

ORIGINAL RESEARCH

Relative Measures of Association for Binary Outcomes: Challenges and Recommendations for the Global Health Researcher

John A. Gallis^{*†} and Elizabeth L. Turner^{*†}

Background: Binary outcomes—which have two distinct levels (e.g., disease yes/no)—are commonly collected in global health research. The relative association of an exposure (e.g., a treatment) and such an outcome can be quantified using a ratio measure such as a risk ratio or an odds ratio. Although the odds ratio is more frequently reported than the risk ratio, many researchers, policymakers, and the general public frequently interpret it as a risk ratio. This is particularly problematic when the outcome is common because the magnitude of association is larger on the odds ratio scale than the risk ratio scale. Some recently published global health studies included misinterpretation of the odds ratio, which we hypothesize is because statistical methods for risk ratio estimation are not well known in the global health research community.

Objectives: To compare and contrast available statistical methods to estimate relative measures of association for binary outcomes and to provide recommendations regarding their use.

Methods: Logistic regression for odds ratios and four approaches for risk ratios: two direct regression approaches (modified log-Poisson and log-binomial) and two indirect methods (standardization and substitution) based on logistic regression.

Findings: Illustrative examples demonstrate that misinterpretation of the odds ratio remains a common issue in global health research. Among the four methods presented for estimation of risk ratios, the modified log-Poisson approach is generally preferred because it has the best numerical performance and it is as easy to implement as is logistic regression for odds ratio estimation.

Conclusions: We conclude that, when study design allows, studies with binary outcomes should preferably report risk ratios to measure relative association.

Introduction

Binary outcomes—which have two distinct levels (e.g., disease yes/no)—are commonly measured in global health research. Examples include depression status [1], disease status [2], and mortality [3], among others. These binary outcomes may either be true “yes or no” variables (e.g., mortality) or be created from an underlying continuous variable (e.g., when depression status is determined by dichotomizing a psychological scale). For example, one recent article created a binary HIV-related knowledge variable by dichotomizing a total HIV-related knowledge score at the median [4].

The relative association of an exposure and binary outcome can be quantified through the use of a ratio measure such as a risk ratio or odds ratio. The risk ratio is defined as the risk of the outcome in the exposed group over the risk

of outcome in the unexposed group, where an exposure could be a treatment (intervention) assignment or some other binary predictor (e.g., obesity yes/no). For example, using data from a randomized controlled trial (RCT) of an intervention to increase the proportion of febrile individuals testing for malaria [5], the estimated “risk” of testing for malaria is higher in treatment than control (**Table 1**; risk ratio = 1.45). The odds ratio is defined as the odds of the outcome in the exposed group over the odds of the outcome in the unexposed group, where the odds of the outcome in a group is the proportion with the outcome over the proportion without the outcome. The odds ratio for this example is 2.7, which is larger than the risk ratio.

The odds ratio (OR) is the only valid measure of relative association in traditional case-control studies, namely cumulative case-control studies, because the sampling of controls (e.g., survivor sampling) does not provide a valid estimate of the risk of exposure in the source population [6]. But for studies that use sampling that is dependent on the exposure of interest—including cohort and cross-sectional studies, and randomized controlled trials—the risk ratio (RR) is a valid alternative measure of relative association. Yet, in many of these studies, the OR is the

* Duke Global Health Institute, Duke University, Durham, North Carolina, US

† Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, US

Corresponding author: Elizabeth L. Turner, PhD (liz.turner@duke.edu)

only relative measure of associated reported [7]. This popularity is likely because it is straightforward to implement the logistic regression approach that is typically used to estimate ORs.

Despite its widespread use, the OR is frequently misinterpreted as an RR by researchers, journalists, policymakers, and the general public [8]. As shown by the example above, interpreting the OR of 2.7 as an RR would considerably overstate the impact of the intervention evaluated in this RCT. Such a large difference in magnitude between the two relative measures of associations arises here because the reference (control) arm risk (51%) shows that malaria testing is a common outcome in the study setting. In contrast, in situations where the reference risk is not large because the outcome is not common (e.g., <10%), the odds ratio would approximate the risk ratio and therefore the potential for misinterpretation is greatly reduced.

The purpose of the current paper is to provide researchers with the tools to be able to obtain appropriate and interpretable measures of relative association in studies with binary outcomes. To do so, we compare and contrast the risk ratio and odds ratio, provide examples of the misinterpretation of odds ratios from the recent global health literature, describe methods for obtaining risk ratios in analyses of binary outcomes, and make recommendations for selecting the most appropriate analysis. In addition, we briefly discuss the merits of including an absolute measure of association along with a relative measure. Our goal is to assist global health researchers in making informed

decisions about when to report the odds ratio or the risk ratio to measure relative association.

Relative Measures of Association

Motivating Example

In the introduction, we presented an example from O’Meara et al. who reported the results of a 2 × 2 factorial RCT, examining the independent and combined effects of two different subsidy interventions (subsidies for rapid diagnostic tests and subsidies for malaria treatment) on the proportion of febrile individuals testing for malaria, a binary outcome [5]. For simplicity, we considered only one of the interventions, namely subsidies for rapid diagnostic tests (RDTs), ignoring the fact that a second intervention was evaluated in the study, and reproduced the reported outcome data from Table 2 of O’Meara et al. [5] (Table 1). The probability (“risk”) of testing for malaria in the RDT subsidy arm is 73.8%, whereas this probability is 51% in the no subsidy (control) arm. Thus, as noted above, the estimated RR is 1.45, and the estimated OR is 2.7. That is, the RDT subsidy is associated with 1.45 times the “risk”, or 2.7 times the odds of malaria testing when compared with no subsidy.

At this stage, it is valuable to make a note on the terminology of “risk” and “risk ratio”. Although in the strictest sense, a risk is defined in epidemiology as the “probability of an event during a specified period of time [9]”, in common usage, “risk” refers more generally to a probability, and the term risk ratio or relative risk is commonly used

Table 1: Example of two relative measures of association, adapted from results of a randomized controlled trial in febrile individuals published in *BMJ Global Health*.^a

Exposure group	Outcome ^b		Relative Measure of Association ^c		
	Tested for malaria	Did not test for malaria	“Risk” of malaria testing	Risk Ratio	Odds Ratio
Intervention	76	27	76/103 = 73.8%	73.8/51.0 = 1.45	$\frac{76}{27} / \frac{51}{49} = 2.7$
Control	51	49	51/100 = 51.0%		

^aRCT by O’Meara et al. (2016) [5] was a 2 × 2 factorial design of two interventions for febrile individuals. Here we have adapted the example to focus on one of those interventions, namely a subsidy for a rapid diagnostic test, where “intervention” denotes the group that received the subsidy and “control” denotes the group that did not receive the subsidy. Specifically, we have extracted outcome data from Table 2 of O’Meara et al. (2016) [5] for the two groups which did not receive the second intervention.

^bRow counts correspond to the number of participants with each level of the outcome within each exposure group.

^cFor intervention group vs. control group, where we note that O’Meara et al. (2016) [5] reported neither of these results in their Table 2 because they instead reported absolute measures of effect.

Table 2: Unadjusted measures of relative association from three articles in the global health literature.

Exposure	Outcome	Unexposed group outcome proportion	Risk Ratio ^a	Odds Ratio ^b	Magnitude of odds ratio relative to risk ratio ^c
1 Surviving Ebola virus [32]	Safe sexual behavior	14%	2.71	3.67	35%
2 Point-of-care testing [33]	Antibiotic use	78%	0.82	0.50	178%
3 Drinking [34]	Feelings of aggression	20%	3.1	6.7	116%

^aRisk ratio (for “exposed” vs. “unexposed”) computed directly from outcome proportions reported in the article as none of the three articles used the risk ratio as a measure of relative association.

^bOdds ratio is obtained from unadjusted logistic regression [32] or directly from outcome proportions reported [33, 34].

^cIn these examples where the outcome is relatively common (i.e., >10%), if the odds ratio were to be incorrectly interpreted as a risk ratio, this is the magnitude of overstatement of relative association.

in research to describe a relative association, even when the probability does not involve an element of time (e.g., cross-sectional prevalence of an outcome). Hence, we use the term “risk ratio” (RR) throughout to refer to a ratio of probabilities or prevalences.

Uses and (Mis)interpretations

Although the authors did not misinterpret the degree of association of the subsidy intervention and the proportion who tested for malaria, the O’Meara example can be used to demonstrate the potential for misinterpretation of the OR. We have stated that the RR of 1.45 can be interpreted as “receiving RDT subsidy is associated with being 1.45 times more likely to test for malaria compared to those not receiving a subsidy”. But, as Schwartz et al. [8] point out in critiquing a prominent study and the media’s interpretation of the results, this is *also* how the OR is commonly interpreted. It is natural to want to interpret the OR of 2.7 as “2.7 times more likely to test for malaria if receiving RDT subsidy.” However, this is *not* the correct interpretation. In the O’Meara et al. example, if the OR is incorrectly interpreted as an RR, it overstates the intervention effect by almost double ($1.45 \times 2 = 2.9$).

As pointed out by many authors, the interpretation of an OR is not intuitive, and ORs are easily misinterpreted [7, 10, 11]. And, importantly, even when authors are careful in their interpretation, for instance by using language

such as “receiving treatment is associated with 2.7 times the odds of outcome compared with control”, it is still natural for the news media and other readers of the research to interpret it as an RR—that is, a ratio of probabilities rather than a ratio of odds [8–12].

In the example above, the OR and RR were so different because the proportion of the study sample testing for malaria (the outcome) was so high (51.0% in control). In the case of a high outcome proportion, the OR is pulled away from the null value (i.e. an OR and an RR of 1.0) more than the RR. In **Figure 1**, we display this relationship by graphing the OR and RR for various levels of the reference probability (e.g., probability of testing for malaria in the no subsidy arm), including for settings where the intervention is associated with a reduction in the probability of the outcome (i.e. with the RR and OR both <1). As can be seen, the higher the reference probability of the event (e.g., tested for malaria) the more the OR overstates the RR (i.e., $OR < RR$ if both are <1 and $OR > RR$ if both are >1), if the OR is incorrectly interpreted as an RR. Similarly, the figure shows that when the outcome is not common (e.g., <10%), the OR closely approximates the RR, and therefore, in such cases the OR may be interpreted as an RR.

Why does this matter?

Although there is some literature on methods to compute RRs for binary outcomes, most of it has appeared in

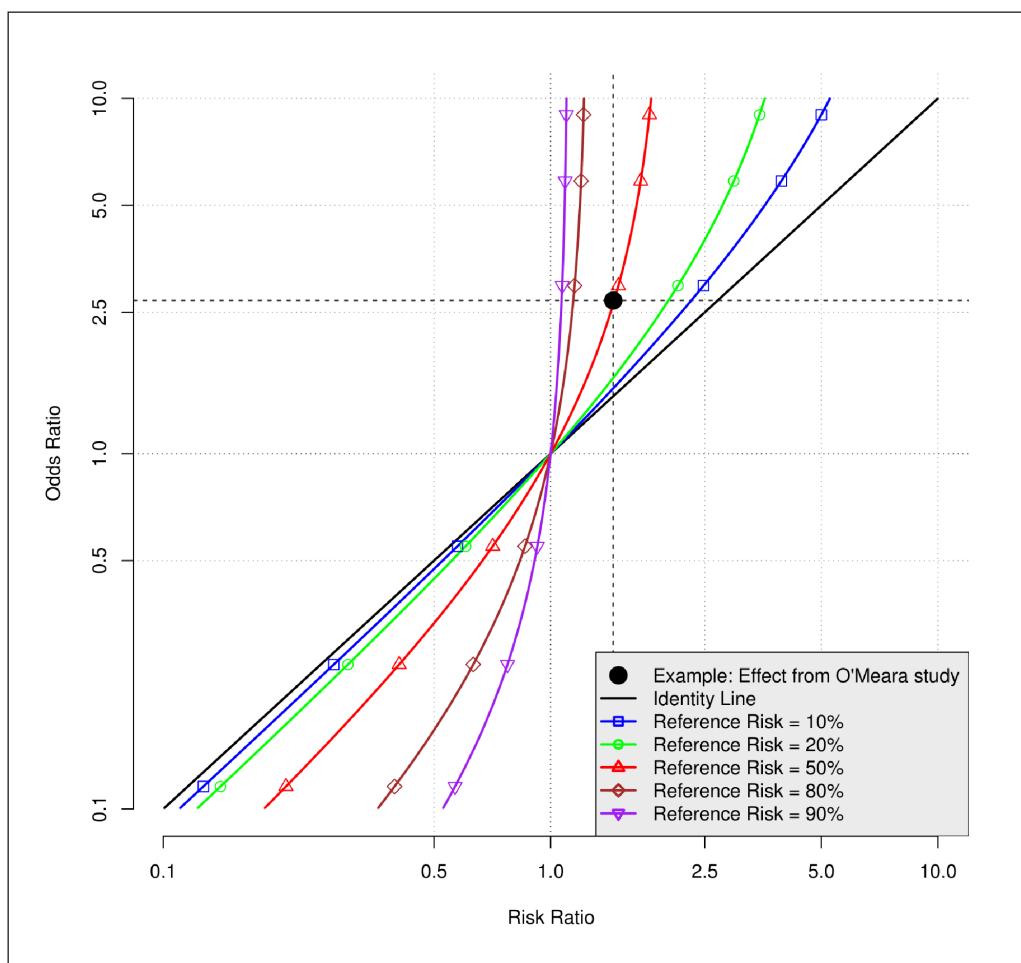


Figure 1: Relationship between the odds ratio and risk ratio at various levels of the reference risk.

epidemiological [13–19], medical [11, 20–27], or statistics journals [28–31]. This may partly explain why ORs are still commonly used and misinterpreted across a wide array of papers and throughout the media. The following three examples from the global health literature, and which are summarized in **Table 2**, present further evidence to the global health research community about why this issue is important.

In the first example, it was reported that Ebola virus disease survivors “were more than five times as likely to engage in safe sexual behavior compared with the comparison group”, based on an adjusted OR of 5.59, obtained from a logistic regression model with adjustment for relevant covariates selected by the authors [32]. Given that the corresponding unadjusted OR is 35% larger than the unadjusted RR (3.67 vs. 2.71), the adjusted RR is expected to be approximately 4.14 (vs. 5.59), which would indicate that survivors were “more than four times as likely to engage in safe sexual behavior”. This example shows that even in studies with a relatively low reference probability of the outcome of interest, interpreting the OR as an RR would lead to an overstatement of the association.

In a second example from the global health literature, an adjusted OR of 0.49 comparing antibiotic use in the intervention (64%) and control (78%) groups was estimated in a RCT of a point-of-care testing intervention to reduce antibiotic use [33]. The unadjusted OR of 0.50 corresponds to an unadjusted RR of 0.82. If this OR were interpreted as an RR, the association of the intervention and outcome would be overstated almost three-fold (i.e., 50% vs. 18% reduction). This article can be used to further emphasize why the distinction between the two relative measures of association is important. Suppose that the intervention is associated with an increase in adverse events. Reporting the OR instead of the RR may make it more challenging to balance the pros of the intervention in terms of reduction in antibiotic overuse vs. the cons due to adverse events. The more high-impact the study, the more likely the conclusions will affect policy and, hence, will affect people’s lives.

The third example from the global health literature is an example of misinterpretation of the OR in the news media. The article reports on the results of a study examining the emotions associated with alcohol consumption [34], in which an adjusted OR of 6.41 comparing heavy drinkers with light drinkers for the outcome of “feelings of aggression” was interpreted as heavy drinkers were “just over six times more likely to report feelings of aggression” than light drinkers [34]. In reporting on the study the news article picked up on this interpretation and stated that “those who showed signs of alcohol dependence were six times more likely to say they felt aggression while drinking [12]”. However, using numbers provided in Table 2 of the paper we find an unadjusted OR for this association of 6.7 while the unadjusted RR is 3.1. Thus, the OR overstates the RR more than twofold and, instead, a more appropriate interpretation would be that “heavy drinkers were just over three times more likely to report feelings of aggression than light drinkers”.

Methods of Obtaining Risk Ratios For Binary Outcomes

Given the difficulty in interpreting the odds ratio, several methods of obtaining risk ratios have been proposed in the literature and have been implemented in a range of studies. For simplicity and brevity, here we describe some advantages and disadvantages of four commonly used methods [29]. Summaries and examples of use of these four methods in the global health literature are given in **Table 3**. Two of the methods are regression-based approaches that directly estimate the RR (log-binomial [35] and modified log-Poisson [16]). In this case, the estimated coefficient of the association between exposure and outcome, when exponentiated, is directly interpreted as a risk ratio. The other two methods indirectly obtain the RR (substitution [25] and standardization [18]) by estimating an OR from a logistic regression model, then computing the RR as a function of the OR through some form of transformation. Although commonly used in many settings, these methods are only necessary when adjusting for covariates. This is because when there is no need to adjust for covariates, simple formulas can be used instead of modeling. Nevertheless, in many cases, researchers will use model-based methods even when not adjusting for covariates as they are similarly valid in unadjusted models and are straightforward to implement.

The first direct estimation method, the log-binomial approach, is a generalized linear model like logistic regression which also uses a binomial outcome distribution but uses a log link rather than a logit link function [35]. While this is an attractive option because it is simple to implement in standard statistical software, the model may fail to converge to a solution, especially when the outcome is common [18, 41].

The second direct estimation method is the modified log-Poisson model [16]. In this approach, a Poisson model with log-link is fitted to the binary data, which is “modified” by using robust standard errors to obtain valid statistical inference. This approach is simple to implement in many statistical software packages, and generally does not suffer from the same convergence issues as the log-binomial.

While the log-binomial and modified log-Poisson regression approaches are appealing, both may estimate individual-level outcome probabilities and/or the upper bound of their 95% confidence intervals above one for binary outcome data. If the intention is inference about associations, this is generally not a major issue. However, if the goal is estimation of individual-level risk, then these two methods will sometimes be inappropriate and estimate individual-level risk above one, especially when the outcome is very common and the variance of adjustment variables is high [11, 42].

Two logistic-regression based approaches that indirectly obtain the RR through transformation are substitution and standardization. Zhang and Yu [25] proposed a substitution method, in which a simple formula—which includes the odds ratio and the prevalence of the outcome in the unexposed group—is used to convert the OR (and its 95%

Table 3: Brief summary of four methods of obtaining risk ratios for binary outcomes.

Name of method	Type of method	Back-ground literature	Some advantages	Some disadvantages	Example of use in the global health literature	Exposure	Binary Outcome
Log-binomial	Direct	Wacholder (1986) [35]	Easy to implement.	May not converge; may estimate individual-level probabilities (and/or the upper bound of their 95% confidence intervals) above 1.	Gibson et al. (2017) [37]	Mobile phone based intervention to improve immunization rates, in a cluster-randomized trial	Full immunization by 12 months of age.
Modified log-Poisson	Direct	Zou (2004) [16]	Easy to implement; almost always converges.	May estimate individual-level probabilities (and/or the upper bound of their 95% confidence intervals) above 1.	Chan et al. (2017) [38]	AIDS-related stigma	Probable depression (PHQ-9 score ≥ 10 or recent suicidal thoughts).
Substitution	Indirect	Zhang and Yu (1998) [25]	Easy to implement. Uses output from logistic regression.	Generally produces biased estimates and 95% confidence intervals are expected to be too narrow, on average [18].	Agweyu et al. (2018) [39]	Various demographics and health-related exposures	Mortality.
Marginal or Conditional Standardization	Indirect	Localio et al. (2007) [18]	Uses output from logistic regression.	May be more difficult to implement and interpret than other methods, especially in certain software packages.	Weobong et al. (2017) [40]	Psychological intervention for depression, in a randomized trial	Remission from depression as measured by the PHQ-9.

Abbreviation: PHQ-9 – Patient Health Questionnaire 9-item [36], a screening tool for depression.

confidence interval [CI]) obtained using a standard logistic regression model to an RR (with 95% CI). Although this method is often cited and used in practice, simulations suggest that 95% CIs obtained using this method suffer from poor coverage, such that they are too narrow and type I error is inflated (too high) [18]. Additionally, several authors have pointed out that such simple substitution methods produce biased RRs [13, 14, 18, 20]. Thus, this method is not recommended and there is currently no simple formula with desirable statistical properties to convert an OR to an RR.

The second indirect logistic regression-based approach is the standardization method proposed by Localio et al. [18] and described in further detail in Muller & MacLehose [43]. This method fits a logistic regression model and uses the estimated regression coefficients to obtain an estimated RR by using marginal standardization whereby the proportions with the outcome in the exposed and unexposed groups are estimated and, from these, the corresponding RRs are estimated. We obtain these proportions as the estimated probability of the outcome within each unexposed and exposed group at specified values of the other covariates in the model (e.g., at the mean of continuous variables). Under certain assumptions, this marginalized effect at each level of the exposure is the prevalence of the outcome we would have observed had everyone been assigned to that level of the exposure and to the specified values of the other covariates in the model. That

is, although termed a “marginalized effect” when adjusted for covariates in this way, it is also conditional on the level of those covariates [43].

Cummings compares these methods with others and finds that, except for the often biased substitution method, they generally produce similar estimates of the RR [29]. Therefore, on balance, we find that the most easily implementable approach with the fewest drawbacks is the modified log-Poisson approach. Given that it is as easy to implement as logistic regression in all major software, there should be no barriers to global health researchers estimating and reporting the RR as a measure of relative association of an exposure and binary outcome, when study design allows. In **Table 4**, we provide code for fitting both the log-binomial and modified log-Poisson models in four commonly used statistical software (R, SAS, Stata, and SPSS), as well as code to implement the marginal standardization approach in both R and Stata. For the marginal standardization approach, SAS and SPSS code are not provided as to our knowledge it is not easily implemented in these two programs.

Additionally, cluster randomized trials (CRTs) are common in global health research. In such cases, statistical models must take into account the fact that the data collected on participants within the same cluster are likely to be correlated. When the outcome is binary, the generalized estimating equations (GEE) approach is an appealing method to analyze CRT data because of its desirable statistical

Table 4: Code to fit the log-binomial and modified log-Poisson models in four commonly used statistical software packages, and to use the marginal standardization method in two of the packages.

Software Program	Data Structure	Log-binomial code ^a	Modified log-Poisson code ^b	Marginal standardization code ^c
Stata ^e	Ind	glm binaryoutcome exposure, family(binomial) link(log) eform	glm binaryoutcome exposure, family(poisson) link(log) vce(robust) eform	logit binaryoutcome i.exposure, or margins exposure, coeflegend post nlcom (RR: _b[1.exposure]/_b[0bn.expo- sure]), post
	Clust ^d	xtset cluster xtgee binaryoutcome exposure, family(binomial) link(log) corr(exchangeable) eform	xtset cluster xtgee binaryoutcome exposure, family(poisson) link(log) corr(exchangeable) eform	xtset cluster xtgee binaryoutcome i.exposure, family(binomial) link(logit) corr(exchangeable) eform margins exposure, post coeflegend nlcom (ratio1: _b[1.exposure]/_b[0bn.expo- sure]), post
SAS	Ind	proc genmod data=temp; class binaryoutcome exposure / param=ref ref=first; model binaryoutcome = exposure / dist=bin link=log; estimate 'Risk Ratio' exposure 1 /exp; run;	proc genmod data=temp; class binaryoutcome exposure participantID / param=ref ref=first; model binaryoutcome = exposure / dist=poisson link=log; repeated subject=participantID / type=Ind; estimate 'Risk Ratio' exposure 1 / exp; run;	
	Clust	proc genmod data=temp descending; class binaryoutcome exposure cluster / desc; model binaryoutcome = exposure / dist=binomial link=log; repeated subject=cluster / corr=exch; estimate 'Risk Ratio' exposure 1 -1 /exp; run;	proc genmod data=temp descending; class binaryoutcome exposure clus- ter / desc; model binaryoutcome = exposure / dist=poisson link=log; repeated subject=cluster / corr=exch; estimate 'Risk Ratio' exposure 1 -1 /exp; run;	

(Contd.)

Software Program	Data Structure	Log-binomial code ^a	Modified log-Poisson code ^b	Marginal standardization code ^c
R ⁱ	Ind	<pre>lglm <- glm(binaryoutcome~exposure, family=binomial(link="log")) exp(cbind(coef(lglm), confint(lglm)))</pre>	<pre>library(gee) pglm <- summary(gee(binaryoutcome~exposure, family=poisson(link="log")), id=participantID, corstr="independent") cbind(RiskRatio=exp(pglm\$coefficients[2]-1.96*pglm\$coefficients[8]), UCI=exp(pglm\$coefficients[2]+1.96*pglm\$coefficients[8]))</pre>	<pre>library(epitools) lglm <- glm(binaryoutcome~exposure, family=binomial) probratio(lglm, method="ML")</pre>
	Clust	<pre>library(gee) lgee<-summary(gee(binaryoutcome~exposure, family=binomial(link="log")), id=cluster, corstr="exchangeable") cbind(RiskRatio=exp(lgee\$coefficients[2]), LCI=exp(lgee\$coefficients[2]-1.96*lgee\$coefficients[8]), UCI=exp(lgee\$coefficients[2]+1.96*lgee\$coefficients[8]))</pre>	<pre>library(gee) pgee<-summary(gee(binaryoutcome~exposure, family=poisson(link="log")), id=cluster, corstr="exchangeable") cbind(RiskRatio=exp(pgee\$coefficients[2]), LCI=exp(pgee\$coefficients[2]-1.96*pgee\$coefficients[8]), UCI=exp(pgee\$coefficients[2]+1.96*pgee\$coefficients[8]))</pre>	
SPSS	Ind	<pre>genlin binaryoutcome (reference=first) by exposure (order=descending) /model exposure intercept=yes distribution=binomial link=Log /Print summary solution(exponentiated)</pre>	<pre>genlin binaryoutcome (reference=first) by exposure (order=descending) /model exposure intercept=yes distribution=poisson link=Log /repeated subject=participantID sort=yes corrtpe=independent adjustcorr=yes covb=robust /Print summary solution(exponentiated)</pre>	
	Clust	<pre>genlin binaryoutcome (reference=first) by exposure (order=descending) /model exposure intercept=yes distribution=binomial link=Log /repeated subject=cluster sort=yes corrtpe=exchangeable adjustcorr=yes covb=robust /Print summary solution(exponentiated)</pre>	<pre>genlin binaryoutcome (reference=first) by exposure (order=descending) /model exposure intercept=yes distribution=poisson link=Log /repeated subject=cluster sort=yes corrtpe=exchangeable adjustcorr=yes covb=robust /Print summary solution(exponentiated)</pre>	

Abbreviations: Ind = Independent (i.e., non-clustered); Clust = Clustered.

Variables: binaryoutcome = the binary outcome; exposure = exposure (e.g., treatment group indicator), assumed to be categorical; participantID = participant identifier; cluster = cluster identifier.

(Contd.)

- ^a The log-binomial code for direct estimation of the risk ratio in the clustered setting is only shown in the generalized estimating equations (GEE) framework. A generalized linear mixed model (GLMM) could also be used.
- ^b For the log-Poisson approach, a robust standard error is needed to account for misspecification of the outcome distribution (i.e., Poisson instead of binomial); GEE is the natural approach to obtain this robust standard error, in both the non-clustered and clustered setting.
- ^c To our knowledge, the marginal standardization method is not as straightforward to implement in SAS or SPSS, so no code is provided. In addition, we are unaware of an easy-to-implement function in R to perform marginal standardization in a clustered setting.
- ^d In the context of GEE to analyze clustered outcome data, we have used an exchangeable working correlation matrix as an example. It is natural to use such a working correlation matrix when the outcome data are measured at a single point in time and the clustering arises through some natural grouping of individuals (e.g., in schools or hospitals). But, if the clustering arises from longitudinal data, other working correlation structures may be preferred.
- ^e The standard errors from Stata may be slightly larger than that obtained from the other programs. This is because Stata multiplies the robust standard errors by $K/(K-1)$, where K is the number of clusters, whereas other programs do not do this.
- ^f The `cbind` R code illustrated here works only for a single binary exposure variable. It will need to be modified for more complex scenarios. Additionally, the `gee` function requires that the outcome be set up as a numeric variable, rather than a factor variable, when specifying the modified log-Poisson model.

properties (e.g., population-average interpretation; robustness to model misspecification; ability to correct for small-sample bias in the case of fewer than 40 clusters enrolled in the CRT) [44–46]. Both the log-binomial and modified log-Poisson [30] models can be easily implemented in the GEE framework, and code to do so is provided in **Table 4**. Similarly, in some software, the marginal standardization procedure can also be easily adapted to clustered binary outcome data. This code could also be used for non-randomized studies with clustering of outcomes.

Absolute Measures of Association

Just as RRs can be directly estimated as a measure of relative association, direct and indirect methods are available to estimate absolute measures of association such as risk differences. Although a discussion of the advantages and disadvantages of such methods is beyond the scope of the current article, it is important to note that absolute measures of association are able to provide important and complementary information about the public health impact of interventions. In fact, both the Consolidated Standards of Reporting Trials (CONSORT) and the Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statements recommend that all RCTs and observational studies reporting on the association of an exposure with binary outcomes provide both a relative measure and absolute measure [47, 48].

As an example of the importance of reporting both types of measures, suppose that the probability of malaria testing uptake in the RDT subsidy arm was 0.02%, while in the no subsidy arm it was 0.01%. In this case, the risk ratio is 2, whereas the risk difference is 0.01 percentage points. From a public health perspective, such a small risk difference may be considered of insufficient magnitude to justify the increased costs and potential adverse effects of providing subsidies. However, if only the relative measure were reported, those involved in scaling up the intervention may be misled to believe the intervention is more effective than it actually is. Thus, we also strongly recommend reporting both a relative and absolute measure of association when reporting on binary outcomes.

The statistical methods for obtaining an absolute measure of association (specifically, the risk difference) are straightforward to implement. For both the modified log-Poisson and log-binomial models, instead of using a log link, an identity link can be used. In this case, the regression coefficients will have a straightforward interpretation as the difference in risk between the levels of the predictor (e.g., intervention and control). Additionally, for the marginal standardization method, once we have obtained the estimated mean risks for the levels of the predictor, we simply subtract the estimated mean risks to obtain the risk difference. In sum, the same methods that can be used to obtain risk ratio can also be used to obtain risk differences with only minor modifications.

Conclusion

We have shown that many methods exist to estimate risk ratios for binary outcome data and that the global health researcher need not feel compelled to present odds

ratios for studies with sampling which depends upon the exposure, such as cohort and cross-sectional observational studies and randomized controlled trials. Overall, the modified log-Poisson regression approach to generate RRs is generally preferred to alternative approaches due to its ease of implementation and desirable statistical properties. While we have not provided an exhaustive review of methods for estimating RRs, Cummings [20] provides an excellent review of alternative methods, and also provides Stata code for implementing various approaches [29]. In addition, Muller & MacLehose discuss marginal standardization (as well as other methods), and provide Stata code for implementing marginal standardization [43]. As noted earlier, sample code for fitting the two direct regression approaches in four statistical programs, as well as performing marginal standardization in two of those programs, is provided in **Table 4**. As with fitting any model, researchers should be aware of and test assumptions underlying the model, and consider how interpretation changes when adjusting for covariates.

Given concerns with interpretation, especially since results of research are commonly used to implement and scale up interventions, we believe that estimation and reporting of odds ratios should be reserved for use mainly when performing case-control studies. In this case, the risk ratio will be directly estimated by the OR if base population sampling (i.e., a case-cohort design) is used. It should be noted, however, that for other forms of control sampling (i.e., risk-set sampling and survivor sampling), the risk ratio would not be validly estimated by the odds ratio [49].

We have provided researchers with the information needed to decide upon the most appropriate and interpretable measures of relative association to present in studies with binary outcomes, while also describing in detail the tools needed to obtain these relative measures. Our hope is that this information will assist researchers in providing the best evidence on the association between exposures and binary outcomes in observational studies, as well as on the effectiveness of interventions evaluated in RCTs in global health settings.

Acknowledgements

The authors wish to thank Joseph Egger, Kristie Kusibab, and Monica Harding of the Duke Global Health Institute Research Design and Analysis Core for their helpful comments on earlier versions of this manuscript. This paper was inspired by the authors' work on the Bachpan study (PI: Dr. Joanna Maselko) and the Innovative public-private partnership to target subsidized antimalarials study (PI: Dr. Wendy Prudhomme O'Meara). The paper received partial funding from National Institutes of Health awards R01AI110478 and K01MH104310 (ELT), and R01HD075875 (both JAG and ELT), and award OPP1171753 from the Bill and Melinda Gates Foundation (ELT). Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Bill and Melinda Gates Foundation.

Funding Information

Partial funding from National Institutes of Health awards R01AI110478 and K01MH104310 (ELT), and R01HD075875 (both JAG and ELT), and award OPP1171753 from the Bill and Melinda Gates Foundation (ELT).

Competing Interests

The authors have no competing interests to declare.

Author Contribution

All authors had access to the data and a role in writing the manuscript.

References

1. **Turner EL, Sikander S, Bangash O**, et al. The effectiveness of the peer delivered Thinking Healthy Plus (THPP+) Programme for maternal depression and child socio-emotional development in Pakistan: study protocol for a three-year cluster randomized controlled trial. *Trials*. 2016; 17(1): 442. DOI: <https://doi.org/10.1186/s13063-016-1530-y>
2. **Ishikuro M, Matsubara H, Kikuya M**, et al. Disease prevalence among nursery school children after the Great East Japan earthquake. *BMJ Global Health*. 2017; 2(2): e000127. DOI: <https://doi.org/10.1136/bmjgh-2016-000127>
3. **Snavely ME, Maze MJ, Muiruri C**, et al. Socio-cultural and health system factors associated with mortality among febrile inpatients in Tanzania: A prospective social biopsy cohort study. *BMJ Global Health*. 2018; 3(1): e000507. DOI: <https://doi.org/10.1136/bmjgh-2017-000507>
4. **Faust L, Yaya S and Ekholuenetale M**. Wealth inequality as a predictor of HIV-related knowledge in Nigeria. *BMJ Global Health*. 2017; 2(4): e000461. DOI: <https://doi.org/10.1136/bmjgh-2017-000461>
5. **O'Meara WP, Mohanan M, Laktabai J**, et al. Assessing the independent and combined effects of subsidies for antimalarials and rapid diagnostic testing on fever management decisions in the retail sector: Results from a factorial randomised trial in western Kenya. *BMJ Global Health*. 2016; 1(2): e000101. DOI: <https://doi.org/10.1136/bmjgh-2016-000101>
6. **Keogh RH and Cox DR**. Case-control studies. Cambridge, New York: Cambridge University Press; 2014.
7. **Knol MJ, Duijnhoven RG, Grobbee DE**, et al. Potential Misinterpretation of Treatment Effects Due to Use of Odds Ratios and Logistic Regression in Randomized Controlled Trials. *PLoS One*. 2011; 6(6): e21248. DOI: <https://doi.org/10.1371/journal.pone.0021248>
8. **Schwartz LM, Woloshin S and Welch HG**. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N Engl J Med*. 1999; 341(4): 279. DOI: <https://doi.org/10.1056/NEJM199907223410411>
9. **Rothman KJ, Greenland S and Lash TL**. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
10. **Persoskie A and Ferrer RA**. A Most Odd Ratio: Interpreting and Describing Odds Ratios. *Am J Prev Med*. 2017; 52(2): 224–28. DOI: <https://doi.org/10.1016/j.amepre.2016.07.030>
11. **Knol MJ, Le Cessie S, Algra A**, et al. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. *Can Med Assoc J*. 2012; 184(8): 895–99. DOI: <https://doi.org/10.1503/cmaj.101715>
12. **NHS Choices**. Different alcoholic drinks may trigger different emotions; many of them negative; 2017 [accessed May 1 2018].
13. **Greenland S**. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol*. 2004; 160(4): 301–05. DOI: <https://doi.org/10.1093/aje/kwh221>
14. **McNutt L-A, Wu C, Xue X**, et al. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol*. 2003; 157(10): 940–43. DOI: <https://doi.org/10.1093/aje/kwg074>
15. **Spiegelman D and Hertzmark E**. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol*. 2005; 162(3): 199–200. DOI: <https://doi.org/10.1093/aje/kwi188>
16. **Zou G**. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004; 159(7): 702–06. DOI: <https://doi.org/10.1093/aje/kwh090>
17. **Chu H and Cole SR**. Estimation of risk ratios in cohort studies with common outcomes: A Bayesian approach. *Epidemiology*. 2010; 21(6): 855–62. DOI: <https://doi.org/10.1097/EDE.0b013e3181f2012b>
18. **Localio AR, Margolis DJ and Berlin JA**. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *J Clin Epidemiol*. 2007; 60(9): 874–82. DOI: <https://doi.org/10.1016/j.jclinepi.2006.12.001>
19. **Sinclair JC and Bracken MB**. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol*. 1994; 47(8): 881–89. DOI: [https://doi.org/10.1016/0895-4356\(94\)90191-0](https://doi.org/10.1016/0895-4356(94)90191-0)
20. **Cummings P**. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med*. 2009; 163(5): 438–45. DOI: <https://doi.org/10.1001/archpediatrics.2009.31>
21. **Penman AD and Johnson WD**. Complementary Log–Log Regression for the Estimation of Covariate-Adjusted Prevalence Ratios in the Analysis of Data from Cross-Sectional Studies. *Biometrical Journal*. 2009; 51(3): 433–42. DOI: <https://doi.org/10.1002/bimj.200800236>
22. **Barros AJ and Hirakata VN**. Alternatives for logistic regression in cross-sectional studies: An empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol*. 2003; 3(1): 1. DOI: <https://doi.org/10.1186/1471-2288-3-21>
23. **Santos CA, Fiaccone RL, Oliveira NE**, et al. Estimating adjusted prevalence ratio in clustered cross-sectional epidemiological data. *BMC Med*

- Res Methodol.* 2008; 8(1): 80. DOI: <https://doi.org/10.1186/1471-2288-8-80>
24. **Davies HTO, Crombie IK and Tavakoli M.** When can odds ratios mislead? *BMJ.* 1998; 316(7136): 989–91. DOI: <https://doi.org/10.1136/bmj.316.7136.989>
 25. **Zhang J and Kai FY.** What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA.* 1998; 280(19): 1690–91. DOI: <https://doi.org/10.1001/jama.280.19.1690>
 26. **Holcomb WL Jr, Chaiworapongsa T, Luke DA, et al.** An odd measure of risk: Use and misuse of the odds ratio. *Obstet Gynecol.* 2001; 98(4): 685–88. DOI: <https://doi.org/10.1097/00006250-200110000-00028>
 27. **Grant RL.** Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ.* 2014; 348: f7450. DOI: <https://doi.org/10.1136/bmj.f7450>
 28. **Fang J.** Using SAS procedures FREQ, GENMOD, LOGISTIC, and PHREG to estimate adjusted relative risks—a case study. *SAS Global Forum 2011.* 4–11 April 2011; 2011. Citeseer.
 29. **Cummings P.** Methods for estimating adjusted risk ratios. *Stata Journal.* 2009; 9(2):175. DOI: <https://doi.org/10.1177/1536867X0900900201>
 30. **Zou G and Donner A.** Extension of the modified Poisson regression model to prospective studies with correlated binary data. *Stat Methods Med Res.* 2013; 22(6): 661–70. DOI: <https://doi.org/10.1177/0962280211427759>
 31. **Freedman DA.** Randomization does not justify logistic regression. *Stat Sci.* 2008; 23(2): 237–49. DOI: <https://doi.org/10.1214/08-STS262>
 32. **Kondé MK, Diop MK, Marie Yvonne C, et al.** Sex practices and awareness of Ebola virus disease among male survivors and their partners in Guinea. *BMJ Global Health.* 2017; 2(3): e000412. DOI: <https://doi.org/10.1136/bmjgh-2017-000412>
 33. **Do NT, Ta NT, Tran NT, et al.** Point-of-care C-reactive protein testing to reduce inappropriate use of antibiotics for non-severe acute respiratory infections in Vietnamese primary health care: A randomised controlled trial. *The Lancet Global Health.* 2016; 4(9): e633–e41. DOI: [https://doi.org/10.1016/S2214-109X\(16\)30142-5](https://doi.org/10.1016/S2214-109X(16)30142-5)
 34. **Ashton K, Bellis MA, Davies AR, et al.** Do emotions related to alcohol consumption differ by alcohol type? An international cross-sectional survey of emotions associated with alcohol consumption and influence on drink choice in different settings. *BMJ Open.* 2017; 7(10): e016089. DOI: <https://doi.org/10.1136/bmjopen-2017-016089>
 35. **Wacholder S.** Binomial regression in GLIM: Estimating risk ratios and risk differences. *Am J Epidemiol.* 1986; 123(1): 174–84. DOI: <https://doi.org/10.1093/oxfordjournals.aje.a114212>
 36. **Kroenke K, Spitzer R and Williams J.** The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001; 16(9): 606–13. DOI: <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
 37. **Gibson DG, Ochieng B, Kagucia EW, et al.** Mobile phone-delivered reminders and incentives to improve childhood immunisation coverage and timeliness in Kenya (M-SIMU): A cluster randomised controlled trial. *The Lancet Global Health.* 2017; 5(4): e428–e38. DOI: [https://doi.org/10.1016/S2214-109X\(17\)30072-4](https://doi.org/10.1016/S2214-109X(17)30072-4)
 38. **Chan BT, Pradeep A, Prasad L, et al.** Association between internalized stigma and depression among HIV-positive persons entering into care in Southern India. *Journal of global health.* 2017; 7(2): 020403. DOI: <https://doi.org/10.7189/jogh.07.020403>
 39. **Agweyu A, Lilford RJ and English M.** Appropriateness of clinical severity classification of new WHO childhood pneumonia guidance: a multi-hospital, retrospective, cohort study. *The Lancet Global Health.* 2018; 6(1): e74–e83. DOI: [https://doi.org/10.1016/S2214-109X\(17\)30448-5](https://doi.org/10.1016/S2214-109X(17)30448-5)
 40. **Weobong B, Weiss HA, McDaid D, et al.** Sustained effectiveness and cost-effectiveness of the Healthy Activity Programme, a brief psychological treatment for depression delivered by lay counsellors in primary care: 12-month follow-up of a randomised controlled trial. *PLoS Med.* 2017; 14(9): e1002385. DOI: <https://doi.org/10.1371/journal.pmed.1002385>
 41. **Williamson T, Eliasziw M and Fick GH.** Log-binomial models: Exploring failed convergence. *Emerging Themes in Epidemiology.* 2013; 10(1): 1–10. DOI: <https://doi.org/10.1186/1742-7622-10-14>
 42. **Yelland LN, Salter AB and Ryan P.** Relative risk estimation in randomized controlled trials: A comparison of methods for independent observations. *Int J Biostat.* 2011; 7(1): 1–31. DOI: <https://doi.org/10.2202/1557-4679.1323>
 43. **Muller CJ and MacLehose RF.** Estimating predicted probabilities from logistic regression: Different methods correspond to different target populations. *Int J Epidemiol.* 2014; 43(3): 962–70. [published Online First: 2014/03/05]. DOI: <https://doi.org/10.1093/ije/dyu029>
 44. **Liang KY and Zeger SL.** Longitudinal data analysis using generalized linear models. *Biometrika.* 1986; 73(1): 13–22. DOI: <https://doi.org/10.1093/biomet/73.1.13>
 45. **Leyrat C, Morgan KE, Leurent B, et al.** Cluster randomized trials with a small number of clusters: Which analyses should be used? *Int J Epidemiol.* 2017 [published Online First: 2017/10/13]. DOI: <https://doi.org/10.1093/ije/dyx169>
 46. **Turner EL, Prague M, Gallis JA, et al.** Review of Recent Methodological Developments in Group-Randomized Trials: Part 2 – Analysis. *Am J Public Health.* 2017; 107(7): 1078–86. DOI: <https://doi.org/10.2105/AJPH.2017.303707>
 47. **von Elm E, Altman DG, Egger M, et al.** Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ.* 2007; 335(7624):

806–8. [published Online First: 2007/10/20]. DOI: <https://doi.org/10.1136/bmj.39335.541782.AD>

48. **Schulz KF, Altman DG and Moher D.** CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; 340: c332. DOI: <https://doi.org/10.1136/bmj.c332>

[published Online First: 2010/03/25]. DOI: <https://doi.org/10.1016/j.ijsu.2010.09.006>

49. **Pearce N.** What does the odds ratio estimate in a case-control study? *Int J Epidemiol*. 1993; 22(6): 1189–92. DOI: <https://doi.org/10.1093/ije/22.6.1189>

How to cite this article: Gallis JA and Turner EL. Relative Measures of Association for Binary Outcomes: Challenges and Recommendations for the Global Health Researcher. *Annals of Global Health*. 2019; 85(1): 137, 1–12. DOI: <https://doi.org/10.5334/aogh.2581>

Published: 20 November 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Annals of Global Health* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 