Percents have perhaps one good use. When they represent fractions and we don't care to present but two decimal places of accuracy, i.e., the percents you calculate are all whole numbers, percents may be OK. But I would still prefer numbers like 0.02, 0.86 and to avoid a symbol (%) when just dealing with numbers.

Links to Other Resources

- [What is a percentage difference?](#) by TJ Cole and DG Altman

metrics 2018

Frank Harrell

Professor of Biostatistics

My research interests include Bayesian statistics, predictive modeling and model validation, statistical computing and graphics, biomedical research, clinical trials, health services research, cardiology, and COVID-19 therapeutics.

Statistical Thinking Comment Policy

Comments are welcomed. Be informed, informative, respectful, and criticize ideas, not people. Please read our Comment Policy before commenting.

Comments on this entire site are premoderated (only moderators can see this message). Change site settings.

7 Comments **Statistical Thinking** **Disqus' Privacy Policy** **Frank Harrell**

Favorite 4 Tweet Share Sort by Best

Join the discussion…

Jim Thompson • 5 years ago

Thank you your thoughts. A typo: 'treatment A multiplied the risk of stroke by 2⁄3 in comparison to treatment A.'
If I understood, the second A should be a B.

1 • Reply • Share ›

Mark Chatfield • 5 months ago

I came across this article by David Small that I think is worth reading. Int. J. Pharmacokinet. (2016) 1(1), 13–15. Let's abolish 'fold higher' and 'fold increase' from our lexicon. https://www.future-science....
Proposed solutions:
• the ratio of the victim drug's AUC with and without the perpetrator drug is 2.4
• the victim drug's AUC with the perpetrator drug was 2.4-fold the AUC without it
• the perpetrator drug increased the victim drug's AUC by 140%
"Admittedly some of these alternatives are not as pithy as a succinct 'fold increase', but they will not

be misinterpreted."

Reply • Share ›

Frank Harrell Mod → Mark Chatfield • 5 months ago

Beautiful - glad to know about this article! The first proposed solution ("the ratio ...") is my favorite. Perhaps another would be: "the drug multiplied the reference AUC by 2.4". I am writing up a statistical report right now and decided to try, in the case of simple proportions, labeling things as e.g. "# diseased / # with symptoms" instead of "proportion of symptomatic individuals who are diseased". Getting explicit and keeping it basic, and using shorter language are my new watchwords.

Reply • Share ›

Tim Cole • 5 years ago

It's important to distinguish between percentages and percentage differences. Percentages are fractions, proportions or probabilities rounded to integers with two significant digits, and this is an efficient way to communicate them. You ask "why not per thousand?", and the answer is that three significant digits of precision would be excessive whereas two are just right.1 It's important in general though to display the percentages as whole numbers - more than two significant digits add spurious precision. This matches the conclusion in your final paragraph.

Whether or not percentages should be applied to probabilities is clearly a matter of taste. But in epidemiological studies proportions are nearly always reported as percentages, and it is here that they are most useful presented as integers.

As regards percentage differences I, like you, dislike them because they are not symmetric nor are they additive. Doug Altman and I discussed this in a recent BMJ statistics note,2 where we pointed out a modified percentage scale which gives both symmetric and additive percentage differences, just like your ratios; the scale involves the 100 loge transformation.

Here is an example. What is the percentage difference between 3 and 4? Is it 25% or is it 33%? The answer is neither, it's 100 loge 4 - 100 loge 3 = 29%, midway between the two. And swapping the 3 and 4 changes just the sign but not the magnitude of the percentage difference - as is the case with the numbers themselves, where the difference is ±1. So it is a symmetric percentage difference, and in cases where the two numbers are not temporally ranked, it has much to commend it. The same transformation is also additive, in just the same way that ratios are multiplicative.

This form of percentage difference may appear somewhat arcane, but it has wide application. Percentage differences also crop up as percentage standard deviations and percentage regression coefficients, where again the 100 loge transformation is useful. The percentage SD is usually calculated as the coefficient of variation (%CV), but the SD of the 100 loge transformed variable is another form of CV. And the regression coefficient of a 100 loge transformed variable is also in percentage units per unit of the independent variable, which makes analyses involving log transformed dependent variables much simpler to present.

So our positions simplify to your preferring ratios, and my preferring log differences.

1. http://dx.doi.org/10.1136/a...
2. http://dx.doi.org/10.1136/b...

Reply • Share ›

Frank Harrell Mod → Tim Cole • 5 years ago • edited

Tim your comments add greatly to the discussion. On the notion of 2 significant digits is enough, that's true in most situations when the proportion is between 0.15 and 0.85 but in other situations we often need 3 digits. I still think one should avoid % in all situations, primarily because you often have to look closely at the context to tell whether the % is absolute or relative, and even more because I don't like adding a punctuation sign to a number. To me the cognitive cost of adding an extra symbol is more when that symbol is not a decimal point.

Reply • Share ›

A Student of Statistics • 5 years ago

You are thinking of numbers like a mathematician. But numbers in research results are meant to communicate information to laypeople, too.

Laypeople understand percent better than probabilities. To the average person, the meaning of "10% of the population have AIDS" is clear, while "the probability for AIDS is 0.1" is almost incomprehensible to most. In school, kids learn about percentages earlier than about probabilities – a clear sign that pedagoges find percentages easier to teach (and kids find them easier to grasp).

And 100 is not random in most cultures. A meter has 100 centimeters. Most currencies have a monetary unit (cent) that is one hundredth of the basic monetary unit. Hundred meters is how far you have to run. 100 is the limit for how far up kids go when they learn to multiply. Small children are proud when they can count to 100 (not to a score or a dozen or a gross). There are other significant numbers, of course, e.g. 60 seconds, but kids conceptualize time and learn about time measurement much later than about lengths, so units of length are much more fundamental to our thinking than units much later than about lengths, so units of length are much more fundamental to our thinking than units of time. (Perhaps those that grew up using the imperial system do not readily perceive the fundamental experience of one hundred in childhood development, as their money and lengths are not divided into "cents" and "centimeters" but in yards, feet, and inches or shilling and pence.)

One is seen by most people as the basic unit ("one person", "one second"), not as a maximum ("everybody", "always"). The idea that "nothing", "no one", and "never" an be expressed by a Zero makes sense, but the idea that "everything", "everybody", and "always" should be expressed by what commonly stands for one single unit doesn't make sense at all. At least 100 is a lot, while 1 is the opposite of much.

In fact there is "per thousand" (called "per mille" or "permille"). Every car driver in Europe is familiar with that term as it is the basis for how much you can drink and drive without going to jail.

Finally, percent is in fact a proportion (per cent = per hundred). Percent is a normalized proportion, and as a statistician you should love normalization ;-)

Reply • Share ›

Frank Harrell Mod ➜ A Student of Statistics • 5 years ago

Thanks for taking the time to write such a full and informative comment. I can see you point about "100". As a statistician I am not always a fan of normalization though. I would find this argument slightly more compelling: normalizing by 100 allows us to deal with integers a good deal of the time. Regarding your statement that "0.1 is almost incomprehensible to most", I find that a bit hard to digest. But I wish I had survey data to back up my feelings. We are taught decimal fractions early in elementary school, and I don't see that understanding 0.1 requires even intermediate math skills. We all have to deal with it when looking at a bank statement and when we pay cash for small items. I would love to learn more about perceptions of 0.1 vs. 10% and of 0.005 vs. 0.5%.

Reply • Share ›

Related

- [Viewpoints on Heterogeneity of Treatment Effect and Precision Medicine]
- [Commentary on Improving Precision and Power in Randomized Trials for COVID-19 Treatments Using Covariate Adjustment, for Binary, Ordinal, and Time-to-Event Outcomes]
- [If You Like the Wilcoxon Test You Must Like the Proportional Odds Model]
- [Violation of Proportional Odds is Not Fatal]
- [The Burden of Demonstrating HTE]