Converting Odds Ratio to Relative Risk in Cohort Studies with Partial Data Information

Zhu Wang

Connecticut Children's Medical Center

Abstract

In medical and epidemiological studies, the odds ratio is a commonly applied measure to approximate the relative risk or risk ratio in cohort studies. It is well known tha such an approximation is poor and can generate misleading conclusions, if the incidence rate of a study outcome is not rare. However, there are times when the incidence rate is not directly available in the published work. Motivated by real applications, this paper presents methods to convert the odds ratio to the relative risk when published data offers limited information. Specifically, the proposed new methods can convert the odds ratio to the relative risk, if an odds ratio and/or a confidence interval as well as the sample sizes for the treatment and control group are available. In addition, the developed methods can be utilized to approximate the relative risk based on the adjusted odds ratio from logistic regression or other multiple regression models. In this regard, this paper extends a popular method by Zhang and Yu (1998) for converting odds ratios to risk ratios. The objective is novelly mapped into a constrained nonlinear optimization problem, which is solved with both a grid search and a nonlinear optimization algorithm. The methods are implemented in R package **orsk** which contains R functions and a Fortran subroutine for efficiency. The proposed methods and software are illustrated with real data applications.

Keywords: odds ratio, relative risk, nonlinear optimization, grid search, multiple roots, R.

1. Introduction

Investigators of medical and epidemiological studies are often interested in comparing a risk of a binary outcome between a treatment and control group, or between exposed and unexposed. Such an outcome can be an onset of a disease or condition. In this context, the study results may be summarized in Table 1 and the odds ratio and relative risk are the important measures in cohort studies. In a case-control study, the odds ratio is often used as a surrogate for the relative risk. The odds ratio is the ratio of the odds of outcome occurring in the treatment

Group	Number of outcome	Number of outcome free	Total
Treatment	n_{11}	n_{10}	$ntrt$
Control	n_{01}	n_{00}	$nctr$

Table 1: Compute the odds ratio and the relative risk.

group to the odds of it occurring in the control group. The odds of outcome in the treatment group is $\frac{n_{11}}{n_{10}}$ and the odds of outcome in the control group is $\frac{n_{01}}{n_{00}}$. The odds ratio thus becomes

$$\theta(n_{01}, n_{11}) = \frac{n_{11}n_{00}}{n_{10}n_{01}}. \tag{1}$$

The odds ratio evaluates whether the probability of a study outcome is the same for two groups. An odds ratio is a positive number which can be 1 (the outcome of interest is similarly likely to occur in both groups), or greater than 1 (the outcome is more likely to occur in the treatment group), or less than 1 (the outcome is less likely to occur in the treatment group). The odds ratio can approximate the relative risk or risk ratio, which is a more direct measure than the odds ratio. In fact, the most direct way to determine if an exposure to a treatment is associated with an outcome is to prospectively follow two groups, and observe the frequency with which each group develops the outcome. The relative risk compares the frequency of an outcome between groups. The risk of the outcome occurring in the treatment group is $\frac{n_{11}}{n_{11}+n_{10}}$ and the risk in the control group is $\frac{n_{01}}{n_{01}+n_{00}}$. The relative risk is the ratio of the probability of the outcome occurring in the treatment group versus a control group, and is naturally estimated by $\frac{n_{11}}{n_{11}+n_{10}} / \frac{n_{01}}{n_{01}+n_{00}}$. It can be easily shown that the odds ratio is a good approximation to the relative risk when the incidence or risk rate is low, for instance, in rare diseases, and can largely overestimate the relative risk when the outcome is common in the study population (Zhang and Yu 1998; Robbins, Chao, and Fonseca 2002). Although it is well-known that the two measures evaluate different quantities in general, the odds ratio has been mis-interpreted as the relative risk in some studies, and thus contributed to incorrect conclusions (Schulman *et al.* 1999; Schwartz, Woloshin, and Welch 1999; Holcomb, Chaiworapongsa, Luke, and Burgdorf 2001). For this reason, many methods have been proposed to approximate the risk ratio, particularly in logistic or other multiple regression models. For instance, see a popular method in Zhang and Yu (1998). The formula in Zhang and Yu (1998) requires the proportion of control subjects who experience the outcome. Specifically, derived from the definition of the odds ratio and the relative risk, the approximated risk ratio is

$$\widehat{RR} = \frac{\text{odds ratio}}{1 - \text{risk}_0 + \text{risk}_0 \times \text{odds ratio}}, \tag{2}$$

where risk_0 is the risk of having a positive outcome in the control or unexposed group (i.e., $\text{risk}_0 = \frac{n_{01}}{nctr}$). Equation 2 can be utilized for both the unadjusted and adjusted odds ratio. The formula can also be employed to approximate the lower and upper limits of the confidence interval. For an interested reader, thus, the formula provides a conversion between the relative risk and odds ratio from the published data. However, it may not be possible to convert the estimate when risk_0 is unknown or cannot be estimated from the data.

To convert the adjusted odds ratio, this paper extends the work in Zhang and Yu (1998) to the scenario when risk_0 cannot be trivially estimated using the published data. In addition,

	Unadjusted odds ratio	95% CI
Use of epidural anesthesia ($n = 2601$)	2.61	2.25–3.03
Non-use of epidural anesthesia ($n = 1636$)	Reference	Reference

Table 2: Unadjusted odds ratio for the first stage of labor lasting > 12 hours and the 95% confidence interval (CI) (Szal *et al.* 1999).

	Adjusted odds ratio	95% CI
Use of epidural anesthesia ($n = 2601$)	2.25	1.92–2.63
Non-use of epidural anesthesia ($n = 1636$)	Reference	Reference

Table 3: Adjusted odds ratio from multiple logistic regression for the first stage of labor lasting > 12 hours and the 95% confidence interval (CI) (Szal *et al.* 1999).

the proposed methods can be applied to the unadjusted odds ratio. The problem under investigation will be described using a concrete example. A retrospective cohort study collected data on 4237 primiparous women (Szal, Croughan-Minihane, and Kilpatrick 1999). Of interest is the association between the use of epidural anesthesia and prolonged first stage of labor (> 12 hours). Often the published results include both unadjusted and adjusted estimates, as in Tables 2 and 3, so that readers "can compare unadjusted measures of association with those adjusted for potential confounders and judge by how much, and in what direction, they changed" (Vandenbroucke *et al.* 2007, item 16(a)). Sometimes the results are mis-interpreted in that women who used epidural anesthesia had 2.61 times (or 2.25 times, adjusting for other factors) the risk of the first stage of labor lasting > 12 hours than those who did not use epidural anesthesia. However, Szal *et al.* (1999) did not describe how many epidural anesthesia users and non-users had the first stage of labor lasting > 12 hours. Thus $risk_0$ is not conveniently available to approximate the relative risk. If we can reconstruct Table 1 based on Table 2, then it is possible to estimate the risk of the study outcome in the control and treatment groups. Completely or at least partially reconstructing Table 1 is also relevant to other applications. For instance, when Holcomb *et al.* assessed 112 clinical research articles in obstetrics and gynecology to determine how often the odds ratio differs substantially from the relative risk estimates, they had to exclude five articles due to lack of information on risk of study outcome in the control group, using Equation 2. More importantly, it remains unclear how accurate the odds ratios approximate the relative risks in the omitted studies. To the author's knowledge, methodologies have not been proposed to estimate $risk_0$ when not all data information is directly available. The proposed methods can reconstruct Table 1, and consequently estimate $risk_0$. In this sense, we extended the work in Zhang and Yu (1998) in the event where $risk_0$ is not directly available. Table 2 will be utilized in this paper to demonstrate the approximation of the risk ratio based on partial data information. Furthermore, with the estimated $risk_0$ and the adjusted odds ratio from the multiple logistic regression in Table 3, we can approximate the risk ratio.

The method developed in this paper is implemented in R (R Core Team 2013) package **orsk** (odds ratio to relative risk; Wang 2013). The paper is organized as follows. Section 2 proposes a nonlinear objective function which measures the similarity between the calculated odds ratio and the reported odds ratio. Two methods are proposed to optimize the nonlinear objective function. Section 3 outlines the implementations in **orsk**. Section 4 illustrates the capabilities of **orsk** with real data provided in Tables 2 and 3. Finally, Section 5 concludes the paper.

2. Methods

We briefly review some additional results of the odds ratio, which form the basis for the methodology introduced in this section. The **orsk** procedure relies on the fact that odds ratios have been reported based on the normal approximation, which is the most common practice in many statistical software programs. An asymptotic $(1 - \alpha)$ confidence interval for the log odds ratio is $\log(\theta(n_{01}, n_{11})) \pm z_{\alpha/2}SE$, where $z_{\alpha/2}$ is the $\alpha/2$ upper critical value of the standard normal distribution and the standard error SE can be estimated by $\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}}$. The lower bound of the confidence interval of the odds ratio can be calculated by $\theta_L(n_{01}, n_{11}) = \exp(\log(\theta(n_{01}, n_{11})) - z_{\alpha/2}SE)$. Therefore,

$$\theta_L(n_{01}, n_{11}) = \theta(n_{01}, n_{11}) \exp\left\{-z_{\alpha/2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}}\right\}. \tag{3}$$

Similarly, the upper bound of the confidence interval of the odds ratio is

$$\theta_U(n_{01}, n_{11}) = \theta(n_{01}, n_{11}) \exp\left\{z_{\alpha/2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}}\right\}. \tag{4}$$

Now, the problem to be solved can be stated as follows. In the context of Table 1, suppose $\theta^{(0)}, \theta_L^{(0)}, \theta_U^{(0)}$ are calculated by Equations 1, 3 and 4, respectively, and $nctr, ntrt$, and α are fixed. The aim is to estimate (n_{01}, n_{11}) and subsequently estimate the relative risk and its corresponding confidence interval. In the layout of Table 2, we have $nctr = 1636, ntrt = 2601, \theta^{(0)} = 2.61, \theta_L^{(0)} = 2.25, \theta_U^{(0)} = 3.03, \alpha = 0.05$. The task is to solve different sets of nonlinear equations for two unknowns (n_{01}, n_{11}) given that $n_{01} + n_{00} = nctr$ and $n_{11} + n_{10} = ntrt$: (i) Equations 1 and 3; (ii) Equations 1 and 4; (iii) Equations 3 and 4; (iv) Equations 1 to 4. The proposal is to select (n_{01}, n_{11}) through minimizing the sum of squared logarithmic deviations between the reported estimates $\theta^{(0)}, \theta_L^{(0)}, \theta_U^{(0)}$ and the corresponding would-be-estimates based on assumed n_{01} and n_{11}. For instance, in scenario (iv), consider the sum of squares SS defined below:

$$SS(n_{01}, n_{11}) = \left\{\log(\theta(n_{01}, n_{11})) - \log(\theta^{(0)})\right\}^2 + \left\{\log(\theta_L(n_{01}, n_{11})) - \log(\theta_L^{(0)})\right\}^2$$
$$+ \left\{\log(\theta_U(n_{01}, n_{11})) - \log(\theta_U^{(0)})\right\}^2. \tag{5}$$

Similar sums of squares can be considered with point estimate and lower or upper confidence interval bounds, or with confidence interval bounds only. The goal now is to solve the following optimization problem:

$$\min_{n_{01}, n_{11}} SS(n_{01}, n_{11}) \text{ for integer } n_{01}, n_{11}, 1 \leq n_{01} \leq nctr - 1, 1 \leq n_{11} \leq ntrt - 1. \tag{6}$$

Apparently SS will be very close to 0 for the true value of (n_{01}, n_{11}), and a smaller SS implies a better solution. Thus SS plays a role similar to the residual sum of squares in the linear regression. Implementing different objective functions in a variety of scenarios provides a means of cross-checking results. Ideally, the solutions should be insensitive to the choice of the objective function.

To solve the constrained optimization problem, we consider two approaches: the exhaustive grid search and a numerical optimization algorithm. In the first algorithm, the minimization

can be performed as a two-way grid search over the choice of (n_{01}, n_{11}). In other words, one can evaluate all the values $SS(n_{01}, n_{11})$, for $n_{01} \in \{1, 2, \ldots, nctr-1\}, n_{11} \in \{1, 2, \ldots, ntrt-1\}$. This will result in a total number of $(nctr-1)(ntrt-1)$ SS values to be sorted from the smallest to the largest; this implies that the computational demand can be high when $(nctr-1)(ntrt-1)$ is large. To make the algorithm more efficient, we adopt a filtering procedure. Specifically, we filter out SS if $SS > \delta$ for a prespecified small threshold value δ, with a default value 10^{-4}. As a result, a smaller threshold value δ can lead to sparser solutions; however, the algorithm may fail to obtain a solution if δ is too close to 0. The optimization problem (6) can also be solved by applying numerical techniques. Here we consider a spectral projected gradient method implemented in R package **BB** (Varadhan and Gilbert 2009). This package can solve large scale optimization with simple constraints. It takes a nonlinear objective function as an argument as well as basic constraints. In particular, the package can find multiple roots if available, with user specified multiple starting values. To this end, starting values for n_{01} are randomly generated from 1 to $nctr - 1$. Similarly, starting values for n_{11} are randomly generated from 1 to $ntrt - 1$. We then form $\min(nctr - 1, ntrt - 1)$ pairs of random numbers and select 10% as the starting values to find multiple roots. Once the solutions (n_{01}, n_{11}) are determined, the odds ratio and the relative risk can be computed, and the results are arranged in the order of the magnitude of SS.

3. Implementation

The proposed methods in Section 2 have been implemented in R package **orsk**. The main function `orsk` returns an object of class `orsk`, for which `print` and `summary` methods are available to extract useful statistics, such as the reported odds ratio, estimated odds ratio and relative risk, with corresponding confidence intervals. Function `orsk` has an argument `type` which specifies the optimization objective function. With the default value `type = "two-sided"`, function SS (5) is minimized. Other objective functions based on Equations 1 and 3, 1 and 4, 3 and 4 have been implemented with argument `type = "lower"`, `type = "upper"` and `type = "ci-only"`, respectively. The optimization algorithm can be called with argument `method`. If `method = "grid"`, the grid search algorithm in Fortran is called. Otherwise, the constrained nonlinear optimization algorithm in R package **BB** is employed. The results from function `orsk` can be illustrated using the `summary` function in which argument `nlist` controls the maximum number of solutions displayed (the default value is 5). The source version of the **orsk** package is freely available from the Comprehensive R Archive Network at http://CRAN.R-project.org/package=orsk. The reader can install the package directly from the R prompt via

```
R> install.packages("orsk")
```

All analyses presented below are contained in a package vignette. The rendered output of the analyses is available by the R command

```
R> library("orsk")
R> vignette("orsk_demo", package = "orsk")
```

To reproduce the analyses, one can invoke the R code

```
R> edit(vignette("orsk_demo", package = "orsk"))
```

4. Example

The data in Tables 2–3 are used to illustrate the capabilities of package **orsk**. These analyses were conducted using R version 3.0.0 (2013-04-03). We applied both grid search and optimization algorithms for minimizing objective function (5) and the solutions are similar for other objective functions discussed in Section 2. Table 2 was first evaluated with the orsk function. As seen below, the output includes two parts: the configurations of the optimization problem and the estimated results. The results include the solutions n_{01} and n_{11}, named ctr_yes and trt_yes, respectively. The risk rates in the control and treatment groups are labeled ctr_risk and trt_risk, respectively. In the ascending order of SS, the output also includes the estimated odds ratio with confidence interval derived from the estimate (n_{01}, n_{11}). The estimated odds ratios and confidence intervals in the output are very close to the reported values in Table 2. However, the derived relative risks and confidence intervals are quite different from the corresponding counterpart of the odds ratios. The results indicate that the estimated relative risks are 2.02 or 1.24 and the confidence intervals can be divided into two groups as well. These two groups correspond to different assumptions on the incidence rates:

- If 18% non-users of epidural anesthesia had the first stage of labor lasting > 12 hours (i.e., $risk_0 = 0.18$), and about 37% users had the first stage of labor lasting > 12 hours, then the relative risk is 2.02 (95% confidence interval 1.8–2.27).

- Alternatively, if the corresponding risks are increased to 68% and 85%, respectively, then the relative risk is 1.24 (95% confidence interval 1.2–1.29).

```
R> library("orsk")
R> res1 <- orsk(nctr = 1636, ntrt = 2601, a = 2.61, al = 2.25,
+    au = 3.03, method = "grid")
R> summary(res1)

        Converting odds ratio to relative risk

Call:
orsk(nctr = 1636, ntrt = 2601, a = 2.61, al = 2.25, au = 3.03,
    method = "grid")

type:  two-sided              method:  grid
threshold value:  1e-04
The odds ratio utilized: 2.61, confidence interval utilized: 2.25-3.03


The following odds ratios and relative risks are for the scenarios created
with different numbers of events in control and treatment group that lead
to comparable results for the above odds ratio and confidence interval
  ctr_yes ctr_no ctr_risk trt_yes trt_no trt_risk   OR OR_lower
1     297   1339    0.182     954   1647    0.367 2.61     2.25
2     295   1341    0.180     949   1652    0.365 2.61     2.25
3     299   1337    0.183     959   1642    0.369 2.61     2.25
4     298   1338    0.182     956   1645    0.368 2.61     2.25
5    1116    520    0.682    2207    394    0.849 2.61     2.25
```
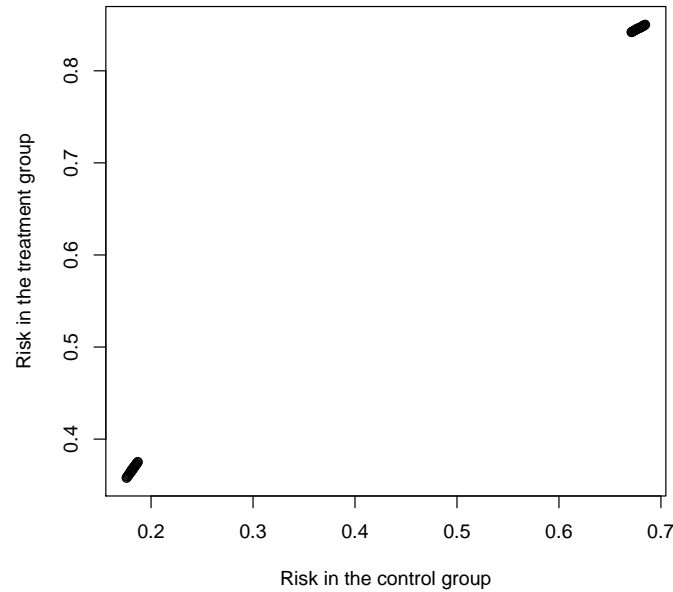
Figure 1: Distribution of risk of the outcome among scenarios for which the calculated odds ratio and confidence interval coincide with the published values.

```
  OR_upper   RR RR_lower RR_upper       SS
1     3.03 2.02      1.8     2.27 3.54e-07
2     3.03 2.02      1.8     2.27 5.79e-07
3     3.03 2.02      1.8     2.26 6.10e-07
4     3.03 2.02      1.8     2.26 9.20e-07
5     3.03 1.24      1.2     1.29 9.76e-07


R> plot(res1, type = "OR")
R> plot(res1, type = "RR")
```

In either case, the odds ratio in Table 2 overestimates the relative risk, and the displayed incidence rates are high ($> 18\%$). Since only the five best solutions are shown, one important question remains: are there any less accurate but still acceptable solutions? To answer this question, we obtain the rounded odds ratio and confidence interval if they coincide with the published values, then plot the corresponding risk of the study outcome in the control and treatment groups, respectively. Figure 1 suggests that, although the incidence rate is unknown from the published data, there is a clear evidence that the incidence is high ($>18\%$) in both the control and treatment groups. Consequently, the reported odds ratio in Table 2 can potentially overestimate the true risk ratio. Figure 2 displays the distribution of relative risk among scenarios for which the calculated odds ratio and confidence interval coincide with the published values. Clearly, the relative risk can be quite different for the same published odds ratio.

Next, utilizing the estimation of risk_0 and Equation 2, we approximate the risk ratio based on the adjusted odds ratio in Table 3. The results can be summarized briefly. Among non-users of epidural anesthesia, if 18% women had the first stage of labor lasting > 12 hours, then
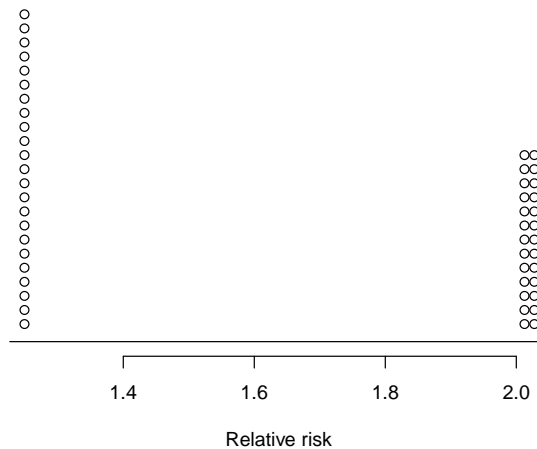
Figure 2: Distribution of relative risk among scenarios for which the calculated odds ratio and confidence interval coincide with the published values.

the approximated risk ratio is 1.84 (95% confidence interval 1.65–2.03). If $risk_0$ was increased to 68% instead, then the approximated risk ratio is 1.22 (95% confidence interval 1.18–1.25). Taking into account the incidence rate, we obtained quite different risk ratios compared with Table 3.

In the situations under consideration it can be expected that there is often no unique solution. As such, the user should carefully review the results. It may be unclear which of the computational results can be taken for further analysis, but this is not unusual for an exploratory study. Alternatively, one may reasonably hope that a subject matter expert can provide valuable insights to the situation and may help make a decision.

When applying the numerical optimization algorithm, the estimated results typically have larger *SS* than the grid search algorithm. Note the solutions may not be replicated if the starting values are generated from different random numbers. It was found that the estimated relative risks range from 1.40 to 2.19, which does not contain value 1.24 as in the grid search algorithm. Additionally, the displayed *SS* values in the **BB** algorithm are larger than those in the grid search algorithm. This example suggests that the grid search algorithm outperforms the numerical optimization algorithm as one might expect. (The result is made exactly reproducible by setting the random seed via **setRNG**, Gilbert 2012.)

```
R> library("setRNG")
R> old.seed <- setRNG(list(kind = "Mersenne-Twister",
+    normal.kind = "Inversion", seed = 579))

R> res2 <- orsk(nctr = 1636, ntrt = 2601, a = 2.61, al = 2.25, au = 3.03,
+    method = "optim")

R> summary(res2)

        Converting odds ratio to relative risk
```

```
Call:
orsk(nctr = 1636, ntrt = 2601, a = 2.61, al = 2.25, au = 3.03,
    method = "optim")

type:  two-sided              method:  optim
threshold value:  NA
The odds ratio utilized: 2.61, confidence interval utilized: 2.25-3.03

The following odds ratios and relative risks are for the scenarios created
with different numbers of events in control and treatment group that lead
to comparable results for the above odds ratio and confidence interval
  ctr_yes ctr_no ctr_risk trt_yes trt_no trt_risk   OR OR_lower
1     301   1335    0.184     963   1638    0.370 2.61     2.25
2     313   1323    0.191     993   1608    0.382 2.61     2.25
3     311   1325    0.190     989   1612    0.380 2.61     2.26
4     312   1324    0.191     990   1611    0.381 2.61     2.25
5     313   1323    0.191     994   1607    0.382 2.61     2.26
  OR_upper   RR RR_lower RR_upper       SS
1     3.02 2.01     1.80     2.25 5.21e-06
2     3.02 2.00     1.79     2.23 1.18e-05
3     3.03 2.00     1.79     2.24 1.36e-05
4     3.02 2.00     1.79     2.23 1.40e-05
5     3.03 2.00     1.79     2.23 1.82e-05

R> summary(res2$res$RR)

  Min. 1st Qu.  Median   Mean 3rd Qu.    Max.
 1.401   1.660   1.848  1.818   1.946   2.187
```

We now compare the computing speed between the two methods of estimation. With the grid search and optimization algorithm in the above example, it took 2.2 and 1.6 seconds, respectively, on an ordinary desktop PC (Intel Core 2 CPU, 1.86 GHz). Although the optimization method has some computational advantage, the grid search method can generate more accurate results with smaller *SS* and can detect multiple (local) minima. In the light of the computing time difference, there is no real benefit of using the optimization based method.

5. Conclusion

In this article we outlined the methods and algorithms for converting the odds ratio to the relative risk when only partial data information is available. As an exploratory tool, R package **orsk** can be utilized for this purpose. In addition, the methods may be used in the formula in Zhang and Yu (1998) to approximate the risk ratio obtained from logistic regression or other multiple regression models, when the risk of having a positive outcome in the control or unexposed group is not directly available. Specifically, once the cells in Table 1 are reconstructed with the aid of the `orsk` function, $risk_0$ can then be estimated. The validity of results

depends on whether the published confidence intervals have or have not been calculated with Equations 3 and 4. One restriction is that the Zhang and Yu method can only be supported in case unadjusted estimates have been published in parallel to logistic regression estimates.

Acknowledgments

References

Gilbert P (2012). *setRNG: Set (Normal) Random Number Generator and Seed.* R package version 2011.11-2, URL http://CRAN.R-project.org/package=setRNG.

Holcomb WL, Chaiworapongsa T, Luke DA, Burgdorf KD (2001). "An Odd Measure of Risk: Use and Misuse of the Odds Ratio." *Obstetrics & Gynecology*, **98**(4), 685–688.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Robbins AS, Chao SY, Fonseca VP (2002). "What's the Relative Risk? A Method to Directly Estimate Risk Ratios in Cohort Studies of Common Outcomes." *The Annals of Epidemiology*, **12**(7), 452–454.

Schulman KA, Berlin JA, Harless W, Kerner JF, Sistrunk S, Gersh BJ, Dube R, Taleghani CK, Burke JE, Williams S, Eisenberg JM, Escarce JJ (1999). "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization." *New England Journal of Medicine*, **340**(8), 618–626.

Schwartz LM, Woloshin S, Welch HG (1999). "Misunderstandings about the Effects of Race and Sex on Physicians' Referrals for Cardiac Catheterization." *New England Journal of Medicine*, **341**(4), 279–283.

Szal SE, Croughan-Minihane MS, Kilpatrick SJ (1999). "Effect of Magnesium Prophylaxis and Preeclampsia on the Duration of Labor." *American Journal of Obstetrics and Gynecology*, **180**(6), 1475–1479.

Vandenbroucke JP, Elm EV, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M (2007). "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration." *PLoS Medicine*, **4**(10), 1628–1654.

Varadhan R, Gilbert P (2009). "**BB**: An R Package for Solving a Large System of Nonlinear Equations and for Optimizing a High-Dimensional Nonlinear Objective Function." *Journal of Statistical Software*, **32**(4), 1–26. URL http://www.jstatsoft.org/v32/i04/.

Wang Z (2013). *orsk: Converting Odds Ratio to Relative Risk in Cohort Studies with Partial Data Information.* R package version 1.0-0, URL http://CRAN.R-project.org/package=orsk.

Zhang J, Yu KF (1998). "What's the Relative Risk? A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes." *Journal of the American Medical Association*, **280**(19), 1690–1691.

Affiliation:

Zhu Wang
Department of Research
Connecticut Children's Medical Center
and
Department of Pediatrics
University of Connecticut School of Medicine
Connecticut 06106, United States of America
E-mail: zwang@ccmckids.org