Editorial

Open Access

# Comparing two quantities by using a ratio

It is often useful to compare multiple subgroups to assess meaningful differences. The purpose of this editorial is to summarize a method for using a ratio to detect significant differences in associations across multiple population subgroups. In Statistics, two quantities can be compared by taking difference (for example, differences between two means: $\mu_1 - \mu_2$, or difference between two proportions: $p_1 - p_2$), or taking ratios (for example, ratios of two proportions: $\dfrac{p_1}{p_2}$, also known as a relative risk: (RR)). Also, odds ratio (OR), incidence rate ratio (IR) or hazard ratio (HR) can be used to compare two groups, such as treatment vs control, or exposed vs unexposed. The standard error of the ratio should be computed if we conduct ratio analysis. For example, to compute the confidence interval for the estimated odds ratio we need to compute the standard error (SE) of log(OR):

$$\hat{\sigma}\left(\log\left(OR\right)\right) = \sqrt{\left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)},$$

where $n_{ij}$ are sometimes replaced by $n_{ij} + 0.5$, when some of the $n_{ij}$ are zero. This result can be derived under the assumption of multinomial sampling by using the Delta Method. When RR is used to compare two quantities, the log transformation is conducted, that is $\log\left(\hat{\pi}_{1|1} / \hat{\pi}_{1|2}\right)$ is often considered instead of $\left(\hat{\pi}_{1|1} / \hat{\pi}_{1|2}\right)$, since the former has a sampling distribution which is closer to normal than that of the latter. The estimated asymptotic standard error (ASE) of log(RR):

$$\hat{\sigma}\log\left(\hat{\pi}_{1|1} / \hat{\pi}_{1|2}\right) = \sqrt{\left[\frac{\left(1 - \hat{\pi}_{1|1}\right)}{\hat{\pi}_{1|1} n_{1+}} + \frac{\left(1 - \hat{\pi}_{1|2}\right)}{\hat{\pi}_{1|2} n_{2+}}\right]}$$

This result can be derived under the assumption of independent binomial sampling using the Delta Method, where $\hat{\pi}_{1|1}$ and $\hat{\pi}_{1|2}$ are sample proportions based on independent binomial samples with success probabilities $\pi_{1|1}$ and $\pi_{1|2}$, respectively. The confidence interval (CI) of log(RR) (Wald CI for log $\left(\pi_{1|1} / \pi_{1|2}\right)$) can be calculated as follows:

$$\log\left(\hat{\pi}_{1|1} / \hat{\pi}_{1|2}\right) \pm z_{\alpha/2}\left(\hat{\sigma}\left(\log\left(\hat{\pi}_{1|1} / \hat{\pi}_{1|2}\right)\right)\right).$$

The CI tends to be slightly conservative (i.e., the actual coverage probability tends to be higher than the nominal level). Exponentiating the endpoints provides a CI for $\left(\pi_{1|1} / \pi_{1|2}\right)$.

Now we discuss applying Delta Method to estimate the SE of a trans-formed parameter. The delta method, in its essence, expands a function of a random variable about its mean. Usually with a one-step Taylor approximation, and then takes the variance. For example, if we want to approximate the variance of G(X) where X is a random variable with mean $\mu$ and function G is differentiable, we can try

$$G\left(X\right) \approx G\left(\mu\right) + \left(X - \mu\right)G'\left(\mu\right),$$

$$\mathrm{V}ar\left(\mathrm{G}\left(\mathrm{X}\right)\right) \approx \mathrm{G}'\left(\mu\right)Var\left(\mathrm{X}\right)\left[\mathrm{G}'\left(\mu\right)\right]^T.$$

This idea can easily be expanded to vector-valued functions of random vectors.

Shimin Zheng,[1] Michael Smith[2]
[1]Department of Biostatistics and Epidemiology, East Tennessee State University, USA
[2]Department of Health Services Management & Policy, East Tennessee State University, USA

**Correspondence:** Shimin Zheng, Department of Biostatistics and Epidemiology, East Tennessee State Uni-versity, Box 70259, Johnson City, TN 37614, USA, Tel 423-439-4443, Email ZHENGS@mail.etsu.edu

$$\mathrm{V}ar\left(\mathrm{G}\left(\mathrm{X}\right)\right) \approx \mathrm{G}'\left(\mu\right)Var\left(\mathrm{X}\right)\left[\mathrm{G}'\left(\mu\right)\right]^T.$$
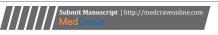
The Delta Method can be applied to Random effects meta-regression analysis, which can be used to investigate factors associated with the magnitude of the ratio of RR (or OR IR HR) (RRR). The triple R method can easily be extended to the quadruple R, the magnitude of the ratio of the ratio of RR (or OR IR HR), or RRRR.

To compare subgroups within selected studies, the natural logarithm transformation of the ratio of RR (or OR IR HR) values (RRR; or analogous estimates of association) for the two compared subgroups should be used. Since the logarithm of RRR has a sampling distribution which is closer to normal than RRR. To find the estimate of the RRR and its CI, the SE of RRR can be derived using the Delta Method. On the other hand, for meta-analysis fixed-model, we assume that there is no heterogeneity between the studies. The model assumes that within-study variances may differ, but that there is homogeneity of effect size across all studies. Often the homogeneity assumption is unlikely and variation in the true effect across studies is to be expected. Therefore, caution is required when using this model.[1,2]

For meta-analysis random-effects model (the most commonly used), we assume that models heterogeneity between the studies, or we assume that the true effect can be different for each study. For example, the effect estimates of urban and rural subpopulation was evaluated. First, the log transformation was conducted: the natural logarithm of the ratio of RR values (RRR; or analogous estimates of association) for the two subgroups, i.e., RR(rural)/RR(urban), a method given by Benmarhnia et al.[3] The formula used to calculate the standard errors of the ratios is as follows (adopted from Benmarhnia et al.[3]):

$$SD\left(ratio\right) = ratio \times \sqrt{\left(\frac{SD\,RR_{rural}^{2}}{RR_{rural}}\right) + \left(\frac{SD\,RR_{urban}^{2}}{RR_{urban}}\right)}$$

Based on the formula above, Li et al.[4] compared the heat-related mortality between rural and urban populations using the RRR method and found that there was no statistically significant difference between the two subgroups.

## Acknowledgments

None.

## Conflicts of interest

None.

## References

1. Oehlert G W. A note on the delta method. *American Statistician*. 1992;27–29.

2. Rice J. A Mathematical Statistics and Data Analysis, (1994; 2nd ed. Duxbury.)

3. Benmarhnia T. A Vulnerability to Heat-related Mortality - A System-atic Review, Meta-analysis, and Meta-regression Analysis. *Epidemiology.* 2015;26:781–793.

4. Li Y, Odame EA, Silver K, et al. Comparing Urban and Rural Vulnerability to Heat-Related Mortality: A Systematic Review and Meta-Analysis. *Journal of Global Epidemiology and Environmental Health*. 2017;9-15.