Reporting only relative effect measures was potentially misleading: some good practices for improving the soundness of epidemiological results

Title Page

Type of manuscript: Original Article

Title: Precision in reporting risk ratios - a mathematical and biological perspective

Authors and affiliation at the time the work was done:

Marco Novelli[1]

Alessandro Baldi Antognini[1]

Paolo Boffetta[2,3]

John PA Ioannidis[4]

Giovanna Spatari[5]

Francesco S Violante[3]

[1] Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Bologna, Italy

[2] Stony Brook Cancer Center, Stony Brook University, Stony Brook, New York, USA

[3] Department of Medical and Surgical Sciences (DIMEC), University of Bologna, Bologna, Italy

[4] Stanford Prevention Research Center, Department of Medicine, Stanford University, Stanford, USA

[5] Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of Messina, Messina, Italy

Words count: 1902 words

Corresponding author:

Marco Novelli

Department of Statistical Sciences "Paolo Fortunati", University of Bologna,

Via Belle Arti 42, 40127, Bologna, Italy

Tel: +39 051 20 9 8263

E-mail: marco.novelli4@unibo.it

Conflict of interest: No author has any conflict of interest to be declared, in relation to the content of this manuscript.

Sources of financial support: No author has received any financial support for the preparation of this manuscript.

Data and computing code required to replicate the results reported in this manuscript: The content of the manuscript allows anyone to replicate the results presented.

Acknowledgments: Francesco S. Violante would like to acknowledge the assistance by Simone Mattei in preparing part of this manuscript.

Abstract

Objective: In the medical and epidemiological literature there is a growing tendency to report an excessive number of decimal digits (often three, sometimes four), especially when measures of relative occurrence are small; this can be misleading.

Study Design and Setting: We combined mathematical and statistical reasoning about the precision of relative risks with the meaning of the decimal part of the same measures from biological and public health perspectives.

Results: We identified a general rule for minimizing the mathematical error due to rounding of relative risks, depending on the background absolute rate, which justifies the use of one or more decimal digits for estimates close to 1.

Conclusions: We suggest that both relative and absolute risk measures (expressed as a rates) should be reported, and two decimal digits should be used for relative risk close to 1 only if the background rate is at least 1/1,000 py. The use of more than two decimal digits is justified only when the background rate is high (i.e., 1/10 py).

Keywords: relative risk; observational studies; statistical precision; rounding error, odds ratio, confidence intervals

Running Title: Precision in reporting risk ratios

Background

Nowadays, reporting measures of relative occurrence, such as relative risks (RR), close to 1 is very common.[1] For example, a search in PubMed (3/2021) with ("HR" OR "RR" OR "OR") AND ("1.001" OR "1.002" OR "1.003" OR "1.004" OR "1.005" OR "1.006" OR "1.007" OR "1.008" OR "1.009" OR "0.991" OR "0.992" OR "0.993" OR "0.994" OR "0.995" OR "0.996" OR "0.997" OR "0.998" OR "0.999") yields 24487 hits, almost all of which reflect relative risks reported in the abstracts. There is a growing concern about false or exaggerated findings in the medical literature, and very small effects are particularly prone to these problems.[2,3] In large studies, very small effects can be observed with formal statistical significance: they are particularly sensitive to bias and their clinical importance is often unclear.[1] With small effects, it is important to decide how many decimal digits should be reported for RRs (i.e. the number of digits after the decimal point to which the number is rounded). The problem is not new, but so far it has been addressed without considering its statistical perspective as well as its biological and public health implications.

Opinions about the number of decimal digits to be reported differ. In a point-counterpoint debate two authors agreed on reporting RR only to the first decimal digit;[4,5] other authors suggested to use a number of decimal digits that maintains constant the mathematical rounding error.[6 9]

We instead suggest a simple strategy for choosing the adequate number of digits in terms of to the maximum amount of error that the researcher is willing to accept and justify. Such a method does not prespecify the number of decimal digits to use, since it may vary from case to case, but focuses on the sources of errors (rounding, statistical uncertainty, etc.) that must be combined with a consideration of the biological/public health measure of interest.

In the following discussion, we propose some practical rules for deciding how many decimal digits should be used and the additional information that should be presented to readers to correctly interpret the study results. Although in what follows we will use risk ratios to illustrate ideas, the same approach can also be used with other relative measures (e.g. odds ratios).

Mathematical/statistical perspective

The approximation/rounding of a real number x in the decimal parts induces a symmetric structure based on the standard additive-scale. Let x_k be the value of x approximated with k decimal digits; then, for any choice of k, $E_k = 5 \times 10^{-(k+1)}$ represents the maximum rounding error, $e_k = |x - x_k|$ is the absolute rounding error, and $r_k = e_k/|x|$ is the relative error.

Since the rounding error represents a subjective-instrumental noise induced by the observer, the choice of k depends on a trade-off between precision and clarity of interpretation. Therefore, k is chosen as the smallest non-negative integer satisfying a suitable constraint reflecting the selected degree of precision.

Consider the standard two by two table describing the association between a determinant and on outcome. The occurrence of an event D (e.g. a disease) in a given group A (e.g. exposed) represents the probability of D (over a given period of time, if incidence is considered), i.e., $AR_A = Pr(D|A) \in [0; 1]$. Let A and B denote the two subpopulations of interest; then, the risk ratio of A with respect to B is defined by $RR = AR_A \,/\, AR_B = Pr(D|A)/Pr(D|B)$, and the estimator of RR is $\widehat{RR} = \left[\frac{a}{a+b}\right]/\left[\frac{c}{c+d}\right] = (ac + ad)/(ac + bc)$, with $s = \sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}}$ being the estimator of its standard error and $CI_{1-\alpha}(\ln[RR]) = \left[\ln\left[\widehat{RR}\right] \pm sz_{1-\alpha/2}\right]$ the asymptotic confidence interval (CI) of level $1 - \alpha$. Any rounding of $\ln\left[\widehat{RR}\right]$ induces only a shift of the corresponding $CI_{1-\alpha}(\ln[RR])$ but its width remains unchanged since $\ln\left[\widehat{RR}\right] + sz_{1-\alpha/2} - \left(\ln\left[\widehat{RR}\right] - sz_{1-\alpha/2}\right) = 2sz_{1-\alpha/2}$ (namely, the width depends only on the rounding of the quantity $sz_{1-\alpha/2}$). Unfortunately, the same does not hold for the CIs for the RR. Indeed,

$$CI_{1-\alpha}(RR) = \left[\widehat{RR}\exp\{-sz_{1-\alpha/2}\};\ \widehat{RR}\exp\{sz_{1-\alpha/2}\}\right] \qquad (1)$$

is not symmetric around $\widehat{RR}$, since the asymptotic distribution of $\widehat{RR}$ is log-Normal. Moreover, the ratio between the width of $CI_{1-\alpha}(RR_k)$ with $\widehat{RR}_k$ rounded to k decimal digits and that obtained without this approximation is equal to the ratio $\widehat{RR}_k/\widehat{RR}$. This means that, for example, if the rounding is made such that $\widehat{RR}_k > \widehat{RR}$, for example by rounding 1.17 to 1.2, then the $CI_{1-\alpha}(RR_k)$ obtained from the rounded quantity $\widehat{RR}_k$ will be wider than that obtained from the non-approximated relative risk, i.e. $\widehat{RR}$. Note that, when $\widehat{RR}$ deviates much from 1, the rounding error gradually loses its impact on the width of the confidence interval, since the ratio $\widehat{RR}_k/\widehat{RR}$, which lies in the interval $\left(1 - \frac{e_k}{RR}; 1 + \frac{e_k}{RR}\right)$, tends to 1 regardless of k.

It is therefore possible to derive some general guidelines for the choice of k:

- the rounding error should not mask the estimated RR; in practice, this corresponds to the requirement that the relative error associated with $\widehat{RR}$ should be lower than a given threshold $\varepsilon < 1$ (usually, $\varepsilon = 10\%$ or 1%) expressing the maximum relative error that is admissible; this implies that k should be chosen such that $\left|\widehat{RR} - \widehat{RR}_k\right| \leq \varepsilon\widehat{RR}$;
- since the rounding error should not mask the statistical one, the chosen k should also guarantee that $\left|\widehat{RR} - \widehat{RR}_k\right| \leq s$. Clearly, if $s \geq 0.5$, the condition is satisfied (since $e_k \leq E_k \leq E_0 = 0.5$).

By combining these two requirements,

$$\left|\widehat{RR} - \widehat{RR}_k\right| \leq \min\{s, \varepsilon\widehat{RR}\} \qquad (2)$$

A less stringent condition could be expressed in terms of the maximum rounding error, by replacing in (2) the absolute rounding error with E_k. Under this approach, k should be chosen as the lowest non-negative integer such that

$$k \geq \max\{-\log_{10}[2s]; -\log_{10}[2\varepsilon\widehat{RR}]\}. \qquad (3)$$

Note that the previous requirements could become quite stringent when the sample sizes of the two groups (namely, $a + b$ for A and $c + d$ for B) grow, since s tends to vanish.

We emphasize that it is crucial to adopt a coherent strategy in reporting risk measures and their CI adapting the confidence level to the precision required for the analysis.

As an example, suppose that $a = 495, b = d = 5,\ c = 395$, so that $\widehat{RR} = 1.002532$ and $s = 0.0072$. Letting $\varepsilon = 1\%$, it follows from (2) that $k = 0$, namely RR should be reported as 1. Clearly, the same reasoning applies to the CI's bounds, so that from (2), and letting $1 - \alpha = 0.99$, we obtain $k = 2$ with a corresponding confidence interval $(0.98\,;1.02)$.

Instead, if $a = 95, b = 5, c = 6$ and $d = 94$, then $\widehat{RR} = 15.833333$. In this case $s = 0.396476$ and if we set $\varepsilon = 1\%$, we obtain $k = 1$, while letting $\varepsilon = 5\%$ then $k = 0$. This corresponds to a $CI_{0.99} = (5.7\,;44.0)$ and $CI_{0.95} = (7.3\,;34.4)$, for $1 - \alpha = 0.99$ and 0.95, respectively. Table 1 shows some examples applying condition (3) with two different choices of the maximum relative error ε. For RR close to 1, even with $\varepsilon = 0.1\%$, the number of required digits is either 2 or 3, while for RR values much higher than 1 the number of chosen digits becomes smaller and often 1 digit is enough.

Clinical perspective

We will now examine the issue of reporting RRs from the viewpoint of their clinical and public health significance. Clearly, for large effects, one or even zero decimal digit suffices to convey all relevant information (practically, we would give the same meaning to an RR equal to 8, 8.4 or 8.44). However, when RRs are close to 1, chasing statistically significant results may induce reporting RRs (or the lower limit of the *CI*) as small as 1.0001 or 0.9999.[10,11]

A general guideline may come from the concept of "*minimal clinically important difference*", a cornerstone in clinical studies,[12] rather neglected in non-clinical-epidemiology.[13] For instance, suppose studying a specific exposure in relation to mortality in elderly patients, with an incidence of 1/10 person-years (py). Then, with a RR of 1.11, the first decimal digit means one additional case in 100 py (no doubt, a sizable effect); while the second decimal indicates one additional case in 1,000 py (a certainly less impressive effect). When studying a rarer outcome, e.g. thyroid cancer, with an incidence in the order of 1/10,000 py, the first decimal digit of the RR approximately identifies 1 event out of 100,000 py and the second digit about 1 case out of 1,000,000 py. In summary, the underlying incidence rate should be estimated along with RR whenever possible: by analogy, the concept of minimal *epidemiologically* important difference should drive the decision on how many decimal digits to be reported.

In clinical studies, the number needed to treat and the number needed to harm are generally reported to give a straightforward representation of the absolute effects of the investigated intervention. In particular, the number needed to harm is defined as the number of persons needed to be treated, on average, to produce one more adverse event than would not have occurred without the treatment.[14] This concept has been expanded and applied in epidemiology, although seldom used.[15] Researchers could complement the information implicit in the RR with the calculation of the Exposure Impact Number (EIN, defined as the number of people with the exposure among whom one case is attributable to the risk factor) and the Exposed Cases Impact Number (ECIN, the number of exposed people with the disease or outcome among whom one case is attributable to the risk factor).[15]

Conclusions

Regardless of the number of digits, reporting only relative effect measures is potentially misleading.[16-18] We recommend researchers, reviewers and journal editors to follow a shortlist of good practices (see Table 2) in reporting both relative and absolute risk (expressed as rates) measures, and not to exceed a maximum number of decimal digits when RR is close to 1. Following these practices could improve the soundness of epidemiological results.

Table 1. Examples of approximations induced in the RR and the corresponding $CI_{1-\alpha}$ by maximum relative error of 10%, 1% and 0.1%

							$\varepsilon = 10\%$			$\varepsilon = 1\%$			$\varepsilon = 0.1\%$		
a[a]	b[a]	c[a]	d[a]	RR[b]	EIN	ECIN	k	RR_k	$CI_{0.90}$	k	RR_k	$CI_{0.99}$	k	RR_k	$CI_{0.999}$
50	450	49	451	1.020408	500.0	50.0	1	1.0	(0.7; 1.4)	2	1.02	(0.62; 1.67)	3	1.020	(0.545; 1.912)
100	400	49	451	2.040816	9.8	2.0	1	2.0	(1.6; 2.7)	2	2.04	(1.34; 3.10)	3	2.041	(1.196; 3.484)
250	250	49	451	5.102041	2.5	1.2	1	5.1	(4.0; 6.5)	1	5.1	(3.5; 7.4)	2	5.10	(3.19; 8.16)
41	59	8	892	46.125000	2.5	1.0	1	46.1	(25.0; 85.0)	1	46.1	(17.7; 120.2)	1	46.1	(13.6; 156.8)

Abbreviations: EIN, Exposure Impact Number; ECIN, Exposed Cases Impact Number; RR, relative risk;

[a] Notation: *a*, exposed cases; *b*, exposed non-cases; *c*, unexposed cases; *d*, unexposed non-cases.

[b] Reported with 6 decimal digits

Table 2. Some recommendations on decimal digits for RR

1) Whenever possible, avoid reporting RR alone. In order to provide a straightforward interpretation, absolute measures of risk (expressed as rates) should be reported. When such a calculation is hampered by the study design (e.g., case-control studies), at least a qualitative perspective on the magnitude of the phenomenon under investigation should be provided.
2) Establish the number of decimal digits to be reported with RR before conducting the analysis, whenever possible. If a study protocol is presented in advance, the number of decimal digits should be stated. The RR cannot be known in advance, but clinical and public health considerations should be well known in advance.
3) Consider reporting the Exposure Impact Number and the Exposed Cases Impact Number. This information can provide a perspective on the relevance of the findings from observational studies.

Reference list

1. Siontis GC, Ioannidis JP. Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol*. 2011;40:1292-307.
2. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.
3. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008;19:640-8.
4. Shapiro S. Bias in the evaluation of low-magnitude associations: an empirical perspective. *Am J Epidemiol*. 2000;151:939-45.
5. Hertz-Picciotto I. Invited commentary: shifting the burden of proof regarding biases and low-magnitude associations. *Am J Epidemiol*. 2000;151:946-8.
6. Cole TJ. Setting number of decimal places for reporting risk ratios: rule of four. *BMJ*. 2015;350:h1845.
7. Cochrane Collaboration. Cochrane style guide. 2010. http://community.cochrane.org/style-manual/numbers-statistics-and-units/statistical-and-mathematical-presentation.
8. Gerstman BB. Data Analysis with Epi Info. 2000. http://www.sjsu.edu/faculty/gerstman/EpiInfo/.
9. Hopkins WG, Batterham AM, Pyne DB, Impellizzeri FM. Misplaced decimal places. *Scand J Med Sci Sports.* 2011;21:867-8.
10. Jacobs J, Kreutzer R, Smith D. Rice burning and asthma hospitalizations, Butte County, California, 1983-1992. *Environ Health Perspect*. 1997;105:980-5.
11. Shen Y, Huang X, Zhang W. Platelet-to-lymphocyte ratio as a prognostic predictor of mortality for sepsis: interaction effect with disease severity—a retrospective study. *BMJ open*. 2019;9:e022896.
12. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-15.
13. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemio*l. 2017;82:128-136.
14. Porta, M, ed. A dictionary of epidemiology. Oxford: Oxford University Press; 2014
15. Heller RF, Dobson AJ, Attia J, et al. Impact numbers: measures of risk factor impact on the whole population from case-control and cohort studies. *J Epidemiol Community Health* 2002;56:606–610.

16. Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, Costiniuk C, Blank D, Schünemann H. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst Rev*. 2011;16:CD006776.
17. King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. *BMJ*. 2012;345:e5774.
18. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine*. 2010;8:18.