BACKGROUND: Health economic evaluation systematically analyzes all inputs and outputs of treatment, and suggests the most effective alternatives for given resources. Results of a cost-effectiveness analysis are summarized in a series of cost-effectiveness (C/E) ratios. Little is known about the distributions of C/E ratio estimates. Hence, there is a lack of statistical tests for the C/E ratios.

OBJECTIVE: This study is to describe the characteristics of C/E ratio distribution. Since the cost and effect distributions frequently follow the normal, lognormal, and gamma distributions, the ratio distributions are formed by some combination of them. METHODS: In describing the ratio distribution, the probability density functions (PDF) of the ratio distributions are derived if they exist. If the closed form of the PDF does not exist, the ratio distributions are presented graphically. RESULTS: First, the ratio distributions take on a variety of shapes depending on the coefficient of variation of their denominator distribution. Most of the time, the ratio distributions have the bell shape with, or without, a heavy right tail. However, the ratio distribution could even be bimodal if the coefficient of variation of denominator was very large. Second, the correlation between numerator and denominator of the C/E ratio significantly affects the distribution shapes and locations. Since the numerator and denominator are allowed to be correlated, this derived PDF is more general and practical than the Cauchy distribution. Third, the ratio distributions formed from the combination of gamma and lognormal distributions are all skewed to the right. CONCLUSIONS: The C/E ratio distributions are not always bell shape nor symmetric. The ratio distributions take on a variety of shapes depending on the coefficient of variation of their denominator distribution and the correlation coefficients between numerator and denominators.

COMPARISON OF STATISTICAL TESTS FOR THE COST/EFFECTIVENESS RATIOS
Kim S, Moeschberger M, Pathak D
The Ohio State University, Columbus, OH, USA

BACKGROUND: Results of a cost-effectiveness analysis are summarized in a series of cost-effectiveness (C/E) ratios. As with other statistics, the C/E ratio is subject to sampling variation. However, constructing a confidence interval for the cost-effectiveness ratio is complicated because both numerator and denominator of the ratio are stochastic in nature. Several statistical methods have been proposed lately, yet, the systematic method of handling uncertainty is generally underdeveloped in economic evaluation. OBJECTIVE: This study is to compare the statistical methods proposed in constructing confidence intervals for the various ratio distributions, for the various correlation coefficients between numerator and denominator, for the various coefficient of variations, and the various sample sizes. METHODS: The ratio distributions are formed from the combinations of normal, lognormal, and gamma distributions, which frequently arise in cost-effective studies. In evaluation the performance of statistical methods, a simulation study was conducted for the various ratio distributions. For each sample, the confidence intervals were constructed by five statistical procedures; the Box, Taylor’s, Fieller’s, bootstrap, and jackknife methods. RESULTS: First, since the ratio distributions are largely dependent on the distribution of its denominator, none of the statistical tests work if the mean of denominator is close to zero. Second, if the sample size is small, none of the statistical tests perform well enough regardless of correlation and coefficient of variation. Third, for the large sample size, all methods, except the box method, constructed the confidence interval well. Among them, Fieller’s method is the first choice of selection for the estimation of the confidence interval. CONCLUSION: None of the statistical tests work if the mean of denominator is close to zero. This is problematic for Incremental Cost-effectiveness Ratio (ICER). If the “net effect” of the new procedure is insignificant, then the statistical test for the ICER is not stable.

RELATIONSHIP BETWEEN NONPARAMETRIC RECEIVER OPERATING CHARACTERISTIC ANALYSIS AND A LIKELIHOOD-RATIO TEST FOR MODEL SELECTION: I. A MONTE CARLO SIMULATION USING CONTINUOUS DATA
Shaw JW, Horrace W, Coons SJ
The University of Arizona, Tucson, AZ, USA

BACKGROUND: Receiver operating characteristic (ROC) analysis is frequently used to assess the accuracy of diagnostic tests. The area under an ROC curve (AUC) is indicative of the extent to which a measure correctly classifies true-positive and true-negative subjects. DeLong et al. (Biometrics 1988; 44: 837–845) have proposed a method for comparing nonparametric ROC curves derived from the same set of cases. In a small study using data from a naturalistic investigation, DeLong’s test yielded results that were consistent with those of a likelihood-ratio test for model selection developed by Vuong
MODEL LL using Vuong’s test. The random number seed was set so that identical samples were compared with each test. RESULTS: In general, the two tests yielded similar statistical conclusions. The observed power of Vuong’s test was slightly less than that of DeLong’s test. There were only two cases in which the tests yielded different results, and these occurred in small samples. CONCLUSIONS: Though slightly less powerful than DeLong’s test, Vuong’s test may be applied in cases where the dependent variable has more than two levels. It is also easier and takes much less time to perform than DeLong’s test. Given these advantages, Vuong’s test may be preferred to ROC-based tests in larger samples.

PMA13

RELATIONSHIP BETWEEN NONPARAMETRIC RECEIVER OPERATING CHARACTERISTIC ANALYSIS AND A LIKELIHOOD-RATIO TEST FOR MODEL SELECTION: II. A MONTE CARLO SIMULATION USING DISCRETE DATA
Shaw JW, Shahriar J, Horrace W
The University of Arizona, Tucson, AZ, USA

BACKGROUND: In a simulation using continuous data, we compared the performance of DeLong’s test for nonparametric receiver operating characteristic (ROC) curves (Biometrics 1988; 44: 837–845) with that of Vuong’s test for model selection (Econometrica 1989; 57: 307–333). Both tests were found to yield similar results regardless of sample size. Diagnostic tests are often measured on an ordinal rating scale, and nonparametric methods tend to underestimate the area under the ROC curve (AUC) when used with discrete data. Thus, it was conceivable that level of measurement might influence the performance of DeLong’s test. OBJECTIVE: A second Monte Carlo simulation was performed to determine whether DeLong’s and Vuong’s tests behave differently when used with discrete data. METHODS: One thousand observations were randomly generated for a Bernoulli dependent variable and 11 binomial independent variables. The independent variables were generated such that realizations were integers and Vuong’s tests behave differently when used with discrete data. Thus, it was conceivable that level of measurement might influence the performance of DeLong’s test. Both simulations, Vuong’s test appears to present a useful alternative to ROC analysis for comparing the accuracy of diagnostic tests.

PMA14

BAYESIAN DECISION ANALYSIS IN OUTCOME STUDIES WITH SMALL NUMBERS OF EVENTS: A SIMULATION BASED PREDICTION APPROACH
Wang J, Davey P, MacDonald T
University of Dundee, Dundee, Scotland

BACKGROUND: Decisions are often based on relative risk and their asymptotic properties, which is not reliable when the number of events is small. Moreover, clinical decision-making primarily depends on individual risk of adverse outcome rather than relative risk. Bayesian decision analysis predicts individual outcome, is valid for small samples and can include decision-maker’s prior knowledge into the analysis. METHOD: We analysed data about gastrointestinal adverse events of medium and high doses of ibuprofen in a population of 46,249 patients. We used a Bayesian method based on expected utility with a utility function EFF = q L(No. Events). Where EFF is the efficiency, L is a quadratic function representing the risk of Adverse Outcome and q represents relative importance of the risk. Bayesian value of information (VOI) of additional observations of a particular subgroup was calculated. Markov Chain Monte Carlo procedure and software BUGS were used to fit a Poisson regression model to adjust for confounders. RESULT: There were 1 and 5 G.I. events in high and medium dose groups (relative risks, RR 5.26 and 2.36 respectively). The Bayesian mean log-RR between high and medium dose was 0.41 (95% CI = 2.72, 2.58). Assuming that the higher dose had 20% higher efficiency, we found that medium dose is preferable when q is larger than 15. VOI of additional observations was calculated for a range of q and showed that additional observations of the higher dose would be most valuable. For example, when q = 50 the VOI of an additional subgroup of 1000 person-years exposure was 15% for high doses but only 3% for medium doses. CONCLUSION: In comparison with the classical approach for drug safety or other outcome studies, Bayesian methods provides much more information to assist decision-making, especially when the number of events is small.