An Easy to Use Method to Approximate Poisson Confidence Limits
Author(s): Bernard Bégaud, Karin Martin, Abdelilah Abouelfath, Pascale Tubert-Bitter, Nicholas Moore, Yola Moride
Source: *European Journal of Epidemiology*, Vol. 20, No. 3 (2005), pp. 213-216
Published by: Springer
Stable URL: http://www.jstor.org/stable/25047436
Accessed: 21/10/2009 23:54

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=springer.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Springer is collaborating with JSTOR to digitize, preserve and extend access to *European Journal of Epidemiology*.

http://www.jstor.org

European Journal of Epidemiology (2005) 20: 213–216
DOI 10.1007/s10654-004-6517-4
© Springer 2005

METHODS

An easy to use method to approximate Poisson confidence limits

Bernard Bégaud[1,2,3], Karin Martin[1,2,3], Abdelilah Abouelfath[1,2], Pascale Tubert-Bitter[4,5], Nicholas Moore[1,2,3] & Yola Moride[6]

[1]*Département de Pharmacologie, Université Victor Segalen, 33076 Bordeaux cedex, France;* [2]*INSERM U657, 33076 Bordeaux cedex, France;* [3]*CHU, 33076 Bordeaux cedex, France;* [4]*INSERM U472, 94807 Villejuif, France;* [5]*Hôpital Paul Brousse, 94807 Villejuif, France;* [6]*Université de Montréal, Montréal, Canada*

Accepted in revised form 22 November 2004

Abstract. Despite the ever larger choice of softwares and statistical packages allowing fast and accurate computation of binomial and Poisson confidence limits, there is always a need for a simple and reliable formula allowing non-computerized computations. The method proposed in this paper is derived from the Freeman and Tukey's variance stabilizing transformation for a random Poisson variable and adjusted for giving the best fit with the exact Poisson values. Despite its simplicity, allowing its use in any circumstances, this method provides very satisfactory results and a much better fit than classical formula based on the normal approximation, even if a continuity correction is used. It allows computation of Poisson confidence limits both for count or rates and proportions.

Key words: Biostatistics, confidence intervals, Poisson approximation, Poisson count

Introduction

As the basis of statistical inference, computation of confidence intervals is of major importance in biostatistics and epidemiology. For dichotomous outcomes, the most common in epidemiology, the reference distribution is binomial. Provided that the number of attempts n tends to infinite and p the probability of success approaches 0, the Poisson distribution constitutes a very satisfactory approximation of the binomial distribution [1–5].

Before statistical packages being available, computing exact binomial distributions was fairly complex, so that approximations were commonly used. The simplest are based on the normal approximation but do not provide satisfactory results when the expected or observed number of events of interest becomes small, i.e., lower than 10 [6–9] or 30 [10–11], even if a continuity correction is used [1, 12]. Much more accurate results can be derived from more complex formula, e.g., those based on the F or X^2 distributions [13], however such computations are relatively troublesome if not made with the help of a computerized program.

Despite the ever greater set of statistical packages available today, there is undisputably a need for a simple formula providing satisfactorily accurate results when a computer or powerful calculator is not available or when its use appears not time effective, e.g. if only a single estimate is needed.

This paper proposes a handy method developed for this purpose and compares the results with exact computation and those obtained from normal approximation.

Methods

If k is the number of events, during an observation period and thus assumed to be a Poisson variable, $\sqrt{k}$ and preferably, $\sqrt{k+0.5}$ are expected to roughly fit the normal distribution [2]. It was the basis of the square root variance stabilizing transformation for a Poisson variable proposed by Freeman and Tukey [14] from which it can be derived that the 100 $(1-\alpha)$% two-sided confidence interval, on the basis of k occurrences, is [15]:

$$\left(\frac{Z_{1-\alpha}}{2} \pm \sqrt{k+0.5 \pm \varepsilon}\right)^2, \qquad (1)$$

ε being 0.5.

For k varying from 1 to 100 and using a program specifically developed on the STATA® software [16], we searched which values of ε led to the best fit with the Poisson distribution, i.e., resulted in the lowest sum of differences between the computed values and the exact Poisson limits.

Moreover, we propose to use formula (1) to derive confidence limits for rates and proportions by dividing the computed values by the number n of trials (e.g., number of subjects, number of units of follow-up).

By using the optimized values of ε, we computed the 95% two-sided confidence interval for 3 theoretical proportions: 3, 10 and 70 per 1000, respectively. The

results were compared to exact Poisson and binomial limits, computed with the STATA® statistical package [16], and to those obtained by the classical formula based on the normal approximation [1]:

$$p \pm Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}. \quad (2)$$

Results

As shown in Figure 1, the values of ε which provided the best fit with the Poisson distribution, i.e. the lowest sum of differences between the computed values, were 0.48 for the lower bound and 0.46 for the upper one. Therefore, for a two-sided interval, formula (1) became:

$$\left(\frac{Z_{1-\alpha}}{2} - \sqrt{k+0.02}\right)^2 \text{ for the lower limit, and}$$

$$\left(\frac{Z_{1-\alpha}}{2} - \sqrt{k+0.96}\right)^2 \text{ for the upper limit.}$$

Both are not too far from the approximation initially proposed by Freeman and Tukey (0.05 and 1, respectively).

Table 1 shows the values obtained from these formula compared with the exact 95% two-sided Poisson limits when k varies from 1 to 100.

Table 2 compares the 95% two-sided confidence intervals for a rate of 3, 10 and 70 per 1000, computed by using the proposed approach and

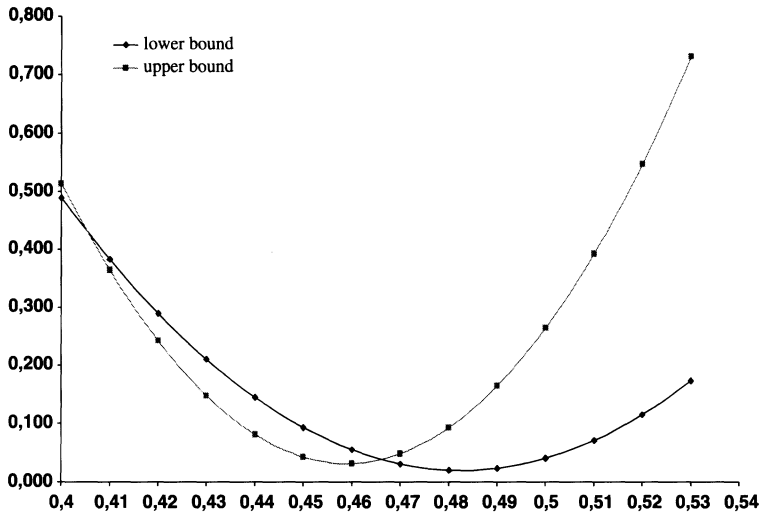

Figure 1. Values of ε providing the best fit (the lowest sum of squares of differences) with the exact Poisson two-sided 95% confidence interval: difference when k varies from 1 to 100.

Table 1. Exact 95% two-sided confidence limits for a Poisson count compared to those calculated by using the proposed formula and the Freeman and Tukey approach when k varies from 1 to 100

k	Exact Poisson confidence limits	Proposed method ε = 0.46 and 0.48	Freeman and Tukey's formula
1	[0.03–5.57]	[0.001–5.66]	[0.0004–5.73]
2	[0.24–7.23]	[0.19–7.29]	[0.19–7.36]
3	[0.62–8.77]	[0.57–8.82]	[0.57–8.88]
4	[1.09–10.24]	[1.05–10.29]	[1.04–10.34]
5	[1.62–11.67]	[1.59–11.71]	[1.58–11.76]
6	[2.20–13.06]	[2.17–13.09]	[2.16–13.15]
7	[2.81–14.42]	[2.79–14.45]	[2.77–14.50]
8	[3.45–15.76]	[3.43–15.79]	[3.42–15.84]
9	[4.12–17.08]	[4.09–17.11]	[4.08–17.16]
10	[4.80–18.39]	[4.78–18.41]	[4.76–18.46]
–	–	–	
30	[20.24–42.83]	[20.24–42.83]	[20.23–42.87]
–	–	–	
50	[37.11–65.92]	[37.12–65.91]	[37.10–65.96]
–	–	–	
70	[54.57–88.44]	[54.58–88.43]	[54.56–88.48]
–	–	–	
100	[81.36–121.63]	[81.38–121.61]	[81.36–121.66]

Table 2. 95% confidence two-sided, limits for an observed number of events of 3, 10 and 70 for 1000 units of observation, computed with the proposed approach and compared to the normal approximation (formula 2) and to exact Poisson and binomial limits

	0.003	0.010	0.070
Exact binomial	0.000619; 0.008742	0.004805; 0.018313	0.054972; 0.087618
Exact Poisson	0.000619; 0.008765	0.004796; 0.018389	0.054568; 0.088440
Proposed approach	0.000574; 0.008821	0.004776; 0.018409	0.054579; 0.088431
Normal approximation	−0.000390; 0.006390	0.003833; 0.016167	0.054186; 0.085814

formula (2) to exact binomial and Poisson limits calculated by the STATA® software [6].

Discussion

Table 1 shows that substituting 0.5 by 0.48 and 0.46, respectively, in the Freeman and Tukey's formula greatly improves the fit with the exact Poisson limits. The improvement is more marked for the upper bound estimates and constant over the 0–100 range for the k values. This is confirmed by Figure 1 which shows that the upper bound estimate is more sensitive to the value of ε than the lower bound estimate.

The only one limitation of the proposed approach appears to be the marked underestimation of the lower limit when k is unity. It is observed both for the original Freeman and Tukey's formula and for the improved one, the discrepancy being lower for the later.

Despite its intrinsic simplicity making computations by hand or with the help of a basic calculator straightforward, the proposed formula gives considerably better results than other 'simple' formulae such as those based on the normal approximation. For example, the confidence limits obtained from the formula classically used for approximating confidence intervals for a Poisson count ($k \pm Z_{1-\alpha/2} \times \sqrt{k}$) were [3.43; 15.79] vs. [3.45; 15.76] for $k = 8$ and [37.12; 65.91] vs. [37.11; 65.92] for $k = 50$, the approximation becoming acceptable only when k is large and meaningless when below 5.

The same can be said for proportions and incidence rates, the results obtained from the classical normal approximation being acceptable only for the larger values of k, i.e., 70 (Table 2). Again, it can be noted (Table 2) that the estimates from the proposed approach are quite close to the exact Poisson limits, particularly for $k = 10$ ($p = 0.01$). In the range of selected values, the Poisson and binomial intervals are almost perfectly superimposable, an expected result since n is large and p remains small, i.e., $\leqslant 0.1$.

As previously said, for the lowest value, i.e., $k = 3$ and $p = 0.03$, the proposed formula fits less satisfactorily the exact Poisson limits. However, even in this case, the results are much better than those that would have been obtained from the formula derived from the normal approximation.

Finally, the main criticism that could be opposed to the present proposal is about its practical relevance. Indeed, as mentioned in the introduction, some indirect methods based on the F or X^2 distribution [13] give very satisfactory results, as good as or even better than the proposed approach. However, in this case, computation is relatively complex and tedious if not using an *ad hoc* computerized program. Therefore, giving that a large choice of statistical packages is available today, some of them giving exact binomial and Poisson intervals, the practical relevance of these methods is today more than questionable.

On the other hand, there is still a niche for a simple computation method allowing to derive one-sided and two-sided confidence intervals without the need of a computer or a specialized program. It could be the case when working out of the office, travelling or when the use of statistical packages is not justified by other calculations and analyses. In these circumstances, the proposed approaches, because very handy and giving much more reliable results, are to be preferred to the classical formula based on the normal approximation [2] and still proposed in the majority of statistical textbooks.

Acknowledgements

The authors are indebted to Professor Joseph Lellouch (INSERM U472, Paris) for his very valuable help and comments. This work has not been supported by a specific funding or grant and, therefore, is not subject to any conflict of interest.

References

1. Snedecor GW, Cochran WG. Statistical methods, 8th edn. Ames: Iowa State University Press, 1989.
2. Miller I, Freund JE. Probability and statistics for engineers. Englewood Cliffs: Prentice Hall, 1965.
3. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. Biostatistics in clinical medicine, 2nd edn. New York: Macmillan Publishing, 1987.
4. Machin D, Gardner MJ. Calculating confidence intervals for survival time analyses. Br Med J 1988; 296: 1369–1371.

5. Gardner MJ, Altman DG. Statistics with confidence. Confidence intervals and statistical guidelines. London: British Medical Journal Editions, 1989.
6. Fisher LD, van Belle G. Biostatistics. A methodology for the health sciences. New York: John Wiley & Sons, 1993.
7. Kyburg HE. probability theory. Englewood Cliffs: Prentice Hall, 1969.
8. Ross S. A First course in probability, 2nd edn. New York: Macmillan Publishing Company, 1984.
9. Machin D, Campbell MJ. Statistical tables for the design of clinical trials. Oxford: Blackwell Scientific Publications, 1987.
10. Walker AM. Observation and inference. An introduction to the methods of epidemiology. Chestnut Hill: Epidemiology Resources Inc., 1991.
11. Tubert-Bitter P, Bégaud B, Moride Y, Abenhaim L. Sample size calculations for single group post-marketing cohort studies. J Clin Epidemiol 1994; 47: 435–439.
12. Fleiss JL. Statistical methods for rates and proportions, 2nd edn. New York: John Wiley & Sons, 1981.
13. Armitage P, Colton TH. (eds), Encyclopedia of biostatistics, Six volume set. New York: Wiley, 1998.
14. Freeman MF, Tukey JW. Transformations related to the angular and the square root. Ann Math Stat 1950; 21: 607–611.
15. Documenta Geigy. Scientific tables, 6th edn. Basel: JR Geigy, 1962.
16. Intercooled Stata 7.0 software. College Station (TX), USA.

Address for correspondence: Bernard Bégaud, Département de Pharmacologie, INSERM U657, Université Victor Segalen, 146, rue Léo Saignat, 33076 Bordeaux, France
E-mail: bernard.begaud@pharmaco.u-bordeaux2.fr