Biostatistics (2010), **11**, 2, *pp.* 373–374
doi:10.1093/biostatistics/kxp050
Advance Access publication on November, 30 2009

Confidence intervals that match Fisher's exact or Blaker's exact tests

MICHAEL P. FAY
Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892-7609, USA
mfay@niaid.nih.gov

SUMMARY

When analyzing a 2×2 table, the two-sided Fisher's exact test and the usual exact confidence interval (CI) for the odds ratio may give conflicting inferences; for example, the test rejects but the associated CI contains an odds ratio of 1. The problem is that the usual exact CI is the inversion of the test that rejects if either of the one-sided Fisher's exact tests rejects at half the nominal significance level. Further, the confidence set that is the inversion of the usual two-sided Fisher's exact test may not be an interval, so following Blaker (2000, Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics* **28**, 783–798), we define the "matching" interval as the smallest interval that contains the confidence set. We explore these 2 versions of Fisher's exact test as well as an exact test suggested by Blaker (2000) and provide the R package *exact2×2* which automatically assigns the appropriate matching interval to each of the 3 exact tests.

Keywords: Conditional Exact Test; Confidence Set; Fisher's Exact Test; Odds Ratio; Two-by-Two Table.

1. MOTIVATING EXAMPLE

Lim *and others* (2009) explore whether being homozygous for the CCR5Δ32 mutation, which causes complete loss-of-function in the chemokine receptor CCR5, effects the probability of having clinical symptoms given infection with West Nile virus. They test for significance of a genetic recessive model on 16 specific symptoms, giving both the two-sided Fisher's exact test p-values and the 95% CIs on the odds ratios based on the asymptotic normality of the log transformation. A test–CI inconsistency occurs for the symptom tremors since the two-sided Fisher's exact test was not significant at the 0.05 level but the 95% CI excluded 1. One might expect better test–CI consistency from using the exact conditional tail interval (ECTI) derived from the one-sided Fisher's exact test which is the only exact CI for odds ratios available from standard software (see, e.g. SAS version 9.2, StatXact 8 Procs, or R version 2.10.0). However, if we use these ECTIs for all 16 symptoms, then 2 other symptoms have the test–CI inconsistency problem in the opposite direction: they are significant by two-sided Fisher's exact test but are not significant by the ECTI. For example, $4/15 = 26.7\%$ of the homozygous subjects and $50/619 = 8.1\%$ of the others had abdominal pain, giving a two-sided Fisher's exact p-value of $p = 0.032$ but 95% ECTI of (0.92, 14.58).

2. MATCHING CONFIDENCE INTERVALS

For a 2×2 table, essentially all exact conditional one-sided tests are equivalent to the one-sided Fisher's exact test, but for the two-sided tests, there are 3 exact conditional tests: (i) the two-sided Fisher's exact test defines the p-value as the sum of all probabilities equal to or less than that of the observed table, (ii) another version of a two-sided Fisher's exact test, which we call the "central" Fisher's exact test, defines the p-value as the minimum of 1 and twice the smallest of the one-sided p-values, and (iii) Blaker's (2000) exact test defines the p-value as the sum of the observed minimum one-sided tail probability and the largest tail probability on the other side that is not larger than the observed one. Using the noncentral hypergeometric distribution, each of these 3 tests may be generalized to test that the null equals any specific odds ratio; therefore, each test can be inverted to create its own $100(1-\alpha)\%$ confidence set which contains all null odds ratios whose corresponding test fails to reject at the α level.

The inversion of the central Fisher's exact test is the ECTI which is easily calculated; however, for the two-sided Fisher's exact or Blaker's exact test, the corresponding inversion confidence set may be the union of 2 disjoint intervals. Blaker (2000) suggested defining a CI, which we call the matching CI of the test, as the smallest interval that contains all members of the inversion. Blaker (2000) suggested a simple algorithm for calculating matching CIs, but the bounds on the precision from algorithms of that type are not clear. In the supplementary material at *Biostatistics* online, we review the ECTI and the 3 tests, give details on the algorithmic precision problem, give a new algorithm for the matching CIs, show that the test–CI inconsistencies of the motivating example are not rare, and compare the 3 tests. For the abdominal pain example, the 3 p-values and 95% matching CIs calculated from our *exact2×2* R package are: two-sided Fisher's exact, $p = 0.032$, (1.17, 14.17); central Fisher's exact, $p = 0.063$, (0.92, 14.58); and Blaker's exact, $p = 0.032$, (1.17, 14.22).

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

The author thanks Dean Follmann for providing the data example and Mike Proschan for discussions on the noncentral hypergeometric distribution. *Conflict of Interest:* None declared.

REFERENCES

BLAKER, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics* **28**, 783–798.

LIM, J. K., MCDERMOTT, D. H., LISCO, A., FOSTER, G. A., KRYSZTOF, D., FOLLMANN, D., STRAMER, S. L., MURPHY, P. M. (2009). CCR5 deficiency is a risk factor for early clinical manifestations of West Nile virus infection, but not for infection per se. *Journal of Infectious Diseases* **201**, pp. 178–185.

[*Received September 4, 2009; revised September 21, 2009; accepted for publication October 27, 2009*]