On-line supplement

1. Definitions and models

The risk ratio, also known as the relative risk (RR), is defined as the ratio of the risk, probability, or prevalence of a health outcome of interest in the exposed/treated/intervention group (r_1), divided by the same in the unexposed or control group (r_0). The risk difference (RD) subtracts the health outcome risk in the control group from the health outcome risk in the intervention group. That is,

$$RR = r_1/r_0 \text{ and } RD = r_1 - r_0.$$

As a relative measure of effect, the relative risk is most directly estimated by the multiplicative model, when it fits the data. The risk difference is an absolute measure of effect, most directly estimated by the additive model, when it fits the data. Although much of what is discussed here will also apply to prevalence ratios and differences, the prevalence as the measure of disease frequency is susceptible to reverse-causation bias, although the occurrence of this may be less common than often taught in standard epidemiology texts. Cumulative incidences, risks and proportions are synonyms, although the cumulative incidence is typically estimated using survival data analysis methods to allow for censoring, staggered enrollment and/or competing risks. Similarly, rates, e.g. mortality rates or disease incidence rates, are used as outcome measures when censoring, staggered enrollment and/or competing risks are in play. Their primary disadvantage is more difficulty in interpretability, as they require units of person-time, e.g. person-years or person-months. These units can be difficult to explain to non-technical audiences, the target audiences for the results of public health interventions. However, the use of risks as the primary outcome measure does not completely get us out of the woods concerning subtlety in interpretation, as the value of a risk depends critically upon the duration of follow-up over which it is calculated. The classic, and most extreme example of this phenomena, is when the risk is that of mortality. In the limit, regardless of how beneficial or harmful an intervention is, the ratio will go to 1, and the difference to 0, since we will all die sometime. When other outcomes are eventually possible – that is, in the presence of competing risks – this limiting property will likely not occur, but the dependency of risks based on different durations of follow-up, and the resulting potential for non-comparability, remains.

In an individually randomized intervention, straightforward methods for a single 2x2 table can be used to estimate relative risks and risk differences, since there is no need to adjust for confounding. Alternatively, in an individually randomized intervention design, the risk ratio can be modeled on the multiplicative scale as

$$\log[\mathrm{E}(Y_i = 1|X_i)] = \log[\Pr(Y_i = 1|X_i)] = \beta_0 + \beta_1 X_i, \text{ [1]}$$

where Y_i is the binary outcome upon which the intervention is focused, X_i is 1 if the participant was randomized to the intervention and 0 otherwise, e^{β_1} is the relative risk, which will be <1 if the outcome is undesirable and the intervention is preventive, e^{β_0} is the risk in the control group, and as is common in science, log is used interchangeable with ln to denote the natural

logarithm on the base e scale. $E[\cdot]$ denotes the expected value, or mean, function, and for binary data is equivalent to the outcome model probability function. Alternatively, if the difference measure of is of interest, the risk difference can be modeled on the additive scale as

$$\mathrm{E}(Y_i = 1|X_i) = \Pr(Y_i = 1|X_i) = \alpha_0 + \alpha_1 X_i, \text{ [2]}$$

where the risk difference is α_1. The parameters of models [1] and [2] have a one to one correspondence, so by standard statistical theory, the corresponding transformation of the maximum likelihood estimates of one model to the parameters of the other will also be maximum likelihood. Thus, from the point of view of validity, in individually randomized studies with no loss to follow-up, staggered entry or competing risks, the choice between the ratio or difference measure as the parameter of interest, that is, the choice between model [1] and [2] doesn't matter, and $\hat{\alpha}_0 = e^{\hat{\beta}_0}$ and $\hat{\alpha}_1 = e^{\hat{\beta}_0+\hat{\beta}_1} - e^{\hat{\beta}_0}$.

Things change when confounding needs to be considered. As discussed in a previous column in this series(1), in cluster-randomized studies as would typically be utilized in the evaluation of public health interventions when randomization of any kind is possible, unless there are a large number of clusters, which is unusual, or outcome rates between clusters are relatively constant, as might be more commonplace, residual between-cluster confounding is likely. Then, in order to validly estimate the intervention effects, models [1] and [2] need to be expanded:

$$\log\left[E(Y_{ij} = 1|X_i, C_{1ij}, \ldots, C_{pij})\right] = \beta_0 + \beta_1 X_i + \beta_2 C_{1ij} + \cdots + \beta_{p+1} C_{pij} \text{ [3]}$$

where $C_{1ij}, \ldots, C_{pij}$ are the p covariates measured in the study that are needed to validly estimate the intervention effect, the relative risk, e^{β_1}, for the i^{th} participant in cluster j. A similar model could be fit if the risk difference were the parameter of interest

$$E(Y_{ij} = 1|X_i, C_{1ij}, \ldots, C_{pij}) = \alpha_0 + \alpha_1 X_i + \alpha_2 C_{1ij} + \cdots + \alpha_p C_{p+1ij} \text{ [4]}.$$

2. Logistic regression – rarely, does non-collapsibility matter; ditto for the Cox model

Fitting binary and nominal outcome variables to models with log and identity link functions have well-known numerical instabilities, meaning that the models often do not converge or provide extreme estimates with inflated variances. Solutions to these issues have been provided(2-4), yet it is our experience that even with these, numerical issues can prove to be insurmountable and the logistic link function is needed to obtain any estimate at all. In addition to the above-mentioned lack of intrinsic interest in the odds ratio as an estimate of effect, a more recent concern about logistic regression models has been raised – that of non-collapsibility. Non-collapsibility is the undesirable property that the on the logistic scale, a study in which the intervention and control groups are exactly balanced in their distributions of other risk factors – the usual condition for a lack of confounding – the crude odds ratio unadjusted for these covariates will not equal the fully adjusted one (5). This occurs because covariate balance is scale-dependent, and the criteria for achieving collapsibility on the logistic scale does not, for mathematical reasons related to the intrinsic non-linear nature of the logistic function, equate

with covariate balance. Fortunately, it is our experience that this concern about non-collapsibility is over-rated in practice, and that, unless the outcome under study is quite common and/or the intervention effect is strong, non-collapsibility generally does not result in substantively important bias, even greater than 10% (6, 7). However, important examples where the logistic approximation led us astray have been given (8, 9).

Nevertheless, we always recommend 'doing the right thing' whenever possible, and in this case, there is no reason, given modern software capabilities, to fit the model that provides an estimate of the parameter of interest, rather than an approximation to it (10). Importantly, neither the identity link nor the log link functions have this collapsibility issue, with respect to the parameter of interest, the risk difference and ratio, respectively (11).

Hazard ratios estimated from the Cox model for survival data analysis have features similar to those described above for logistic regression. Hence, it appears that the non-collapsibility issue is again typically not important when survival data analysis methods are needed for estimation of the rate ratio or difference measure (12), because there is staggered entry into the study, competing risks, or loss to follow-up. Under these circumstances, parameters of the survival model, such as baseline rates/hazards, and rate/hazard ratios and differences, can be converted into risk ratio and differences for pre-specified durations of follow-up. As long as there is no effect modification by the time scale on which the survival model has been fit, rate ratios and differences have the useful property of being independent of the duration of follow-up, unlike risk ratios and differences.

1. Spiegelman D. Evaluating Public Health Interventions: 2. Stepping Up to Routine Public Health Evaluation With the Stepped Wedge Design. Am J Public Health 2016;106(3):453-7.
2. Yelland LN, Salter AB, Ryan P. Relative Risk Estimation in Randomized Controlled Trials: A Comparison of Methods for Independent Observations. The International Journal of Biostatistics 2011;7(1):1-31.
3. Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RH. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ 2012;184(8):895-9.
4. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005;162(3):199-200.
5. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Statistical Science 1999;14(1):29-46.
6. Holcomb WL, Jr., Chaiworapongsa T, Luke DA, Burgdorf KD. An odd measure of risk: use and misuse of the odds ratio. Obstet Gynecol 2001;98(4):685-8.
7. Katz KA. The (relative) risks of using odds ratios. Arch Dermatol 2006;142(6):761-4.
8. Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. New England Journal of Medicine 1999;341(4):279-283.
9. Schulman KA, Berlin JA, Harless W, Kerner JF, Sistrunk S, Gersh BJ, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. N Engl J Med 1999;340(8):618-26.

10. D S, E H. The SAS %RELRSK macro. In; 2012.
11. Ritz J, Spiegelman D. Equivalence of conditional and marginal regression models for clustered and longitudinal data. Statistical Methods in Medical Research 2004;13(4):309-323.
12. Nevo D, Liao X, Spiegelman D. Estimation and inference for the mediation proportion. Submitted 2017.