Failed "Normalization": Dividing by the Control Average

Tatsuki Koyama, Ph.D. and Alex Zhao, M.S.

Cancer Biostatistics Center, Vanderbilt-Ingram Cancer Center

Vanderbilt-Ingram Cancer Center

Abstract

Experiments in basic science often have a specific control group for each of the treatment groups. The observed values in treatment groups are divided by the average of the corresponding control group in an attempt to make the treatment groups comparable. The resulting values are sometimes called fold changes. Despite the overwhelming popularity of this method, dividing by the control group average has multiple serious issues from a statistical point of view. A statistically sound, yet simple, alternative is the use of regression models with indicator variables for group of interest and for treatment. Then the test for difference in treatment effects, accounting for differences between the control groups, is simply a test for interaction terms in the regression model.

1 Dividing by the control average is a problem.

1.1 Multiplicative or additive effect?

Dividing by the control average may mask an additive effect. See section 2.1.

1.2 Ignored variability

The average of the control group is a random variable that varies from sample to sample. When it is treated like a constant, its variability (uncertainty) is ignored, and as a result, type I error rate gets inflated. See section 2.2.

Q: *"But variability in the control group is small and negligible!"*

A: It may be small, but it is rarely small enough to be negligible. See section 2.2.

1.3 Weird distribution

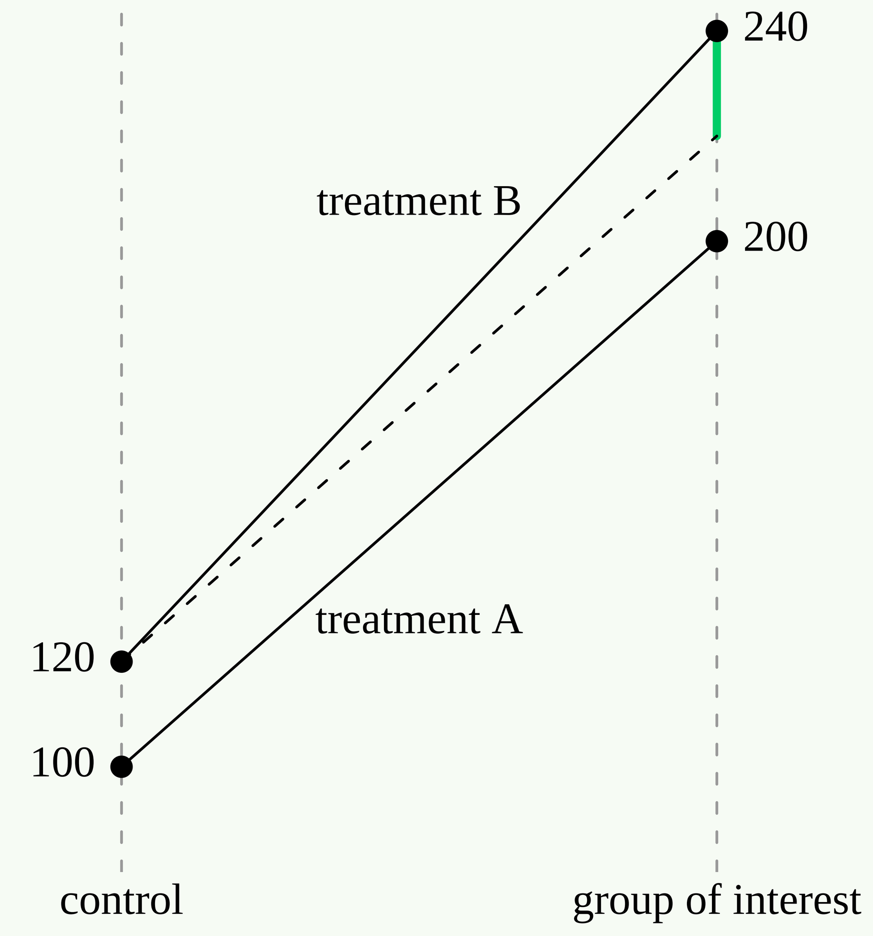
When a random variable is divided by another random variable, the resulting values may have a strange distribution. The ratio of 2 independent normal distributions is some variant of a Cauchy distribution, which is not very nice.

For one thing, the central limit theorem does not apply to a Cauchy distribution; having a large sample size does not make the estimate of the average any better. See section 2.3.

2 How bad are these problems?

2.1 Preliminary question

"Do you really want to look at fold change, not the difference?"



If the true effects are as shown in this plot, the green segment is the difference in the group of interest after taking into consideration the control group difference. By contrast, the fold change is 2 for both treatments.

2.2 Ignored variability in the control group

In this simple simulation study, we counted the number of times the null hypothesis, "Fold changes are the same.", was incorrectly rejected by a t-test in 100,000 simulations.

Control Group		Group of Interest		type I error	
n	sd	n	sd	fold change	regression
3	4	10	4	58.0%	9.5%
3	2	10	4	32.4%	7.6%
3	1	10	4	13.5%	5.6%

Type I error rate of 58.0%! The main reason for such a severely inflated type I error rate is ignored variability in the control group. As the standard deviation of the control group decreases, the type I error rate decreases; however, even when it is only $1/4$ of the standard deviation of the group of interest, the type I error rate is still 13.5%.

2.3 Increasing sample size won't help.

Because the central limit theorem does not apply, increasing the sample size does not lead to a better estimate of the true fold change. Even with the sample sizes 100 (control) and 200 (group of interest), the type I error rate is not much better.

Control Group		Group of Interest		type I error	
n	sd	n	sd	fold change	regression
100	4	200	4	51.2%	5.0%
100	2	200	4	25.9%	5.0%
100	1	200	4	11.0%	5.0%

3 A solution

A simple linear regression model can be fit to test for the difference between the groups of interest taking into account the difference in the control groups. It is simply a test for an interaction effect. If you really, truly believe effects are multiplicative, you can take the logarithm of all the numbers before fitting a regression model.

The type I error rates for this approach are also shown in the tables above. The regression approach seems much better, but the type I error rates are still somewhat inflated in the first table. The reason?

Sample size of 3 is not enough to estimate the control group average!

Unlike the case with fold change method, increasing the sample sizes will improve the regression approach. Doubling from 3 to 6 will result in some improvement as seen below.

Control Group		Group of Interest		type I error	
n	sd	n	sd	fold change	regression
6	4	10	4	45.4%	6.1%
6	2	10	4	21.4%	5.5%
6	1	10	4	9.0%	5.0%

4 In Closing

Do not divide by a number that is not a constant.

Even if you think the variability is (should be) miniscule, it is probably not small enough.

Think about all the false-positive findings that have resulted from the 'dividing by the control average' method!

Dividing by a baseline measurement to compute percent change probably suffers from a similar problem. Body mass index (weight / height^2) probably suffers from a similar problem, and so does waist-to-hip ratio. And so does any kind of ratio of random variables. That's too big a problem to tackle for now. So, ...

Do not divide by the control average.

tatsuki.koyama@vanderbilt.edu