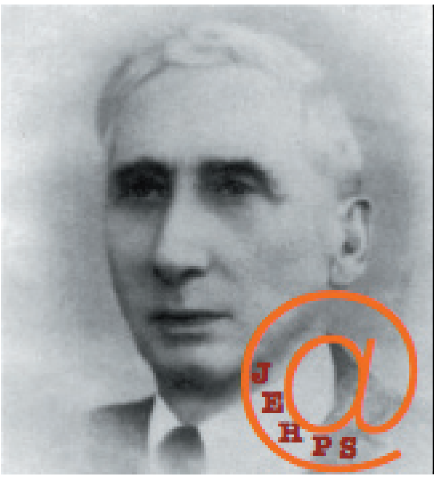

Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 5, n°2; Décembre/December 2009

www.jehps.net

Yates and Contingency Tables: 75 Years Later

David B. HITCHCOCK[1]

Résumé

Il y a soixante-quinze ans, Yates [1934] a publié un article présentant sa correction de continuité au test du χ^2 pour l'indépendance dans les tables de contingence. L'article était aussi une des premières introductions au test exact de Fisher. Nous discutons l'importance historique de Yates et de son article de 1934. Le développement du test exact et de la correction de continuité est étudié de façon détaillée. Les discussions ultérieures sur le test exact et la correction de continuité sont brièvement décrites en essayant de cerner l'importance de l'article de 1934.

Abstract

Seventy-five years ago, Yates [1934] presented an article introducing his continuity correction to the χ^2 test for independence in contingency tables. The paper also was one of the first introductions to Fisher's exact test. We discuss the historical importance of Yates and his 1934 paper. The development of the exact test and continuity correction are studied in some detail. Subsequent disputes about the exact test and continuity correction are briefly recounted, as we attempt to ascertain the 1934 paper's place in history.

[1]University of South Carolina, Department of Statistics, Columbia, SC 29208, hitchcock@stat.sc.edu

1 Introduction

The year 2009 marks the 75th anniversary of the publication of Frank Yates' paper [Yates 1934] discussing options for testing for association in contingency tables: the Pearson χ^2 test, Fisher's exact test, and the well-known continuity correction that bears Yates' name. While it may not be commonly given credit as a colossally influential article that has shaped statistical science, this medium-length paper introduced some lasting ideas and methods (as well as some perpetually controversial ones) to contingency table analysis.

The purpose of Yates' paper is twofold: firstly, to introduce statisticians to Fisher's exact test (a very new procedure at the time), in large part to use the exact test as a sort of gold standard against which the small-sample performance of the (at that point in time) well-established χ^2 test of Pearson may be judged. Secondly, Yates presents his continuity correction, which results in the χ^2 test better approximating the exact test.

Yates [1934] motivates the discussion by immediately asserting, "The χ^2 test is admittedly approximate, for in order to establish the test it is necessary to regard each cell value [i.e., count] as normally distributed with a variance equal to the expected value, the whole set of values being subject to certain restrictions." Note that the variance equals the mean in the archetypal count model, the Poisson, and that the normal approximates a Poisson with large mean. This heuristic argument was also used by Fisher in a revised footnote within a paper on the χ^2 statistic [Fisher 1922].

Of course, the χ^2 approximation requires a large sample size, and Yates quotes the rule of thumb that is still most commonly used today: the χ^2 test is "sufficiently accurate if no cell has an expectancy of less than 5." In Section 5 we deal with Yates' discussion of the performance of the χ^2 test for small to moderate samples.

After some brief background information about Frank Yates, we explore in Sections 3 through 5 his 1934 paper, outlining its major statistical contributions. In Section 6 we briefly discuss controversies and criticisms of Yates' correction that arose in later years and revisit Yates' 1984 reply to those critics. The final section details the historical importance of the 1934 paper.

2 Background information about Frank Yates

There are a number of good biographical articles about Frank Yates in the statistical literature, including those by Nelder [1997], Dyke [1995] and Healy [1995a and 1995b]. Here we briefly summarize some highlights from his career, including those relevant to the 1934 paper on contingency tables.

Yates came to statistics when he began working at Rothamsted Experimental Station in 1931 as an assistant to R. A. Fisher, who was already highly prominent at that time. When Fisher left Rothamsted two years later, Yates rose to head of the Statistics Department, where he remained for 35 years, while still continuing collaborations with Fisher [Nelder 1997]. It is natural, albeit somewhat unfortunate, that Yates' legacy is so closely tied to Fisher. Healy [1995b], while noting that Yates was "undoubtedly Fisher's follower and stood in [Fisher's] shade," suggested that Yates' work was a major impetus for Fisher's statistical insights spreading through the larger scientific community. Healy rated Yates (as a *practicing* statistician) at least as highly as Fisher.

The contributions of Yates to the field of experimental design are well known and well recounted in many of the references listed above. However, Yates' greatest contribution to statistics is perhaps his embrace of the use of computing to solve statistical problems. This philosophy is of course of primary importance today, and is quite relevant to the article which we discuss here.

3 Development of Fisher's exact test

As Yates points out in his first paragraph, the χ^2 test was famously introduced by Pearson [1900], with Fisher [1922] modifying the degrees of freedom of the test statistic. The origin of the exact test is somewhat murkier. The first appearance of the exact test in Fisher's book *Statistical Methods for Research Workers* (*SMRW*), whose first edition had been published in 1925, was in the fifth edition, published in 1934, the same year as Yates' paper. (The fifth edition of *SMRW* also included, for the first time, discussion of Yates' continuity correction.) Clearly, *SMRW* was quite expeditious in reflecting the then state-of-the-art methods in contingency table analysis.

While Fisher was likely the first person to derive the exact test, he was not the first to attempt to publicly disseminate it. That honor may belong to

Joseph Oscar Irwin (who had previously worked with Fisher at Rothamsted Experimental Station), who in 1935 published a paper on comparing two binomial proportions using an exact test. In a footnote to that article, Irwin [1935] wrote, "This paper was concluded in May 1933, but its publication has been unavoidably delayed. Meanwhile a paper dealing with the same subject, in some respects more completely, has been published by F. Yates." In fact, Yates' 1934 paper was "more complete" than Irwin's only in certain aspects. Irwin's paper dealt solely with the case of 2×2 contingency tables, whereas Yates mentioned more general tables. Also, the introduction of the exact test was merely a part of Yates' paper, with the introduction of the continuity correction occupying the major portion. On the other hand, Irwin focused solely on the development of the exact test in his paper, and in that particular respect, provided much more description than did Yates. Armitage [1982] and Greenberg [1983] provide information on Irwin's career.

Since Irwin and Yates each worked with Fisher at Rothamsted, a reasonable conclusion is that Fisher influenced each of them to use the exact test, and in published work they each described the procedure slightly differently. Yates [1984, p. 429] would later suggest that Fisher, "as is indicated by a cryptic passage" from a *Eugenics Review* paper [Fisher 1926], had essentially thought of the exact test as early as 1926, noting that a probability Fisher gave there had been derived via exact methods. Other early work provides evidence of Fisher's primacy in the development of the exact test. Bartlett [1935] proposed using the χ^2 statistic as a means of testing for second-order interactions in $2 \times 2 \times 2$ tables (and extensions thereof). Bartlett credited Fisher with suggesting the method of calculating expected cell counts and relied on Fisher's [1922] large-sample theory for χ^2. For small samples, Bartlett borrowed, presumably also from Fisher, the idea of using exact multinomial probabilities to calculate significance.

Interestingly, while Yates [1934] refers in the first sentence of the paper to "statistical tests of independence in contingency tables," when he motivates the exact test in the 2×2 case, he assumes that the data come from "two samples of $N - n$ and n respectively," each from a binomial distribution. This would actually correspond to a "test for homogeneity," which is based on a different sampling scheme [Bock et al. 2007]: In this setting, samples are taken from multiple populations and a single categorical variable is observed, rather than, say, two categorical variables being observed on a single sample in the test for independence. Of course, the implementations of both the exact test and the χ^2 test are identical whether we are testing for independence

Table 1: A cross-classification table following the notation of Yates [1984].

	B_1	B_2	Total
A_1	a	b	n_1
A_2	c	d	n_2
Total	m_1	m_2	N

NOTE: N observations, on which two binary variables A and B are measured, are cross-classified.

or homogeneity, so we may use either setting to motivate the test statistic calculation. Anyway, Yates makes no distinction between the two settings in the article.

We note that in the 1934 paper Yates denotes the marginal totals by $N - n$, n, $N - n^{'}$, and $n^{'}$, respectively. In a later paper [Yates 1984], he adapts an arguably less confusing notation for the 2×2 table, shown in Table 1.

In the 1934 paper, Yates (giving credit to a suggestion by Fisher) derives the fact that the P-value of the exact test is a hypergeometric probability. Yates' explanation of the exact test is somewhat more theoretically detailed than Fisher's in *SMRW* (although Fisher illustrates the method with a numerical example). While Yates states that the method is due to Fisher, Healy [1995b] wonders – with a touch of generous hyperbole, perhaps – if it might be more appropriately called the Fisher-Yates Exact Test, given Yates' role in disseminating the method.

In the case of a 2×2 table, the P-value of the exact test represents the probability that the count in a particular cell is as or more favorable to the alternative hypothesis (of association) than the observed count for that cell, when the margins of the table are fixed at their observed values. Is it reasonable to assume the "constancy of marginal totals"? This is a question that many practitioners gloss over, but Yates devotes nearly a page and a half to justifying this assumption. As it turns out, through the years Yates' arguments in this regard were not universally accepted, and this led in part to much controversy through the years, briefly discussed in Section 6.

Yates' first, fairly intuitive, argument was that the conclusion of the test is in no way affected by which variable represents the rows and which represents the columns; therefore, since the row totals are fixed (as is the case when taking two independent binomial samples), we can just as well view the columns as fixed. A second argument Yates presents is that we may view the observations as a *single* random sample of size N, and then classify those sampled observations according to the row variable (say, A). We may then randomly select m_1 of these N observations and assign them to class 1 of the column variable (say, B), with the other m_2 of the observations being assigned to class 2 of B. While this establishes the constancy of the column totals, it may not intuitively mirror the way data are gathered for 2×2 tables, in many cases. Finally, Yates mentions that "the marginal totals are in the nature of ancillary statistics." As shown in e.g., Little [1989], the column totals m_1 and m_2 are not truly ancillary but merely "approximately ancillary, in the sense that (m_1, m_2) usually contains very little information" about the odds ratio [Little 1989, p. 286]. This fact never concerned Yates, but did lead to conflicting perspectives about the appropriateness of the exact test.

Throughout the paper, Yates focuses mostly on the 2×2 case, although Fisher's exact test can be extended to general $r \times c$ tables. (In that case, one must choose a method of ordering possible table configurations according to how severely they depart from the null. Yates [1934, p. 234] suggests that the χ^2 statistic itself serve "as a criterion by which deviations from expectation may be arranged in order of magnitude" but notes, "Whether χ^2 is really the best criterion of a deviation from expectation we do not propose to study here.") Having derived the exact test, Yates concludes, "even when the marginal totals are quite small the evaluation of χ^2 is much more expeditious," and he focuses on the properties of the χ^2 test. In fact, one suspects that Yates discusses the exact test here primarily to motivate exact probability calculations with which the various χ^2 approximations may be compared.

4 Yates' continuity correction

The part of the Yates [1934] paper that is most well-known today may be his continuity correction recommendation, for which, according to Nelder [1997], "Yates is probably most widely known." The ordinary (uncorrected) χ^2 statistic commonly given as

$$\chi^2 = \sum_{i,j} \frac{(Obs_{ij} - Exp_{ij})^2}{Exp_{ij}}$$

where Obs_{ij} is the observed count in cell (i, j) and Exp_{ij} the estimated expected count in cell (i, j) given no association, results in a uniformly lower P-value than that of the exact test, as Yates illustrates with a simple table. His recommendation is to adjust the expected counts 1/2 unit closer to the observed counts for each cell, resulting in:

$$\chi^{2'} = \sum_{i,j} \frac{(|Obs_{ij} - Exp_{ij}| - 0.5)^2}{Exp_{ij}}.$$

Yates points out that this adjusted test statistic produces a P-value sometimes greater and sometimes less than the exact P-value, but which is typically much closer to the exact than is the P-value from the uncorrected χ^2. To illustrate this, Yates first deals with the simple case of a 1×2 contingency table, calculating some exact binomial probabilities and comparing these with the corresponding (one-tailed) P-values of the ordinary χ^2 test and the χ^2 test with his continuity correction. For example, if the true probability of success $p = 0.5$, in 10 independent trials the exact probability of 4 or fewer successes is 0.3770. The P-value of the uncorrected test (with a "less than" alternative) is 0.2635, and the P-value of the corrected test is 0.3759, far closer to the "exact" value. Yates points out that this is because the χ^2 distribution is continuous, "whereas the distribution it is endeavouring to approximate is discontinuous." With several other such examples, Yates provides evidence for the continuity correction's improvement to the approximation.

5 Yates' study of the small sample behavior of the χ^2 test

In the section of his paper labeled, "Discrepancies of the χ^2 Test after correcting for Continuity," Yates [1934] computes discrepancies between the 0.025 and 0.005 cutoff values of the χ^2 distribution and the corresponding values of the sampling distribution of the continuity-corrected test statistic $\chi^{2'}$. This is by far the most technical section of the paper, and given the style of exposition of Yates' day (heavy on wordy explanations and light on mathematical notation), it is probably the most difficult to follow.

Whereas in the previous section Yates compared P-values arising from the uncorrected and corrected χ^2 tests, in this section he attempts to compare specific cutoff points for quantities having a discrete distribution. This is a bit more awkward: Yates [1934, p. 223] notes, "There are, of course, in general no discrepancies corresponding to the exact 2.5 per cent and 0.5 per cent points, but it is possible to determine approximate hypothetical discrepancies ... by interpolation." These interpolations give rise to some unusual-looking plots with multiple axes that require some study to understand. Yates later apparently regretted this focus on nominal levels in this discrete type of test, noting that "concentration on single-tail nominal levels of 2.5 and 0.5 per cent is a defect in my 1934 paper, which reflects the current thinking of the time" [Yates 1984, p. 437].

Specifically, Yates orders Table 1 so that $n_1 > n_2$ and $m_1 > m_2$ (if this is not so, one can simply switch the category labels). He then examines the distribution of d, which in this formulation is the random value for the cell with the smallest expected value. He notes that its value may range from 0 to n_2 (i.e., across $n_2 + 1$ terms) and its expected value will be $E(d) = n_2 m_2/N$. If we fix $E(d)$ and n_2+1, we obtain a series of distributions for d as the overall total N varies. Yates calls the distribution in such a series with the smallest possible N (for a given $E(d)$ and n_2+1) the *limiting contingency distribution*, and notes that when $N \to \infty$, the distribution approaches a binomial. By examining the discrepancies associated with the continuity-corrected χ^2 for each of these extreme distributions, Yates could generally characterize the discrepancies for any distribution in that series.

In terms of practically applying his findings, Yates suggests using the continuity correction whenever the smallest expected cell count is less than 500 (in other words, in any case other than when the correction has a negligible effect). When the smallest expected cell count is less than 100, Yates suggests forgoing the ordinary χ^2 table when obtaining the critical value for the test. Rather, the analyst should find an adjusted cutoff value based on the table Yates [1934, Table III] presents. If interpolation within the levels given in Yates' table is not precise enough, Yates recommends that the exact test be used.

Again, these small-sample conclusions and recommendations Yates provides are philosophically based on the exact test being a gold standard; Yates is determining what application of a χ^2-type test can best approximate the exact-test results. As the years went by, the general trust in the exact test fell into doubt in some statisticians' eyes. Yates [1934] ignores the question of whether the null distributions of χ^2 and $\chi^{2'}$ are nearly χ^2_1 for small samples, which seems to be a more natural gold standard against which to measure performance. Later work [e.g., Larntz 1978] investigated this question, albeit not necessarily with respect to the continuity correction. Starmer et al. [1974, p. 377] flatly stated, "There seems to be no good reason to use the exact test as the standard of comparison for competing tests" of association, instead suggesting as a gold standard a randomized version of the exact test, which would counteract the exact test's conservatism. In hindsight, Yates' usage of the exact test as a benchmark is a notable weakness in his 1934 study.

6 Later Criticism and Controversy

We now give a brief overview of the substantial controversy engendered by the methods in Yates [1934], without attempting a comprehensive investigation into the foundational issues. In the 1970s and 1980s, simulation-based evaluations of Yates' prescribed methods revealed their inherent conservatism. Berkson [1978], Haber [1980], Upton [1982] and Garside and Mack [1976] showed, for a variety of testing situations, that the power of both the continuity-corrected χ^2 test and the exact test usually failed the reach the nominal significance level. Later, Upton [1992] would reverse his criticism of the exact test, concluding it was unfair, when dealing with a test based on a discrete distribution, to compare the Type I error rate with a nominal α. A more modern way to alleviate the conservatism of such discrete tests is to use the mid-P-value.

In addition, a number of statisticians have opposed the idea of conditioning on both sets of marginal totals and presented competing methods, beginning with Wilson [1941] and Barnard [1945], both of whom, according to Fisher [1956], later recanted their opposition to the exact test. However, later unconditional approaches, including those of Suissa and Shuster [1985] and Haber [1986], were in stark contrast to the conditional approach. Concerning the continuity correction, Conover [1974] stated that Yates' correction would ensure an improvement only when $n_1 = n_2$ and $m_1 = m_2$, and never unless both marginal totals were truly fixed.

Yates [1984] wrote a long and detailed defense of the exact test (and, to a lesser degree, of the continuity-corrected χ^2 test). Using a large series of examples, Yates defended the exact test on mostly philosophical grounds that highlighted his objections to the Neyman-Pearson style of hypothesis testing. Some foundational issues intrinsic to Yates' arguments included:

the interpretation of the significance level as the long-run proportion of rejections when H_0 is true; the role of conditioning on the marginals when testing association in two-way tables; the use (or misuse) of strict nominal significance levels; and the correct approach for adapting one-sided P-values to a two-sided test and vice versa. Many of the questions, of course, have no indisputably correct answer. Moreover, Yates in 1934 could hardly have forseen such criticism. However, Yates [1984], in passionately – even indignantly – refuting the critics' conclusions, was supported in his fundamental viewpoints by a number of the discussants, including Barnard [1984], Cox [1984], and Mantel [1984].

7 Conclusion

The historical significance of Yates' 1934 article is perhaps underrated. Though merely of moderate length, it provided one of the earliest explanations in the statistical literature of Fisher's exact test, at around the same time that Fisher [1934] mentioned the exact test in the fifth edition of *SMRW* (although the idea of the exact test seems to have been known for some time previous by those in Fisher's inner circles). In addition, Yates [1934] formally proposed the continuity correction to the χ^2 test for the first time. Finally, Yates' numerical studies in this paper were the first in a long and often contentious series of investigations into the best method for testing for association in contingency tables. This controversy has carried into modern times, without a definitive resolution, and has led to foundational questions about the meaning of conditional tests and the appropriateness of Neyman-Pearsonian measures of actual levels of significance.

Admittedly, most statisticians today would likely say that the continuity correction to the χ^2 test and its relationship to the exact test is a fairly minor dispute today, as statistical controversies go. While Yates' 1934 paper may be more important historically than as a progenitor of current research, it is more than a mere curio. Since the χ^2 and exact tests are used in numerous everyday applications, its influence, though indirect, is ubiquitous on statistical practice. The acceptance of Yates' continuity correction today is rather mixed in practice, with texts and software packages differing in their recommendations concerning the correction.

Reading Yates' 1934 article and its golden-anniversary counterpart [Yates 1984], one may glean some sharp insights into Frank Yates' personality and statistical philosophy. Yates' deep respect for Fisher shines through, especially in the 1984 paper, in which Yates repeatedly refers to Fisher's work and statistical philosophy. In the examples and justifications Yates [1934] provides, we sense his relatively early reliance on computing machines as vital tools for statistical analysis and his belief in the primacy of applications. From the occasionally almost-scathing rejoinder [Yates 1984] to his critics, on the other hand, we encounter another side of his personality in the acerbic, Fisherian defense of his ideas.

While the 1934 paper is neither the first word nor the last word regarding testing for association in contingency tables, it is historically important for several reasons. It provided a snapshot of the contemporary state of the art at the time, it discussed (probably, for many readers, introduced) Fisher's method for exact testing, and it suggested the continuity correction that would be debated for decades to come. Yates bestowed upon statistics many contributions during his distinguished career, and his 1934 paper is a small gem whose brilliance can be seen three-quarters of a century later.

Acknowledgments

The author gratefully acknowledges the suggestions of an editor and two anonymous referees, which have greatly improved this paper.

References

Armitage, P. (1982), "Joseph Oscar Irwin (1898-1982)," *Journal of the Royal Statistical Society, Series A*, 145, 526-528.

Barnard, G. A. (1945), "A New Test for 2×2 Tables," *Nature*, 156, 177.

Barnard, G. A. (1984), Discussion of "Tests of Significance in 2×2 Tables," *Journal of the Royal Statistical Society, Series A*, by F. Yates, 147, 449-450.

Bartlett, M. S. (1935), "Contingency Table Interactions," *Supplement to the Journal of the Royal Statistical Society*, 2, 248-252.

Berkson, J. (1978), "In Dispraise of the Exact Test," *Journal of Statistical Planning and Inference*, 2, 27-42.

Bock, D. E., Velleman, P.F., and De Veaux, R. D. (2007), *Stats: Modeling the World*, Second Edition, Boston: Pearson/Addison-Wesley.

Conover, W. J. (1974), "Some Reasons for Not Using the Yates Continuity Correction on 2×2 Tables," *Journal of the American Statistical Association*, 69, 374-376.

Cox, D. R. (1984), Discussion of "Tests of Significance in 2×2 Tables," *Journal of the Royal Statistical Society, Series A*, by F. Yates, 147, 451.

Dyke, G. (1995), "Obituary: Frank Yates," *Journal of the Royal Statistical Society, Series A*, 158, 333-338.

Fisher, R. A. (1922), "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P," *Journal of the Royal Statistical Society*, 85, 87-94.

Fisher, R. A. (1926), "Bayes' Theorem and the Fourfold Table," *Eugenics Review*, 18, 32-33.

Fisher, R. A. (1934), *Statistical Methods for Research Workers*, Fifth Edition, Edinburgh: Oliver and Boyd.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Edinburgh: Oliver and Boyd.

Garside, G. R., and Mack, C. (1976), "Actual Type 1 Error Probabilities for Various Tests in the Homogeneity Case of the 2×2 Contingency Table," *The American Statistician*, 30, 18-21.

Greenberg, B. G. (1983) "Joseph Oscar Irwin, 1898-1982, An Obituary Appreciation," *Biometrics*, 39, 527-528.

Haber, M. (1980), "A Comparison of Some Continuity Corrections for the χ^2 Test on 2×2 Tables," *Journal of the American Statistical Association*, 75, 510-515.

Haber, M. (1986), "An Exact Unconditional Test for the 2×2 Comparative Trial," *Psychological Bulletin*, 99, 129-132.

Healy, M. J. R. (1995a), "Obituary: Frank Yates 1902-1994," *Biometrics*, 51, 389-391.

Healy, M. J. R. (1995b), "Frank Yates, 1902-1994: The Work of a Statistician," *International Statistical Review*, 63, 271-288.

Irwin, J. O. (1935), "Tests of Significance for Differences between Percentages Based on Small Numbers," *Metron*, 12, 83-94.

Larntz, K. (1978), "Small-sample Comparisons of Exact Levels for Chi-Squared Goodness-of-fit Statistics," *Journal of the American Statistical Association*, 73, 253-263.

Little, R. J. A. (1989), "Testing the Equality of Two Independent Binomial Proportions," *The American Statistician*, 43, 283-288.

Mantel, N. (1984), Discussion of "Tests of Significance in 2×2 Tables," *Journal of the Royal Statistical Society, Series A*, by F. Yates, 147, 457.

Nelder, J. A. (1997), "Yates, Frank," in *Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present*, Norman L. Johnson and Samuel Kotz (eds.), New York: Wiley.

Pearson, K. (1900), "On a Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such that It Can Be Reasonably Supposed to Have Arisen from Random Sampling," *Philosophical Magazine, Series 5*, 50, 157-175. (Reprinted in *Karl Pearson's Early Papers* (1948), E. S. Pearson (ed.), Cambridge: Cambridge University Press.)

Starmer, C. F., Grizzle, J. E., and Sen, P. K. (1974), Comment on "Some Reasons for Not Using the Yates Continuity Correction on 2×2 Tables," *Journal of the American Statistical Association*, 69, 376-378.

Suissa, S., and Shuster, J. J. (1985), "Exact Unconditional Sample Sizes for the 2 by 2 Binomial Trial," *Journal of the Royal Statistical Society, Series A*, 148, 317-327.

Upton, G. J. G. (1982), "A Comparison of Alternative Tests for the 2×2 Comparative Trial," *Journal of the Royal Statistical Society, Series A*, 145, 86-105.

Upton, G. J. G. (1992), "Fisher's Exact Test," *Journal of the Royal Statistical Society, Series A*, 155, 395-402.

Wilson, E. B. (1941), "The Controlled Experiment and the Four-fold Table," *Science*, 93, 557-560.

Yates, F. (1934), "Contingency Tables Involving Small Numbers and the χ^2 Test," *Supplement to the Journal of the Royal Statistical Society*, 1, 217-235.

Yates, F. (1984), "Tests of Significance for 2×2 Contingency Tables," *Journal of the Royal Statistical Society, Series A*, 147, 426-463.