects with $X < \text{LOD}_x$ and $Y \geq \text{LOD}_y$, and 4 subjects with $X < \text{LOD}_x$ and $Y < \text{LOD}_y$. In their analysis of the data, Amorim and Alvarez-Leite replaced the 39 ND's with the LOD's for X and Y, and then performed inference for the Pearson correlation in the usual way.

In most statistical analyses of data subject to a limit of detection, the ND's are treated as left-censored at the LOD. If both X and Y are subject to left-censoring, the general maximum likelihood (ML) approach due to Lyles et al. (2001) can be used. Other approaches that have been used include: (1) simple substitution, in which each ND is replaced by zero, the LOD, or some simple function of the LOD, such at LOD/2; (2) assuming that the ND's are tied at the LOD, and then calculating Spearman's or Kendall's correlation coefficient adjusted for ties; or (3) analyzing only the complete cases (i.e., only those pairs for which both $X \geq \text{LOD}_x$ and $Y \geq \text{LOD}_y$. It is important to note that ND's *cannot* be assumed to be missing at random; thus, missing-data methods available in PROC MIXED, for example, do not apply.

Figure 1 contains a scatterplot of the Amorim and Alvarez-Leite data, and Table 5 contains a summary of the results of various methods that were used to estimate the correlation between the hippuric acid and *ortho*-Cresol concentrations.

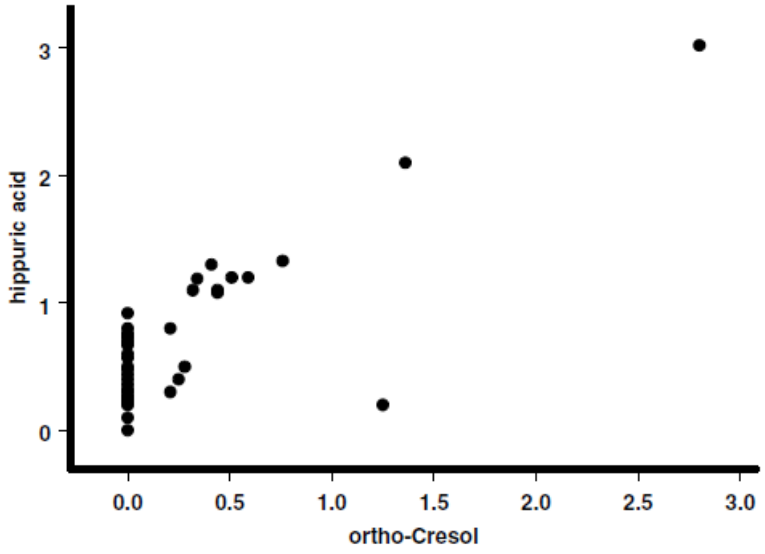

Figure 1. Scatterplot of *ortho*-Cresol vs. hippuric acid concentrations in urine samples of 54 individuals exposed to toluene in shoe factories, painting sectors of metal industries, and printing shops. Observations below the detectable limit of either assay are plotted as zero for purposes of illustration.

Table 5. Comparison of Estimation Results Based the Data of Amorim and Alvarez-Leite

Method	n	$\hat{\rho}$	95% C.I.	Width
Impute zero	54	0.79	(0.67, 0.88)	0.21
Impute LOD/2	54	0.79	(0.65, 0.87)	0.22
Impute LOD	54	0.79	(0.63, 0.86)	0.23
Lyles et al. (2001)	54	0.79	(0.62, 0.88)	0.26
Spearman	54	0.58	(0.36, 0.74)	0.38
Complete cases	15	0.76	(0.40, 0.92)	0.52
Kendall	54	0.49	(0.12, 0.74)	0.62

In her PhD dissertation research, the 2nd author (C. McCracken) is performing a comprehensive comparison of various methods for dealing with ND's when estimating a correlation. The methods she is comparing include (but are not limited to) (1) Pearson's correlation with non-detects set equal to LOD, $LOD/2$ or $LOD/\sqrt{2}$, (2) other imputation methods proposed in the literature; (3) Pearson's correlation estimated using the ML method described by Lyles et al. (2001); (4) Spearman's correlation with non-detects assumed to be tied at LOD; and (5) Kendall's tau with non-detects assumed to be tied at LOD. She is comparing these estimation methods under conditions of large censoring proportions, unbalanced censoring proportions, small to moderate n, and non-BVN data.

As of the date of this presentation (July 29, 2012), Ms. McCracken had obtained results for two sample sizes (n = 100 and 200); two true values of the correlation (ρ = 0.25 and 0.50); the following censoring proportions in (X,Y): (20, 20), (40, 20), (40, 40), (60, 40), (60, 60), and (80, 20); and three distributions [bivariate normal (BVN), bivariate gamma (BVG), bivariate beta (BVB)]. These particular non-bivariate normal distributions were chosen to represent a wide range of departures from BVN, as measured by multivariate skewness (β_{1p}) and multivariate kurtosis (β_{2p}). For the BVN, (β_{1p}, β_{2p}) = (0, 8). For the BVG alternative, (β_{1p}, β_{2p}) = (1, 9) and, for the BVB, the two versions included in the simulation have (β_{1p}, β_{2p})=(3, 10) and (4, 12), respectively. A sample of Ms. McCracken's simulation results for the ML method is summarized in Table 6.

The results in Table 6 indicate that the ML method performs well for moderate to large censoring proportions and moderate departures from BVN. However, with more severe departures from BVN, as in the case of the BVB distribution, no method (including ML) performed acceptably. None of the simple substitution methods (LOD/2, etc.) ever reached coverage probability above 92% for a 95% confidence interval.

Table 6. Empirical Coverage Probabilities of 95% Confidence Intervals Based on the Maximum Likelihood Method

Distribution	Censoring Proportion	Coverage Probability
Bivariate Normal	(80, 20)	95.1
	(40, 40)	94.9
	(60, 60)	94.5
Bivariate Gamma	(80, 20)	94.9
	(40, 40)	94.5
	(60, 60)	94.4
Bivariate Beta	(80, 20)	87.8
	(40, 40)	89.4
	(60, 60)	75.4

6. Conclusion

Even if there are "no" data or extremely limited data, valid methods are available, even when the sample size is quite small. These methods may require specialized software.

References

Agresti, A. (2007), *An Introduction to Categorical Data Analysis*, 2nd Edition, Hoboken, NJ: Wiley.

Agresti, A., and Coull, B. (1998), "Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119-126.

Amorim, L.C.A., and Alvarez-Leite, E.M. (1997). "Determination of *o*-cresol by gas chromatography and comparison with hippuric acid levels in urine samples of individuals exposed to toluene," *Journal of Toxicology and Environmental Health*, 50, 401-407.

Blondell, R.D., Dodds, H.N., Looney, S.W., Lewis, C.M., Hagan, J.L., Lukan, J.K., and Servoss, T.J. (2005), "Toxicology Screening Results: Injury Associations Among Hospitalized Trauma Patients," *Journal of Trauma,* 58, 561-570.

Borkowf, C.B. (2006), “Constructing Binomial Confidence Intervals with Near Nominal Coverage by Adding a Single Imaginary Failure or Success,” *Statistics in Medicine*, 25, 3679-3695.

Carter, R.E., Yan, L., Lipsitz, S.R., Newcombe, R.G., and Hermayer, K.L. (2010). "Relative Risk Estimated From The Ratio Of Two Median Unbiased Estimates," *Applied Statistics*, 59, 657 - 671.

Casella, G. (2001), Comment on “Interval Estimation for a Binomial Proportion,” *Statistical Science*, 16, 120-122.

Clopper, C.J., and Pearson, E.S. (1934), “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial,” *Biometrika*, 26, 404-413.

Helsel, D. R. (2005). *Nondetects and Data Analysis.* New York: John Wiley & Sons, Inc.

Helsel, D.R. (2010). "Much Ado About Next to Nothing: Incorporating Nondetects in Science," *Annals of Occupational Hygiene*, 54, 257- 262.

Leemis, L.M., and Trivedi, K.S. (1996), “A Comparison of Approximate Interval Estimators for the Bernoulli Parameter,” *The American Statistician,* 50, 63-68.

Looney, S.W. and Hagan, J.L. (2006), “On Methods for Handling Biomarker Data below the Analytic Limit of Detection,” *Proceedings of the American Statistical Association*, pp. 2477-2481, presented at the 2006 Joint Statistical Meetings, Minneapolis, MN.

Lyles, R.H., Williams, J.K., and Chuachoowong, R. (2001). Correlating two viral load assays with known detection limits. *Biometrics* 57: 1238-1244.

McCracken, C.E. "Correlation Coefficient Inference for Left-censored Biomarker Data with Known Detection Limits," doctoral dissertation, Georgia Health Sciences University Department of Biostatistics, in progress.

McCracken, C.E., and Looney, S.W. (2011),"A Comparison of Methods for Finding the Upper Confidence Limit for a Binomial Proportion When Zero Successes Are Observed," *Proceedings of the American Statistical Association*, pp. 2876-2890, presented at 2011 Joint Statistical Meetings, Miami Beach, FL.

Meeks, S., Looney, S.W., Van Haitsma, K., and Teri, L. (2008). "BE-ACTIV: A Staff-Assisted, Behavioral Intervention for Depression in Nursing Homes," *The Gerontologist*, 48(1), 105-114.

Tuyl, F., Gerlach, R., and Mengerson, K. (2008), “A Comparison of Bayes-Laplace, Jeffreys, and Other Priors: The Case of Zero Events,” *The American Statistician*, 62, 40-44.

Vos, P., and Hudson, S. (2005), “Evaluation Criteria for Discrete Confidence Intervals: Beyond Length and Coverage,” *The American Statistician*, 59, 137-142.

Wilson, E.B. (1927), “Probable Inference, the Law of Succession, and Statistical Inference,” *Journal of the American Statistical Association,* 22, 209-212.