Converting the Odds Ratio to the Relative Risk with Partial Data Information

Zhu Wang
Connecticut Children's Medical Center
University of Connecticut School of Medicine

Abstract

In medical and epidemiological studies, the odds ratio is a commonly applied measure to approximate the relative risk or risk ratio. It is well known 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 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 a dichotomized length of labor duration. In this context, the study results may be summarized in Table 1 and the odds ratio and the relative risk are the important measures. The odds of outcome in the treatment group is $\frac{n_{11}}{n_{10}}$

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

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

and the odds of outcome in the control group is $\frac{n_{01}}{n_{00}}$. The odds ratio is

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

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 is to approximate the relative risk or risk ratio, which is a more direct measure than the odds ratio. 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 *et al.* 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 led to incorrect conclusions (Schulman *et al.* 1999; Schwartz *et al.* 1999; Holcomb *et al.* 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 $\frac{\text{odds ratio}}{1-\text{risk}_0+\text{risk}_0\times\text{odds ratio}}$, where risk_0 is the risk of having a positive outcome in the control or unexposed group (i.e., $\text{risk}_0 = \frac{n_{01}}{nctr}$). Apparently, the formula can convert the unadjusted odds ratio in the layout of Table 1. The formula can also be employed to approximate the lower and upper limits of the confidence interval. However, the above formula becomes unusable if risk_0 can't be estimated from the data.

This paper extends the work in Zhang and Yu (1998) to the cases when risk_0 can't be estimated trivially, although the proposed methods can be applied to the unadjusted odds ratio as well. The problem under investigation can be described using a concrete example. In a retrospective cohort study, Szal *et al.* (1999) collected data on 4237 women who had nulliparous, term vaginal deliveries. Here we focus on the association between the use of epidural anesthesia and prolonged first stage of labor ($>$ 12 hours), and Table 2 and Table 3 were compiled from the study. It seems to suggest that the 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 didn't use epidural anesthesia. However, Szal *et al.* (1999) didn't describe how many epidural anesthesia users and non-users had the first stage of labor lasting $>$ 12 hours. Thus risk_0 is not available in order to approximate the relative risk. If we can reconstruct Table 1 based on Table 2, then it is simple to estimate the risk of the study outcome in the control and treatment groups. Completely or at least partially reconstructing Table 1 is practically important not only in Table 2 and Table 3. 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 skip five articles due to lack of information on risk of study outcome in the control group, using the formula in Zhang and Yu (1998). More importantly, it remains unclear whether the odds ratio exaggerates a risk association or a treatment effect in the skipped studies. Methodologies have

not been proposed to estimate risk_0 when not all data information is directly available, based on the author's best knowledge. The methods proposed here can reconstruct Table 1, and can not only convert the odds ratio to the relative risk, but can also estimate risk_0. In this sense, we extend the work in Zhang and Yu (1998) to the case where risk_0 is not directly available. Table 2 will be utilized in this paper to demonstrate how to approximate the risk ratio based on partial data information. Furthermore, with the estimated risk_0 and the multiple logistic regression results in Table 3, we will approximate the risk ratio based on the adjusted odds ratio.

Table 2: Unadjusted odds ratio for the first stage of labor lasting > 12 hours.

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

Table 3: Adjusted odds ratio from multiple logistic regression for the first stage of labor lasting > 12 hours.

	Adjusted Odds ratio	95% confidence interval
Non-use of epidural anesthesia (n=1636)	Reference	Reference
Use of epidural anesthesia (n=2601)	2.25	1.92-2.63

In this paper, we develop methods to estimate the relative risk with partial data information and implement the methods in R (R Development Core Team 2011) package **orsk** (odds ratio to relative risk). The paper is organized as follows. Section 2 proposes a nonlinear objective function which measures the closeness between the calculated odds ratio and the reported odds ratio. We also provide two methods to optimize the nonlinear objective function. Section 3 outlines the implementations in the package **orsk**. Section 4 illustrates the capabilities of **orsk** with real data in Table 2 and Table 3. Finally, Section 5 concludes the paper.

2. Methods

We briefly review some additional results of the odds ratio, which form the basis for further proposal development. An asymptotic $(1-\alpha)$ confidence interval (CI) for the log odds ratio is $\log(\theta) \pm z_{\alpha/2} SE$, 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 $\theta_L = \exp(\log(\theta) - z_{\alpha/2} SE)$. Therefore,

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

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

$$\theta_U = \theta \exp\left[z_{\alpha/2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}}\right]. \tag{3}$$

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, 2 and 3, respectively. In addition, $nctr, ntrt$, 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 (2); (ii) Equations (1) and (3); (iii) Equations (2) and (3); (iv) Equations (1) to (3). The proposal is to choose (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 a sum of squares SS defined below.

$$\begin{aligned} SS(n_{01}, n_{11}) = & \left\{ \log \frac{n_{11}(ntrt - n_{01})}{(nctr - n_{01})n_{01}} - \log(\theta^{(0)}) \right\}^2 \\ & + \left\{ \log \frac{n_{11}(ntrt - n_{01})}{(nctr - n_{01})n_{01}} - z_{\alpha/2} \sqrt{\frac{1}{n_{11}} + \frac{1}{ntrt - n_{11}} + \frac{1}{n_{01}} + \frac{1}{nctr - n_{01}}} - \log(\theta_L^{(0)}) \right\}^2 \\ & + \left\{ \log \frac{n_{11}(ntrt - n_{01})}{(nctr - n_{01})n_{01}} + z_{\alpha/2} \sqrt{\frac{1}{n_{11}} + \frac{1}{ntrt - n_{11}} + \frac{1}{n_{01}} + \frac{1}{nctr - n_{01}}} - \log(\theta_U^{(0)}) \right\}^2 \\ & \equiv \left\{ \log(\theta) - \log(\theta^{(0)}) \right\}^2 + \left\{ \log(\theta_L) - \log(\theta_L^{(0)}) \right\}^2 + \left\{ \log(\theta_U) - \log(\theta_U^{(0)}) \right\}^2. \end{aligned} \tag{4}$$

Similar sums of squares can be considered with point estimate and lower or upper confidence interval bounds, or with confidence interval bounds only. As a reviewer pointed out, it is possible to develop a weighted sum of squares combining the different scenarios, which may help attenuate the impact of rounding errors. This paper focuses on the basic idea of sum of squares and potential improvements will be employed in the future version of the software. The goal now is to solve the following optimization problem:

$$\min_{n_{01}, n_{11}} SS(n_{01}, n_{11}) \text{ for integer } n_{01}, n_{11}, 1 \le n_{01} \le nctr - 1, 1 \le n_{11} \le ntrt - 1. \tag{5}$$

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 robust when minimizing any one of the objective functions. However, sometimes data are corrupted and questionable results may occur. Indeed, an application of different objective functions discovered a suspicious odds ratio and confidence interval in Lee *et al.* (2010), which was formally reported in Wang (2011).

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, we can evaluate all the values $SS(n_{01}, n_{11})$, for $n_{01} \in \{1, 2, ..., nctr - 1\}, n_{11} \in \{1, 2, ..., ntrt - 1\}$. This will result in a total number of $(nctr - 1)(ntrt - 1)$ of SS to be sorted from the smallest to the largest and 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}. Apparently,

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 problem can also be solved by applying numerical optimization techniques. Here we consider a spectral projected gradient method implemented in R package **BB** (Varadhan R 2009). This package can solve for 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 selected results can be arranged in the order of the magnitude of SS. It is worth emphasizing that the calculated odds ratios are for the scenarios created with different numbers of events in both treatment and control group that lead to comparable results for the reported odds ratio and confidence interval.

3. Implementation

The above methods have been implemented in R package **orsk**. To make the grid search algorithm computationally efficient, a Fortran subroutine is utilized. Several supporting R functions are available to extract or calculate useful statistics, such as the reported odds ratio, estimated odds ratio and relative risk, with confidence intervals. The function `orsk` returns an object of class `orsk`, for which `print` and `summary` method are available. A detailed description of these functions is available in the online help files. Function `orsk` has an argument `type` which specifies the optimization objective function. With the default value `type="two-sided"`, function SS (4) is minimized. Other objective functions based on Equations (1) and (2), (1) and (3), (2) and (3) 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 is called. Otherwise, the constrained nonlinear optimization algorithm is employed. The estimating results from running function `orsk` can be illustrated using the `summary` function and 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 (http://CRAN.R-project.org). 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 Table 2 and Table 3 are used to illustrate the capabilities of **orsk**. These analyses were conducted using R version 2.14.0 (2011-10-31) and the operating system `i686-pc-linux-gnu (32-bit)`. We applied both grid search and optimization algorithms for minimizing objective function (4) and the solutions are similar for other types of objective function discussed in Section 2. We first apply the `orsk` function to the data in Table 2. As seen below, the output includes two parts: setup and results. The setup describes the configurations of the optimization problem and the results include the solution n_{01} and n_{11}, named as `ctr_yes` and `trt_yes`, respectively. The risk rates in the control group and the treatment group are labeled as `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}), along with the known *nctr* and *ntrt*. 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. The results show that the estimated relative risks are 2.02 or 1.24. The confidence intervals can be divided into two groups as well. These two groups correspond to different assumptions on the incidence rates:

- Among those who didn't use epidural anesthesia, if about 18% women had the first stage of labor lasting > 12 hours (i.e., risk_0=0.18), and among those who used epidural anesthesia, about 37% women had the first stage of labor lasting > 12 hours, then the relative risk is 2.02 (95% confidence interval 1.8-2.3).
- On the other hand, if the corresponding risks are increased to 68% and 85%, respectively, then the relative risk is 1.24 (95% confidence interval 1.2-1.3).

```
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

 estimated results. The calculated odds ratios and relative risks are for
 the scenarios created with different numbers of events in both control and
 treatment group that lead to comparable results for the reported 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
  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
```

In either cases, the odds ratio in Table 2 overestimates the relative risk. When the incidence rate is high, the odds ratio poorly approximates the relative risk. The higher the risk, the worse the approximation. One may ask if there are low incidence rates so that the reported odds ratio does approximate the relative risk well. To address this issue, consider those cases in which the incidence of outcome in both the epidural anesthesia users and non-users are within 10%, leading to reasonable good approximations (Cummings 2009). However, there is no solution (n_{01}, n_{11}) satisfying the constraint $\frac{n_{01}}{nctr} < 0.1, \frac{n_{11}}{ntrt} < 0.1, |\theta - 2.61| < 0.3, |\theta_L - 2.25| < 0.3$, and $|\theta_U - 3.03| < 0.3$. To prove this assertion, first, it is simple algebra to show that the constraint implies $|\log(\theta) - \log(2.61)| < -\log(1 - 0.3/2.61), |\log(\theta_L) - \log(2.25)| < -\log(1 - 0.3/2.25), |\log(\theta_U) - \log(3.03)| < -\log(1 - 0.3/3.03)$. Thus, the choice of $\delta = (\log(1-0.3/2.61))^2+(\log(1-0.3/2.25))^2+(\log(1-0.3/3.03))^2$ can generate results which contain all solutions satisfying the above constraint on the point estimate and its confidence interval. Furthermore, the essentially empty output from the last line in the following code completes the proof of the assertion. Note that this particular δ is one of the numbers such that the generated results contain all solutions under the constraint. With larger δ, many more solutions are expected and the `orsk` function may fail on some systems due to the computing demand. In conclusion, although the incidence rate can't be determined from the published data, there is a clear evidence that the incidence of outcome is above 10%. Otherwise, the derived odds ratio (and its confidence interval) is different by 0.3 to the counterpart in Table 2. Because of the high incidence rate, therefore, a correction may be desirable in order to appropriately interpret the magnitude of the reported association (Zhang and Yu 1998).

```
R> anes <- orsk(nctr = 1636, ntrt = 2601, a = 2.61, al = 2.25,
+      au = 3.03, method = "grid", d = (log(1 - 0.3/2.61))^2 +
+          (log(1 - 0.3/2.25))^2 + (log(1 - 0.3/3.03))^2)
R> tmp <- subset(anes$res, ctr_risk < 0.1 & trt_risk < 0.1 &
+      abs(OR - 2.61) < 0.3 & abs(OR_lower - 2.25) < 0.3 &
+      abs(OR_upper - 3.03) < 0.3)
R> tmp

 [1] ctr_yes  ctr_no   ctr_risk trt_yes  trt_no   trt_risk OR
 [8] OR_lower OR_upper RR       RR_lower RR_upper SS
<0 rows> (or 0-length row.names)
```

Next, utilizing the above results on the estimation of risk_0, we approximate the risk ratio based on the adjusted odds ratio in Table 3. If 18% women had the first stage of labor lasting

$>$ 12 hours among those who didn't use epidural anesthesia, then the approximated risk ratio is 1.84 (95% confidence interval 1.65-2.03). If the risk was increased so that risk_0 is 68% instead, then the approximated risk ratio is 1.22 (95% confidence interval 1.18-1.25). Taking into account the incidence rate, we have obtained quite different risk ratios from Table 3.

In the situations under consideration it can be expected that there is often no unique solution. As such, the user should carefully investigate the results. It can be possible that it is not clear at all which of the computational results can be taken for further analysis. But this is not unusual for an exploratory study. On the other hand, 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. In contrast to the grid search algorithm, the estimated relative risks range from 1.40 to 2.19, which doesn't contain value 1.24 as in the grid search algorithm. Apparently, the SS values are larger than the grid search algorithm in the top solutions. This example suggests that the grid search algorithm outperforms the numerical optimization algorithm as one might expect.

```
R> require("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

 estimated results. The calculated odds ratios and relative risks are for
 the scenarios created with different numbers of events in both control and
 treatment group that lead to comparable results for the reported 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   1.849   1.820   1.951   2.189
```

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 2.4 and 1.4 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. From the code development perspective, the optimization based method is useful since it provides the solutions to which the grid method can be compared.

5. Conclusion

In this article we have 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. Since the `orsk` function is based on the asymptotic distribution of the confidence intervals, the results might not be valid in situations of small sample sizes or low event rates. However, one should recognize that this limitation is inherited from the original study.

References

Cummings P (2009). "The relative merits of risk ratios and odds ratios." *Arch Pediatr Adolesc Med*, **163**, 438–445.

Holcomb WL, Chaiworapongsa T, Luke DA, Burgdorf KD (2001). "An odd measure of risk: use and misuse of the odds ratio." *Obstetrics & Gynecology*, **98**, 685–688.

Lee SL, Islam S, Cassidy LD, Abdullah F, Arca MJ (2010). "Antibiotics and appendicitis in the pediatric population: An American Pediatric Surgical Association Outcomes and Clinical Trials Committee Systematic Review." *Journal of Pediatric Surgery*, **45**(11), 2181–2185.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 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**, 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**, 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**, 279–283.

Szal SE, Croughan-Minihane MS, Kilpatrick SJ (1999). "Effect of magnesium prophylaxis and preeclampsia on the duration of labor." *Am. J. Obstet. Gynecol.*, **180**, 1475–1479.

Varadhan R GP (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). URL http://www.jstatsoft.org/v32/i04/.

Wang Z (2011). "Letter to the editor. 'Antibiotics and appendicitis in the pediatric population: an American Pediatric Surgical Association Outcomes and Clinical Trials Committee Systematic Review'." *Journal of Pediatric Surgery*, **46**(4), 787–788.

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**, 1690–1691.

Affiliation:

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