have an estimate of excess risk of infinity with 1 control, an estimate of excess risk converging to 5.63 with 2 controls, then expected to decrease with the number of controls, converging (as both sample size and the number of controls diverge) to 2.8. Even if (as in the case of [1]) the number of events is not always higher among the treated than in the controls (so that there would not be the problem of an infinite value), taking more than one control would move the estimated excess risk toward 1, due to the issue described above of a similar quantity added to both the numerator and the denominator of a fraction.

Federico Tedeschi
Department of Neurosciences
Biomedicine and Movement Sciences
University of Verona
Verona, Italy
E-mail address: federico.tedeschi@univr.it

Reference

[1] Redelmeier D, Tibshirani R. Methods for analyzing matched designs with double controls: excess risk is easily estimated and misinterpreted when evaluating traffic deaths. J Clin Epidemiol 2018;98:117−22.

https://doi.org/10.1016/j.jclinepi.2018.11.010

Analyzing excess risk from matched designs with double controls: author's response

Tedeschi provides a thoughtful perspective on the interpretation and calculation of excess risk. We largely agree, and this was our rationale for raising the caution that excess risk is easily misinterpreted [1]. This was also our reason for providing a detailed exposition and concrete example to help explain the distinction to a larger audience of general medical readers. By contrast, clinical research studies sometimes publish findings expressed solely as an excess risk without fully considering these concerns [2].

The first concern from Tedeschi relates to the nature of underlying risk, whether outcomes reflect a stochastic or deterministic process, and how to consider the degree of individual heterogeneity. Such uncertainties underpin endless debates about the effect of random chance on patient outcomes and applying frequentist statistics to the care of individual patients [3]. We will not settle these debates anytime soon because the unknowns cannot be directly explored through a testable hypothesis.

DOI of original article: https://doi.org/10.1016/j.jclinepi.2018.11.010.
Conflict of interest: No conflict of interest.

A second concern from Tedeschi is on the idea that excess risk must require a comparator, thereby implying that different comparators can lead to different estimates. We agree and also endorse the additional point that an abundant collection of matched comparators could lead to particularly skewed estimates. We also agree that such discrepancies tend to be accentuated when the baseline count is small, excess count is large, variation is uneven, and comparisons are calculated as ratio statistics [4].

A tangential concern from Tedeschi is whether double controls are sufficient and what happens when a larger collection of matched comparators is available. Typically, an increase from solitary controls to double controls yields more statistical power, whereas an increase beyond seven matched controls provides minimal further gains. Higher assemblies of matched controls also raise pitfalls from mathematical complexity, missing data, and difficulties in visual displays. Double controls may be a useful compromise [5].

We agree with Tedeschi and reinforce the central point of consensus. Namely, that excess risk is easily estimated and easily misinterpreted. In addition, we underscore our main conclusion that, because implications can differ, a conservative approach can be to show results based only on total counts and not excess counts when analyzing a matched study. We thank Tedeschi for joining the conversation and providing a mathematical example that further illustrates the potential for misinterpretations.

Donald A. Redelmeier*
Department of Medicine
University of Toronto, Toronto
Canada

Robert J. Tibshirani
Department of Statistics
Stanford University
Stanford, CA, USA

*Corresponding author. University of Toronto, Sunnybook G-151, 2075 Bayview Ave, Toronto, Ontario, Canada. Tel.: 416-480-6999; fax: 416-480-6048.
E-mail address: dar@ices.on.ca (D.A. Redelmeier)

References

[1] Redelmeier DA, Tibshirani RJ. Methods for analyzing matched designs with double controls: excess risk is easily estimated and misinterpreted when evaluating traffic deaths. J Clin Epidemiol 2018;98:117−22.
[2] Jansen AG, Sanders EA, Wallinga J, Groen EJ, van Loon AM, Hoes AW, et al. Rate-difference method proved satisfactory in estimating the influenza burden in primary care visits. J Clin Epidemiol 2008;61:803−12.
[3] Rietbergen C, Debray TPA, Klugkist I, Janssen KJM, Moons KGM. Reporting of Bayesian analysis in epidemiologic research should become more transparent. J Clin Epidemiol 2017;86:51−8.

[4] Tedeschi F. A clarification on odds ratios as conditional risk ratios. J Clin Epidemiol 2018;97:122−3.
[5] Redelmeier DA, Tibshirani RJ. A simple method for analyzing matched designs with double controls: McNemar's test can be extended. J Clin Epidemiol 2017;81:51−5.

https://doi.org/10.1016/j.jclinepi.2018.11.009

"An error in an old paper illustrates the need for data/code archives"

To motivate their much-cited article, Kernan et al [1] presented their Table 1, herein reproduced. The purpose of the table is to illustrate the probability of covariate imbalance when drawing two samples from a common distribution in which a specific prognostic factor has a specified prevalence. For example, if a sample of 50 observations has 15 with the specified prognostic factor (let us call them "hits" for short) and it is divided into two subsamples of 25 each, it would be most desirable for subsample A to have eight hits and subsample B to have seven (or conversely); no closer balance can be obtained. In this situation, the percentage of hits in subsample A is 32% and the percentage of hits in subsample B is 28%, for a difference of 4%. As the text of the Kernan article makes clear, Table 1 is concerned with the probability that the difference will be more than 10%; a difference of exactly 10% is not considered extreme. Five numbers in the table are wrong, which are denoted by asterisks.

The number of hits in subsample A has a hypergeometric distribution. R code for computing the relevant probabilities is given by:

```
nobs < − 50 # total number of observations
hits < − 15 # number of desired items
myseq < − seq (0, hits)
dhyper (myseq, m = hits, n = nobs − hits,
k = nobs/2).
```

Once one recognizes that this is a hypergeometric problem, it is immediately obvious that the numbers 33% and 24% in Table 1 are problematic. Fifteen percent of 30 is 4.5 and 15% of 50 is 7.5. The hypergeometric distribution is not defined for noninteger numbers.

Let us consider the case when the total number of observations is 50 and the percentage of hits is 30%. In this case, there are 15 hits to be divided between the two subsamples. Only two possible allocations result in an absolute difference less or equal to 0.10: seven hits in subsample A (and therefore eight in subsample B), 7/25−8/25 = −0.04, which occurs with probability 0.231; or eight hits in subsample A (and seven in subsample B), 8/25 - 7/25 = 0.04, which also has a probability of 0.231. Therefore, the overall probability that the difference exceeds 0.10 is 1 - 0.462 = 0.538, markedly different from the claimed probability of 0.38. Similarly, it can be shown that 10% should be 9% and 27% should be 28%.

Table 1
Probability that simple randomization will result in two treatment groups that differ by more than 10% for the proportion of patients with a prognostic factor according the prevalence of the factor in the whole cohort[a]

Total no. of patients in the trial	Probability of 10% difference according to the prevalence of the prognostic factor in the entire study cohort	
	15%	30%
30	33%*	43%
50	24%*	38%*
100	10%*	27%*
200	3%	9%
400	0.3%	2%

Asterisk denotes incorrect number.
[a] Calculated in a simulation study involving 10,000 hypothetical trials.

When contacted by email for the simulation code, the corresponding author replied that he no longer had it. This is reasonable, given that the article is 20 years old. However, lost forever is knowledge of how the probabilities in Table 1 were calculated. Fortunately, the errors in this table do not affect the conclusions of the article. However, we have no assurance that such is the case for as-yet-undiscovered errors in other articles. The published article is just part of the scholarship and research that leads to the conclusions stated in the article. Behind that lies the data and code, which is just as crucial to understanding how results were obtained and enabling other researchers to build on what was published. If the article is really just part of the scholarship, why does not the journal make all the scholarship available to other researchers and require that both data *and* code be available? The replication literature has demonstrated repeatedly that data alone are insufficient to ensure reproducibility; the code must be available, too. There were no data for the Kernan article, but there certainly was code. If data have to be uploaded, it takes no additional effort or resources to also require that code be archived, too.

B.D. McCullough
Department of Decision Sciences & MIS
Drexel University
Philadelphia, PA 19104, USA
Department of Decision Sciences & MIS, Drexel University, GHall 742, Philadelphia, PA 19104, USA.
Tel.: 215 895 2134; Fax: 215 895 2907.
E-mail address: bdmccullough@drexel.edu

DOI of original article: https://doi.org/10.1016/S0895-4356(98)00138-3.
Conflict of interest statement: None.