

Article

Overestimation of Relative Risk and Prevalence Ratio: Misuse of Logistic Modeling

Charalambos Gnardellis ¹, Venetia Notara ², Maria Papadakaki ³, Vasilis Gialamas ⁴ and Joannes Chliaoutakis ^{3,*}

- ¹ Department of Fisheries and Aquaculture, School of Agricultural Sciences, University of Patras, 26504 Patra, Greece
- ² Department of Public and Community Health, School of public Health, University of West Attica, 11521 Athens, Greece
- ³ Department of Social Work, Faculty of Health Sciences, Hellenic Mediterranean University, 71004 Heraklion, Greece
- ⁴ Department of Early Childhood Education, School of Education National and Kapodistrian, University of Athens, 10680 Athens, Greece
- * Correspondence: jchlia@hmu.gr

Abstract: The extensive use of logistic regression models in analytical epidemiology as well as in randomized clinical trials, often creates inflated estimates of the relative risk (RR). Particularly, in cases where a binary outcome has a high or moderate incidence in the studied population (>10%), the bias in assessing the relative risk may be very high. Meta-analysis studies have estimated that about 40% of the relative risk estimates in prospective investigations, through binary logistic models, lead to extensive bias of the population parameters. The problem of risk inflation also appears in cross-sectional studies with binary outcomes, where the parameter of interest is the prevalence ratio. As an alternative to the use of logistic regression models in both longitudinal and cross-sectional studies, the modified Poisson regression model is proposed.

Keywords: etiological studies; relative risk; logistic regression; modified Poisson regression



Citation: Gnardellis, C.; Notara, V.; Papadakaki, M.; Gialamas, V.; Chliaoutakis, J. Overestimation of Relative Risk and Prevalence Ratio: Misuse of Logistic Modeling. *Diagnostics* **2022**, *12*, 2851. <https://doi.org/10.3390/diagnostics12112851>

Academic Editor: Ching-Wei Wang

Received: 25 September 2022

Accepted: 16 November 2022

Published: 17 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Often in analytic epidemiology, the main interest of an investigation focuses on the possible causal relationship between a risk factor and the occurrence of a disease (or another binary outcome). The examination of the relationship between the factor and the disease can be conducted both in prospective and retrospective studies. In prospective studies, the association between the factor exposure and the disease is assessed using the relative risk or risk ratio (RR). Relative risk is defined as the ratio of the probability of the disease (outcome) in those exposed to the factor to the probability of the disease in the unexposed. Therefore, as a measure, RR estimates how much higher (or lower) is the chance of developing the disease for those exposed to the risk factor compared to the unexposed. The data of such an investigation can be summarized in a 2-by-2 table of the following form (Table 1):

Table 1. Notation for entries in a 2-by-2 table.

Factor	Outcome (Disease)		Total
	Event	No Event	
exposed	<i>a</i>	<i>b</i>	<i>a + b</i>
unexposed	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>n = a + b + c + d</i>

The RR in Table 1 is estimated by the equation

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)}$$

In cases where the etiology of rare diseases (or rare events) is being studied, or generally, in cases where the values of a and c are very small compared to those of b and d , the RR can be approximated by the odds ratio (OR).

The OR is defined as the odds of the disease occurrence between exposed and unexposed persons

$$OR = \frac{(a/(a+b))/(b/(a+b))}{(c/(c+d))/(d/(c+d))} = \frac{a/b}{c/d} = \frac{a \cdot d}{b \cdot c}$$

OR is a measure of the strength of the association between a binary factor and an outcome [1–3]. When the probability of the outcome (or the incidence of a disease) is low, the difference between the odds ratio and relative risk is negligible.

$$RR = \frac{a(c+d)}{c(a+b)} \approx \frac{a \cdot d}{b \cdot c}$$

Theoretically, both relative risk and odds ratio can take any positive value.

In case-control studies, the same definitions can be applied, except that RR cannot be assessed directly. In these investigations, the proportion of cases in the entire population at risk is unknown; therefore, one cannot measure the incidence of the outcome or disease. The assessment of relative risk in case-control studies is performed through OR, which adequately approximates RR, as long as the incidence of the disease is low (rare diseases) [2–4].

In cross-sectional investigations, the above definitions are modified and adapted to the cross-sectional concept of prevalence. In these studies, as a measure of the relationship between exposure and disease, the prevalence ratio (PR) is commonly used. That is, the prevalence of the disease in those exposed to a risk factor to the prevalence in the unexposed. The prevalence ratio summarizes the concept of risk in cross-sectional studies but without having the potential causality that exists in longitudinal investigations [2,3]. With the notation presented in Table 1, the prevalence ratio is defined as

$$PR = \frac{a/(a+b)}{c/(c+d)}$$

In the case of rare diseases, the ratio is adequately approximated by the prevalence odds ratio (POR)

$$POR = \frac{(a/(a+b))/(b/(a+b))}{(c/(c+d))/(d/(c+d))} = \frac{a/b}{c/d} = \frac{a \cdot d}{b \cdot c}$$

The odds ratio, as a cross-product ratio in a 2-by-2 table, has the property of “reciprocity”, which means changing the reference category of the binary outcome will yield reciprocal estimates [5–10]. Its assessment as a point estimate of the corresponding population ratio, as well as the assessment of its confidence intervals, are independent of the reference category of the outcome variable. Furthermore, it is known (from sampling theory) that the logarithmic transformation of OR, in either the form $\frac{a \cdot d}{b \cdot c}$ or the reciprocal $\frac{b \cdot c}{a \cdot d}$, is normally distributed with estimated standard error, in both cases, equal to $\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$. These properties make the logarithm of OR the natural measure (time-invariant) in stochastic-risk modeling [2,3].

It must be emphasized that reciprocity does not apply to the assessment of RR because its definition depends on the reference category of the outcome variable. In the case where RR is defined in terms of the occurrence of the disease, it is estimated by the quantity $\frac{a/(a+b)}{c/(c+d)}$, while in the opposite case (non-occurrence) by the quantity $\frac{b/(a+b)}{d/(c+d)}$ (Table 1).

Therefore, the interpretation of a comparison based on RR is critically important on whether the positive outcome or its negative complement is modeled [5–7].

In contrast to the odds ratio, relative risk has a notable arithmetic property called “collapsibility”. This means that a crude (unadjusted) risk ratio will not change if adjusted for a variable that is not a confounder. Due to collapsibility, assuming no confounding, the RR estimates the change in risk, on a ratio scale, for the entire exposed group due to exposure. Odds ratios are not collapsible, and therefore, the interpretation of an OR is more limited. The odds ratio will estimate the average change in odds for exposed individuals only if all individual odds ratios are the same and all individual risks without exposure are the same. With the exception of this unlikely situation, the crude OR will be closer to 1 than the average of stratum-specific or individual ORs. Even in the absence of confounding, the adjusted OR will be further from 1 than the unadjusted OR [11–16].

Odds ratio and relative risk are connected by the relationship $OR = \frac{1 - (\frac{c}{c+d})}{1 - (\frac{a}{a+b})} \cdot RR$. This arithmetic relation indicates that OR tends to overestimate RR by the factor $\frac{1 - (\frac{c}{c+d})}{1 - (\frac{a}{a+b})}$ when it is more than 1 or underestimates it when it is less than 1 [17]. However, in cases where the incidence or prevalence of a disease is low ($\leq 10\%$) OR approximates RR adequately [5]. The overestimation or underestimation of RR by OR tends to be further inflated when it is controlled for potential confounders, as it usually results in logistic regression. A simple method has been proposed to approximate the relative risk from an adjusted odds ratio (obtained from logistic regression) and derive an estimate of an association effect that better represents the true RR [17,18]. However, this method usually produces ratio estimates biased away from 1 when outcomes are common and risk among those not exposed varies substantially, suggesting that the strength of association is greater than is true [19,20]. Moreover, the proposed confidence interval for the adjusted RR, computed by applying the method to the bounds of the adjusted OR’s confidence interval, can also be biased, leading one to believe that the relative risk estimate is more precise than is true. Options exist to obtain unbiased estimates of relative risks in studies of common outcomes. As we will see below, there are methods that have widely available user-friendly software and are statistically appropriate to handle binary common outcomes.

It is common, in statistical analyses of a binary outcome, to use logistic regression as a predominant technique in assessing the risk of exposure or the benefit of a treatment [21,22]. Logistic regression is a linear model with a dependent variable, the log odds of a binary outcome

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1x_1 + B_2x_2 + \dots + B_kx_k$$

The basic advantage of logistic regression is the ability to yield odds ratio estimators adjusted for the effect of covariates. However, in cases where the incidence of a disease (or an outcome) is relatively high, the use of ORs frequently yields inflated estimates of the RRs [5,11,13,23–26].

Modified Poisson regression is an alternative to logistic regression, where the parameters are RRs rather than ORs. It is a generalized linear model with a linking function, the logarithm of the probability of a binary event (instead of the log odds) [27]. For the data summarized in Table 1, the probability of the disease in those exposed to the risk factor is defined by

$$p = \frac{a}{a+b}$$

A simple linear model, with outcome variable, logarithm of p and predictor the exposure variable X , is expressed by the equation

$$\ln(p) = B_0 + B_1x \tag{1}$$

For $x = 1$ (exposed) and $x = 0$ (unexposed) the values of (1) are

$$\ln(p_{(x=1)}) = B_0 + B_1 \cdot 1 = B_0 + B_1 \tag{2}$$

$$\ln(p_{(x=0)}) = B_0 + B_1 \cdot 0 = B_0 \tag{3}$$

The exponential transformation of (2) and (3) gives

$$p_{(x=1)} = e^{(B_0+B_1)} \tag{4}$$

and

$$p_{(x=0)} = e^{B_0} \tag{5}$$

RR is defined as the ratio of the probability of the outcome of the exposed ($x = 1$) divided by the probability of the unexposed ($x = 0$). Using Equations (4) and (5), RR is finally defined by

$$RR = \frac{p_{(x=1)}}{p_{(x=0)}} = \frac{e^{(B_0+B_1)}}{e^{B_0}} = \frac{e^{B_0} \cdot e^{B_1}}{e^{B_0}} = e^{B_1}$$

If a Poisson distribution is assumed, the model parameters are estimated by a maximum likelihood estimation algorithm (quasi-likelihood method, QL) [28]. Estimating the coefficients of the simple model results in

$$e^{B_0} = \frac{c}{c+d} \text{ and } e^{B_1} = \frac{\alpha(c+d)}{c(\alpha+b)}$$

The estimated standard error of RR, after being corrected for data overdispersion (using robust estimates for the covariance matrix), is given by [29]

$$\sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

In the case of multiple predictor variables, the simple Poisson model can be extended by the equation

$$\ln(p) = B_0 + B_1x_1 + B_2x_2 + \dots + B_kx_k \tag{6}$$

or, equivalently

$$p = e^{B_0+B_1x_1+B_2x_2+\dots+B_kx_k} , \tag{7}$$

where x_1, x_2, \dots, x_k are the values of a set of predictor variables X_1, X_2, \dots, X_k .

The interpretation of the model (6) is similar to the multiple logistic regression one, with the difference that its coefficients are referred to the probability (and not to the odds) of an outcome. Specifically, for each variable X_i , the quantity e^{B_i} is the factor by which the probability of an outcome is multiplied when the independent variable X_i increases by one unit (given that the other variables remain constant).

It should be mentioned that the above use of Poisson’s regression is an exceptional use of Poisson’s model, which is commonly applied to count data. This specific model application, with the error variance correction, is generally referred to as modified Poisson regression or robust Poisson regression [27].

2. Materials and Methods

In a clinical study, survival of mixed lobular and ductal carcinomas of breast cancer was compared with a series of patients with pure invasive ductal ones. One hundred and thirty-two patients, ninety-eight with ductal carcinomas and thirty-four with mixed who had been treated by modified radical mastectomy, were followed for 10 years [26]. Survival data were available for all patients with a minimum of five-year follow-up period. At the

end of follow-up time, their survivorship was assessed in relation to the histological type of tumor. The data are presented in Table 2.

Table 2. Survivorship of 132 women at the end of five-year follow-up period.

		Dead	Alive	Total
Histological type	Mixed	26	8	34
		76.5%	23.5%	100.0%
	Ductal	38	60	98
		38.8%	61.2%	100.0%
Total		64	68	132
		48.5%	51.5%	100.0%

The OR and RR of five-year mortality for the mixed versus ductal type were, respectively:

$$OR = \frac{a \cdot d}{b \cdot c} = \frac{26 \cdot 60}{8 \cdot 38} = 5.13,$$

$$RR = \frac{a(c + d)}{c(a + b)} = \frac{26 \cdot 98}{38 \cdot 34} = 1.97$$

Moreover, the 95% Confidence Intervals (CIs) of OR and RR were: 95% CI OR (2.11, 12.50) and 95% CI RR (1.45, 2.69).

3. Results

Using a simple logistic regression model, with the five-year mortality as a binary outcome (dead/alive) and the histological type as a predictor, the estimated OR for mixed versus ductal carcinoma mortality at the end of five-year period is the same as in the simple 2-by-2 table, that is, about 2.5 times the RR (Table 3).

Table 3. Simple logistic and Poisson regression models with five-year mortality as the outcome variable.

Logistic Regression	B	OR	p-Value	95% CI OR
Histological type (mixed vs. ductal)	1.64	5.13	<0.001	2.11–12.50
Poisson Regression	B	RR	p-Value	95% CI RR
Histological type (mixed vs. ductal)	0.68	1.97	<0.001	1.45–2.69

Furthermore, using a multiple logistic regression model controlling for age, grade of malignancy (Grade II, Grade III), possible lymphocytic infiltration, and the number of lymph nodes metastases (≤ 3 , >3), the estimated OR is further inflated to 5.25 (Table 4).

The results of logistic model indicate that adjusted OR tends to overestimate RR, even to a greater extent than unadjusted one. Moreover, simulation studies have shown that estimated CIs for the OR are quite wide, while the point estimation of the ratio appears to have inconsistency problems (that is, the bias does not decrease as the sample size increases) [19,23,29]. When estimating RR of an outcome, the use of logistic regression model is considered abusive if the outcome incidence is rare (<10%). For common outcomes, OR always overstates RR, sometimes dramatically. In these cases, the modified Poisson regression model has been proposed [29–31].

Table 4. Multiple logistic and Poisson regression models with five-year mortality as the outcome variable.

Logistic Regression	B	OR	p-Value	95% CI OR
Histological type (mixed vs. ductal)	1.66	5.25	<0.001	1.94–14.22
Age of patients (in years)	0.01	1.01	0.723	0.95–1.08
Grade of malignancy (Grade III vs. Grade II)	0.44	1.55	0.360	0.61–3.95
Lymphocytic infiltration	0.66	1.93	0.219	0.68–5.51
Lymph node metastases (>3)	0.32	1.38	0.430	0.62–3.10
Poisson Regression	B	RR	p-Value	95% CI RR
Histological type (mixed vs. ductal)	0.66	1.94	<0.001	1.36–2.75
Age of patients (in years)	0.01	1.01	0.688	0.98–1.03
Grade of malignancy (Grade III vs. Grade II)	0.17	1.18	0.380	0.81–1.72
Lymphocytic infiltration	0.29	1.34	0.250	0.81–2.22
Lymph node metastases (>3)	0.16	1.17	0.368	0.82–1.68

The modified Poisson regression model will be used to reanalyze the survival data of breast cancer. The results of the modified Poisson regression, with dependent variable the five-year mortality and predictor the histological type, are presented in Table 3. The table includes the regression-derived coefficient (B), the estimation of RR, the Wald test, and the CI for RR. The point and interval estimation of RR are identical to those calculated directly by the contingency table (RR = 1.97, 95% CI: 1.45–2.69).

In order to adjust for covariates, the survival data were reanalyzed, controlling for histological type, degree of malignancy, possible lymphocytic infiltration, number of metastatic lymph nodes, and age (Table 4).

The adjusted RR for mixed versus ductal adenocarcinoma five-year mortality still remains about twice as high (RR = 1.94, $p < 0.001$). Moreover, the CIs for all predictor variables are robust and consistent compared to those obtained from the logistic regression model. Specifically, the 95% CI for the histological type is (1.36, 2.75) versus (1.94, 14.22) of the logistic model.

It is evident that the assessment of RR in Poisson regression is more coherent compared to the logistic model. The difference from its unadjusted assessment in the 2-by-2 contingency table between the histological type and the present situation is negligible (1.94 vs. 1.97), implying that its adjustment for potential confounding variables is small. Furthermore, the introduction of multiple independent variables in the Poisson model yields more efficient coefficient estimators than the logistic model (more compact CIs).

The estimation of RR through the Poisson model is performed in terms of a specific outcome (death in our example) without the “reciprocity” that appears in logistic model. That is, while in logistic regression, regardless of the outcome against which the risk is estimated, the exponentiated coefficients are reciprocal of each other (with the significance remaining the same), in the Poisson model, both the coefficients and statistical significance are modified according to the estimated outcome. This differentiation may seem strange to those familiar with the use of logistic regression, but it is consistent with the rationale of Poisson model since the dependent variable in the latter (i.e., the logarithm of the relative risk) varies according to the outcome of interest. In extreme cases, there may be statistical significance for one category of the binary outcome and not for the other [5].

The breast cancer data will be reanalyzed to show the differences in the symmetry (reciprocity) between the logistic regression and the Poisson model results. In the new analysis, the binary outcome will be survival and not death (Table 5). Exponentiated coefficients of logistic model (that is, RRs) are reciprocal of the original ones (with death as a reference event), while in the Poisson regression, they differ in both values and statistical significance. For example, the adjusted RR for survival of mixed adenocarcinoma versus ductal is $OR = 0.19 = 1/5.25$, $p = 0.001$, of Grade III versus Grade II, $OR = 0.65 = 1/1.55$, $p = 0.360$, etc. On the contrary, reciprocity does not apply to the Poisson regression coefficients: for mixed adenocarcinoma vs. ductal, $RR = 0.40 \neq 1/1.94 = 0.52$, $p = 0.005$, for Grade III vs. Grade II, $RR = 0.79 \neq 1/1.18 = 0.85$, $p = 0.353$, etc.

Table 5. Multiple logistic and Poisson regression models with five-year survival as the outcome variable.

Logistic Regression	B	OR	p-Value	95% CI OR
Histological type (mixed vs. ductal)	−1.66	0.19	0.001	0.07–0.52
Age of patients (in years)	−0.01	0.99	0.723	0.93–1.06
Grade of malignancy (Grade III vs. Grade II)	−0.44	0.65	0.360	0.25–1.65
Lymphocytic infiltration	−0.66	0.52	0.219	0.18–1.48
Lymph node metastases (>3)	−0.32	0.72	0.430	0.32–1.62
Poisson Regression	B	RR	p-Value	95% CI RR
Histological type (mixed vs. ductal)	−0.92	0.40	0.005	0.21–0.76
Age of patients (in years)	−0.01	0.99	0.659	0.97–1.02
Grade of malignancy (Grade III vs. Grade II)	−0.24	0.79	0.353	0.48–1.30
Lymphocytic infiltration	−0.24	0.78	0.208	0.54–1.15
Lymph node metastases (>3)	−0.12	0.88	0.434	0.65–1.20

4. Discussion

Binary logistic regression is widely used in empirical research, sometimes even uncritically. Based on the functional form of the logistic model, it results that the exponentiated coefficient of a dichotomous independent variable (defining two groups) estimates the OR of the binary outcome. If the investigation simply regards the existence of a significant relationship between the independent variable and the outcome with no interest in unbiased estimation of RR, logistic regression is extremely useful with extensive applications.

However, if the focus is on estimating the RR of a binary outcome, then the use of OR, as derived from the exponentiated coefficients of the logistic regression model, is unsafe. This estimate is very close to the RR when the outcome has a low incidence in the studied data ($\leq 10\%$). Nevertheless, in cases where the incidence is higher, the use of logistic regression tends to overestimate the RR and sometimes misleadingly inflate it. In prospective epidemiological studies, when exposure to a risk factor is evaluated, the logistic regression model should be used with great caution. In these cases, where the estimation of RR can be performed directly from the data, it is suggested to use the modified Poisson model as described previously. Especially in cases of common outcomes with a frequency higher than 10% in the study data, the use of the modified Poisson regression is considered necessary. The same procedure is proposed for the analyses of randomized clinical trials [12]. Meta-analysis studies have estimated that approximately 40% of the RR assessment in prospective investigations through logistic regression models leads to extensive biases of the corresponding population parameters [32].

The binary logistic model can be used in a reliable and consistent manner if the outcome has a low incidence in the sample data. It can also be used in studies with the main interest in predicting an outcome or estimating the strength and direction of the association between the independent variables and the dichotomous outcome. In these latter cases, where there is no need for unbiased estimation of RR, logistic regression works satisfactorily with information-rich output. Furthermore, in case-control epidemiological studies, the use of logistic regression is the only option since the RR cannot be directly estimated from the available data. In these specific investigations, the only possible option is to estimate the ORs.

In cross-sectional studies, the use of logistic regression is subject to the same limitations as in longitudinal studies [5,33–35]. In these cases, since we are interested in unbiasedly estimating the RR of an outcome, the use of the modified Poisson model is suggested (since the outcome has a low incidence in the available data). If the purpose of a cross-sectional study is not to estimate the relationship (in terms of strength and direction) of a dichotomous variable and an outcome but to unbiasedly estimate the prevalence ratio in two groups, then the aforementioned limitations are equally applied.

Apart from the modified Poisson regression, as an alternative procedure for estimating relative risk, the log binomial model has been proposed and is widely used. The latter has similar power to the robust Poisson regression (possibly, in some cases, providing slightly

smaller standard errors of the relative risk) but often exhibits convergence problems of the maximum likelihood algorithm it uses [36,37]. Moreover, it generates incorrect estimates when the covariate values in the data are not bounded by one (that is, the case of continuous covariates) [36]. Results similar to those of Poisson regression are also provided by the Cox proportional hazard model with robust variance estimation. The Cox model is equivalent to Poisson regression if the follow-up time for all subjects in the study is considered constant (in the analysis, we can set the follow-up time to one or any quantity that is the same for all subjects). Another simple alternative for risk estimation is to use the Mantel–Haenszel method for relative risk, which can adjust for one or two confounders [38,39]. However, this method becomes inefficient when several factors, especially continuous variables, are being adjusted for simultaneously. The literature is rich regarding the effectiveness and unbiasedness of all previous methods. Considering the weaknesses and strengths of the methods, as well as their ease of use and availability in commercial data analysis software, Poisson regression seems to have the most advantages [31,40–42]. Nevertheless, researchers should keep in mind that due to the lack of symmetry in Poisson regression coefficients (as with all models with the relative risk as a dependent variable), the choice of reference category is critically important in whether the positive outcome or its negative complement is modeled. Moreover, the study results and inference should not be based solely on statistical significance but also on clinical significance, which sometimes is overlooked, and undue importance to statistical measures is given.

5. Conclusions

The uncritical application of logistic regression in epidemiological studies often creates inflated estimates of the RR. The basic advantage of logistic regression is the ability to yield approximations of the RRs through odds ratios adjusted for the effect of possible covariates. However, in cases where the incidence of a disease or an outcome is relatively high, ORs can substantially overestimate RRs. An alternative to the use of logistic regression is the modified Poisson model. As a basic criterion for choosing the appropriate method, it is suggested primarily to construct a 2-by-2 contingency table between the risk factor and the outcome and then calculate both the OR and the RR. If the two values differ notably, the use of the modified Poisson model is recommended. Since this procedure is implemented in standard statistical packages, there is no justification for relying on logistic regression when the relative risk is the parameter of primary interest.

In a future paper, validation of the performance of modified Poisson regression, in terms of relative bias for point estimation and percentage of confidence interval coverage, will be undertaken through simulated and real data sets.

Author Contributions: Conceptualization, original draft preparation, C.G. and V.N.; supervision, C.G.; writing—review and editing, M.P., V.G. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Macmahon, B.; Trichopoulos, D. *Epidemiology: Principles and Methods*, 2nd ed.; Little Brown and Company: Boston, MA, USA, 1996; pp. 65–84.
2. Rothman, K.J.; Greenland, S. *Modern Epidemiology*, 2nd ed.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 1998; pp. 67–114.
3. Kleinbaum, D.; Kupper, L.; Morgenstern, H. *Epidemiologic Research*; Van Nostrand Reinhold: New York, NY, USA, 1982; pp. 97–148.
4. Greenland, S.; Thomas, D.C. On the need for the rare disease assumption in case–control studies. *Am. J. Epidemiol.* **1982**, *116*, 547–553. [[CrossRef](#)] [[PubMed](#)]

5. Tamhane, A.R.; Westfall, A.O.; Burkholder, G.A.; Cutter, G.R. Prevalence Odds Ratio versus Prevalence Ratio: Choice Comes with *Conseq. Stat. Med.* **2016**, *35*, 5730–5735. [[CrossRef](#)] [[PubMed](#)]
6. Eckermann, S.; Coory, M.; Willan, A.R. Indirect comparison: Relative risk fallacies and odds solution. *J. Clin. Epidemiol.* **2009**, *62*, 1031–1036. [[CrossRef](#)] [[PubMed](#)]
7. Doi, S.A.; Furuya-Kanamori, L.; Xu, C.; Lin, L.; Chivese, T.; Thalib, L. Controversy and Debate: Questionable utility of the relative risk in clinical research: Paper 1: A call for change to practice. *J. Clin. Epidemiol.* **2022**, *142*, 271–279. [[CrossRef](#)]
8. Walter, S.D. Choice of effect measure for epidemiological data. *J. Clin. Epidemiol.* **2000**, *53*, 931–939. [[CrossRef](#)]
9. Senn, S.; Walter, S.; Olkin, I. Odds ratio revisited. *Evid. Based Med.* **1998**, *3*, 71.
10. Cook, T.D. Advanced statistics: Up with odds ratios! A case for odds ratios when outcomes are common. *Acad. Emerg. Med.* **2002**, *9*, 1430–1434. [[CrossRef](#)] [[PubMed](#)]
11. Greenland, S. Interpretation and choice of effect measures in epidemiologic analyses. *Am. J. Epidemiol.* **1987**, *125*, 761–768. [[CrossRef](#)] [[PubMed](#)]
12. Newman, S.C. *Biostatistical Methods in Epidemiology*; Wiley: New York, NY, USA, 2001.
13. Greenland, S. Noncollapsibility, confounding, and sparse-data bias. Part 1: The oddities of odds. *J. Clin. Epidemiol.* **2021**, *138*, 178–181. [[CrossRef](#)]
14. Greenland, S. Noncollapsibility, confounding, and sparse-data bias. Part 2: What should researchers make of persistent controversies about the odds ratio? *J. Clin. Epidemiol.* **2021**, *139*, 264–268. [[CrossRef](#)]
15. Huitfeldt, A.; Stensrud, M.J.; Suzuki, E. On the collapsibility of measures of effect in the counterfactual causal framework. *Emerg. Themes Epidemiol.* **2019**, *16*, 1. [[CrossRef](#)] [[PubMed](#)]
16. Cummings, P. The Relative Merits of Risk Ratios and Odds Ratios. *Arch. Pediatr. Adolesc. Med.* **2009**, *163*, 438–445. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, J.; Yu, K.F. What's the Relative Risk? A method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. *JAMA* **1998**, *280*, 1690–1691. [[CrossRef](#)]
18. Holland, P.W. A note on the covariance of the Mantel–Haenszel log-odds-ratio estimator and the sample marginal rates. *Biometrics* **1989**, *45*, 1009–1016. [[CrossRef](#)] [[PubMed](#)]
19. McNutt, L.A.; Chuntao, W.; Xiaonan, X.; Hafner, J.P. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am. J. Epidemiol.* **2003**, *157*, 940–943. [[CrossRef](#)]
20. Greenland, S.; Holland, P.W. Estimating standardized risk differences from odds ratios. *Biometrics* **1991**, *47*, 319–322. [[CrossRef](#)]
21. Hosmer, W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; John Wiley and Sons: New York, NY, USA, 2000.
22. Collett, D. *Modelling Binary Data*; Chapman and Hall: London, UK, 1991.
23. 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*, 301–305. [[CrossRef](#)]
24. Knol, M.J.; Le Cessie, S.; Algra, A.; Vandenbroucke, J.P.; Groenwold, R.H. Overestimation of risk ratios by odds ratios in trials and cohort studies: Alternatives to logistic regression. *CMAJ* **2012**, *184*, 895–899. [[CrossRef](#)]
25. Richardson, S.T.; Robins, M.J.; Wang, L. On Modeling and Estimation for the Relative Risk and Risk Difference. *JASA* **2017**, *112*, 1121–1130. [[CrossRef](#)]
26. Agnantis, N.J.; Behraki, E.; Gnardellis, C.; Vassilaki, I.; Mahera, H. Comparative study of 5-year survival in no special type ductal and mixed type breast cancer. *Breast* **1992**, *1*, 196–198. [[CrossRef](#)]
27. Winkelmann, R.; Zimmermann, K.F. *Robust Poisson Regression*; Fahrmeir, L., Francis, B., Gilchrist, R., Tutz, G., Eds.; Springer: New York, NY, USA, 1992; pp. 201–206.
28. McCullagh, P. Quasi-Likelihood Functions. *Ann. Stat.* **1983**, *11*, 59–67. [[CrossRef](#)]
29. Zou, G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am. J. Epidemiol.* **2004**, *159*, 702–706. [[CrossRef](#)] [[PubMed](#)]
30. Yelland, L.N.; Salter, A.B.; Ryan, P. Performance of the Modified Poisson Regression Approach for Estimating Relative Risks from Clustered Prospective Data. *Am. J. Epidemiol.* **2011**, *174*, 984–992. [[CrossRef](#)] [[PubMed](#)]
31. Chen, W.; Qian, L.; Shi, J.; Franklin, M. Comparing performance between log-binomial and robust Poisson regression models for estimating risk ratios under model misspecification. *BMC Med. Res. Methodol.* **2018**, *18*, 63. [[CrossRef](#)] [[PubMed](#)]
32. Knol, M.J.; Duijnhoven, R.G.; Grobbee, D.E.; Moons, K.G.M.; Groenwold, R.H. Potential Misinterpretation of Treatment Effects Due to Use of Odds Ratios and Logistic Regression in Randomized Controlled Trials. *PLoS ONE* **2011**, *6*, e21248. [[CrossRef](#)] [[PubMed](#)]
33. Barros, A.J.D.; Hirakata, V.N. 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*, 21. [[CrossRef](#)]
34. Jean-Luis, G.; Turner, A.D.; Seixas, A.; Jin, P.; Rosenthal, D.; Liu, M.; Avirappatu, G. Epidemiologic Methods to Estimate Insufficient Sleep in the US Population. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9337. [[CrossRef](#)]
35. Martinez, B.A.F.; Leotti, V.B.; Silva, G.S.E.; Nunes, L.N.; Machado, G.; Corbellini, L.G. Odds Ratio or Prevalence Ratio? An Overview of Reported Statistical Methods and Appropriateness of Interpretations in Cross-sectional Studies with Dichotomous Outcomes in Veterinary Medicine. *Font. Vet. Sci.* **2017**, *4*, 193. [[CrossRef](#)]
36. Petersen, R.M.; Deddens, A.J. A comparison of two methods for estimating prevalence ratios. *BMC Med. Res. Methodol.* **2008**, *28*, 9. [[CrossRef](#)]

37. Williamson, T.; Misha Eliasziw, M.; Fick, H.G. Log-binomial models: Exploring failed convergence. *Emerg. Themes Epidemiol.* **2013**, *10*, 14. [[CrossRef](#)]
38. Nurminen, M. Asymptotic Efficiency of General Noniterative Estimators of Common Relative Risk. *Biometrika* **1981**, *68*, 525–530. [[CrossRef](#)]
39. Tarone, E.R. On summary estimators of relative risk. *J. Chronic Dis.* **1981**, *34*, 463–468. [[CrossRef](#)]
40. Chen, W.; Shi, J.; Qian, L.; Azen, S.P. Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: A simulation study. *BMC Med. Res. Methodol.* **2014**, *14*, 82. [[CrossRef](#)] [[PubMed](#)]
41. Gallis, A.J.; Turner, L.E. Relative Measures of Association for Binary Outcomes: Challenges and Recommendations for the Global Health Researcher. *Ann. Glob. Health* **2019**, *85*, 1–12. [[CrossRef](#)]
42. Coutinho, L.; Scazufca, M.; Menezes, P.R. Methods for estimating prevalence ratios in cross-sectional studies. *Rev. Saud. Publica* **2008**, *42*, 992–998. [[CrossRef](#)]