

# Effect Sizes for $2 \times 2$ Contingency Tables

Jake Olivier<sup>1\*</sup>, Melanie L. Bell<sup>2</sup>

**1** School of Mathematics and Statistics, University of New South Wales, Sydney, Australia, **2** Psycho-Oncology Co-Operative Research Group, School of Psychology, University of Sydney, Sydney, Australia

## Abstract

Sample size calculations are an important part of research to balance the use of resources and to avoid undue harm to participants. Effect sizes are an integral part of these calculations and meaningful values are often unknown to the researcher. General recommendations for effect sizes have been proposed for several commonly used statistical procedures. For the analysis of  $2 \times 2$  tables, recommendations have been given for the correlation coefficient  $\phi$  for binary data; however, it is well known that  $\phi$  suffers from poor statistical properties. The odds ratio is not problematic, although recommendations based on objective reasoning do not exist. This paper proposes odds ratio recommendations that are anchored to  $\phi$  for fixed marginal probabilities. It will further be demonstrated that the marginal assumptions can be relaxed resulting in more general results.

**Citation:** Olivier J, Bell ML (2013) Effect Sizes for  $2 \times 2$  Contingency Tables. PLoS ONE 8(3): e58777. doi:10.1371/journal.pone.0058777

**Editor:** Fabio Rapallo, University of East Piedmont, Italy

**Received:** December 19, 2012; **Accepted:** February 6, 2013; **Published:** March 7, 2013

**Copyright:** © 2013 Olivier and Bell. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: j.olivier@unsw.edu.au

## Introduction

Sample size calculations are an integral part of scientifically useful and ethical research [1]. A study which is too small may not answer the research question, wasting resources and potentially putting participants at risk for no purpose [2]. Studies which are too large can also waste resources and expose participants to the potential harms of research needlessly, as well as delaying results and their translation into practice. The computation of sample size *a priori* is usually dependent upon predetermined values for power and level of significance, an estimate of the expected variability in the sample and an effect size of practical or clinical importance. By convention, the choice of power and level of significance is usually at least 80% and no more than 5% respectively. When a practically important effect size is unknown, there are several recommendations in the literature to guide the researcher. In his seminal paper, Cohen [3] gives operationally defined small, medium and large effect sizes for various, common significance tests. The use of effect size recommendations should not replace differences of clinical or practical importance [4] and may not be appropriate for all disciplines. In basic science research, for example, large effect sizes by Cohen's criteria are common and, therefore, require small sample sizes. On the other hand, clinical and epidemiological research often deals with small effect sizes and often requires large, population-based studies. While there are some approaches to estimating a minimum important effect [5], there are instances where this information is simply not known. Thus, effect size recommendations assist with the balance between overly small and overly large sample sizes.

When the researcher is interested in  $2 \times 2$  contingency tables, a common measure of effect size is  $\phi$  which, in this instance, is equivalent to Pearson's correlation coefficient [6]. Cohen [3] recommends effect sizes of  $\phi = 0.1, 0.3$  and  $0.5$  for small, medium and large effect sizes respectively and are identical to his recommendations for the correlation coefficient. Although Cohen

[3] denotes this statistic as  $\omega$ , much of the literature uses  $\phi$  [6–10] and the remainder of this manuscript follows this convention. To support his recommended effect sizes for correlation coefficients, Cohen [11] chose equivalent values for the difference in two means through the connection with point biserial correlation. Additionally,  $\phi$  is applicable to logistic regression since it can be converted to an odds ratio (*OR*) when the row (or column) marginal probabilities of the  $2 \times 2$  table are fixed. For example, when the marginal probabilities are uniform (i.e., 0.5 for row and column probabilities), Cohen's recommended effect sizes are equivalent to odds ratios of 1.49, 3.45 and 9.0. It will be demonstrated that the connection between the odds ratio and  $\phi$  is largely dependent on the marginal probabilities and these *OR* values should not be used in general.

A problem arises when using the effect size  $\phi$  for  $2 \times 2$  tables as the full range of correlation coefficients are only possible under very restrictive circumstances and are not justified in general [12]. On the other hand, odds ratios are valid effect size measures that are not constrained by the marginal probabilities. Ferguson [10] recommends small, medium, and large odds ratio effect sizes of 2.0, 3.0 and 4.0, but urges caution in their use as they are not "anchored" to Pearson's correlation coefficient. Although many have pointed out problems with  $\phi$  as an association measure and advocate the use of odds ratios as an alternative, effect size recommendations for odds ratios do not exist in general.

It is common in randomised controlled trials and case-control studies to fix one of the marginal probabilities in the  $2 \times 2$  table as it directly relates to the ratio of participant allocation. For instance, a marginal probability of 0.5 corresponds to a 1:1 case-control ratio while a 2:1 ratio is a marginal probability of 0.67 (or equivalently 0.33 for 1:2).

The aims of this paper are to demonstrate: (1) the equivalence of effect size measures for  $2 \times 2$  contingency tables, in particular the relationship between  $\phi$  and the odds ratio; (2) that recommended odds ratio effect sizes can be derived from Cohen's work using the

**Table 1.** 2x2 contingency table of probabilities.

	$X=0$	$X=1$	Total
$Y=0$	$\pi_{00}$	$\pi_{01}$	$\pi_{0+}$
$Y=1$	$\pi_{10}$	$\pi_{11}$	$\pi_{1+}$
Total	$\pi_{+0}$	$\pi_{+1}$	1.0

doi:10.1371/journal.pone.0058777.t001

maximum value of  $\phi$  as a guideline for fixed marginal probabilities; (3) the shortcomings of  $\phi$  and the strength of the odds ratio as an effect size measure; and (4) that conservative odds ratio effect size recommendations can be derived without relying on fixed margins. We provide an example that investigates the association between helmet wearing by bicyclists and overtaking distance by automobiles.

**Equivalence of Effect Size Measures for 2x2 Contingency Tables**

**2 x 2 Contingency tables**

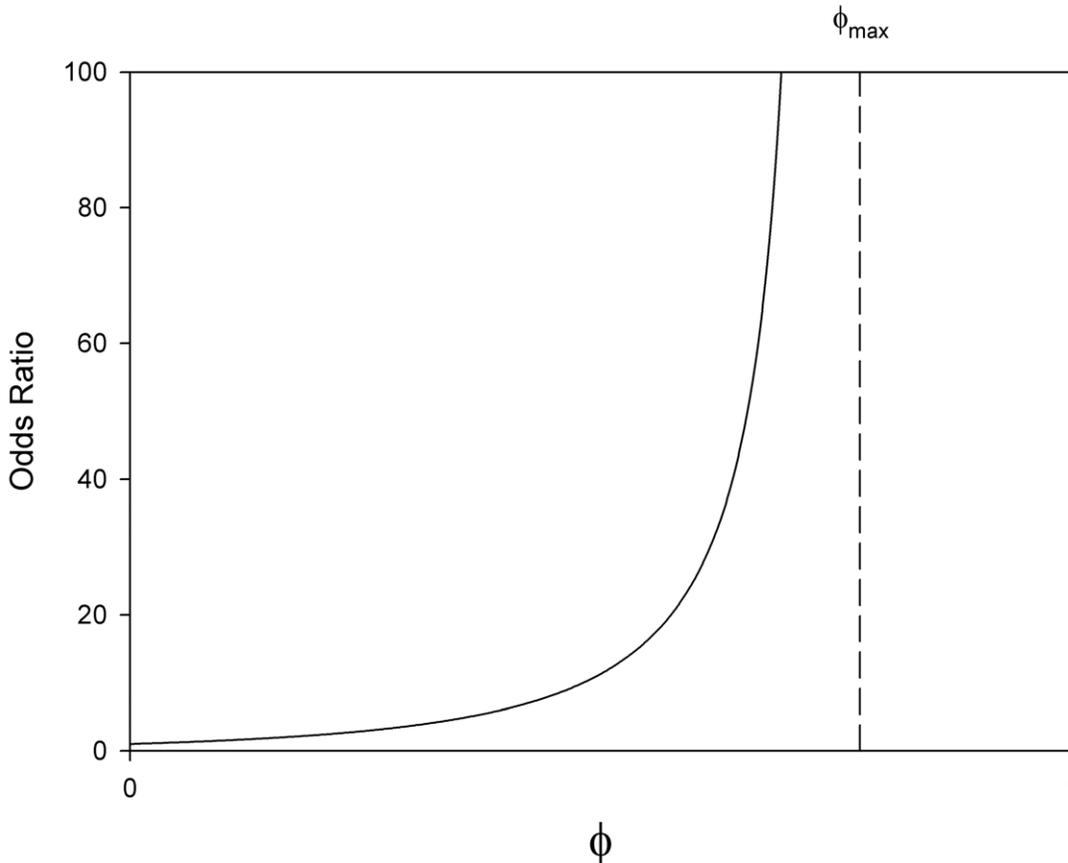
The two-way classification or contingency table is a common method for summarising the relationship between two binary variables, say  $X$  and  $Y$ . Table 1 gives the joint probability distribution of  $X$  and  $Y$  when their individual outcomes are from the set  $\{0,1\}$ .

In this formulation,  $\pi_{ij} = P(X=i, Y=j)$ , for  $i=0,1, j=0,1$ , is the joint probability of  $X$  and  $Y$ ,  $\pi_{i+} = P(X=i)$  is the marginal probability of  $X$ , and  $\pi_{+j} = P(Y=j)$  is the marginal probability of  $Y$ . Under an assumption of independence between  $X$  and  $Y$ , the product of the marginal probabilities equals the cell probabilities, i.e.,  $\pi_{ij} = \pi_{i+} \pi_{+j}$ . Alternatively, the 2x2 table could be represented by the frequency of observations so that  $n_{ij} = n \times \pi_{ij}$  where  $n = \sum_{i,j} n_{ij}$ . Similarly, the marginal frequencies are  $n_{i+} = n_{i0} + n_{i1}$  and  $n_{+j} = n_{0j} + n_{1j}$ . Note that  $\pi_{ij}$  is assumed to be the population proportion as the focus of this paper is the use of effect sizes as a planning tool and not statistical inference per se. In a case-control study, for example,  $X$  may indicate the presence or absence of disease while  $Y$  is an indication of exposure. Thus,  $\pi_{11} = P(X=1, Y=1)$  would represent the joint probability of being diseased and exposed.

**Effect size  $\phi$  and Equivalences for 2 x 2 Tables**

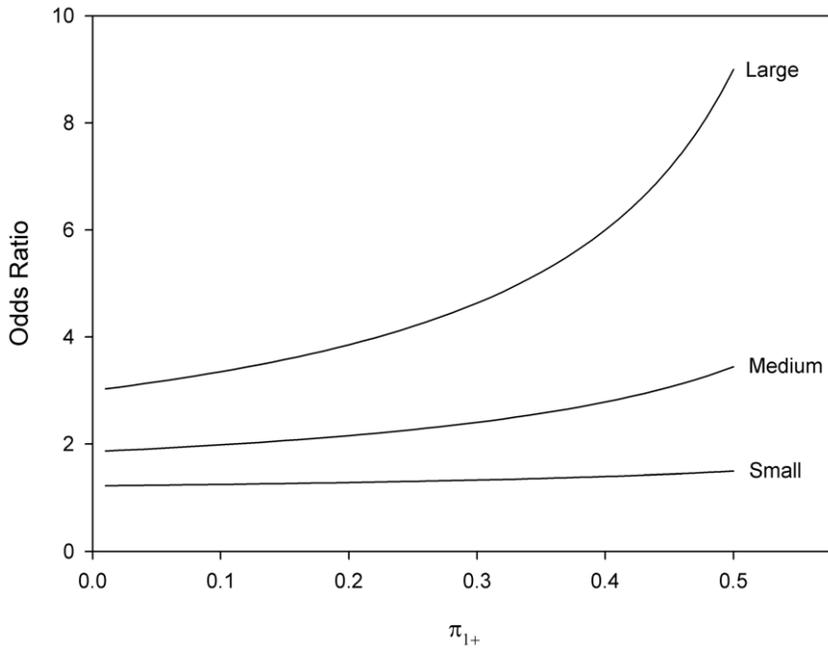
There are many association measures applicable to 2x2 tables which, with the exception of the odds ratio and relative risk, are equivalent or similar to  $\phi$ . The equivalence of some of these association measures is outlined below.

For the random sample  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ , Pearson's correlation coefficient is



**Figure 1.** Relationship between the odds ratio and  $\phi$  for unequal marginal probabilities.

doi:10.1371/journal.pone.0058777.g001



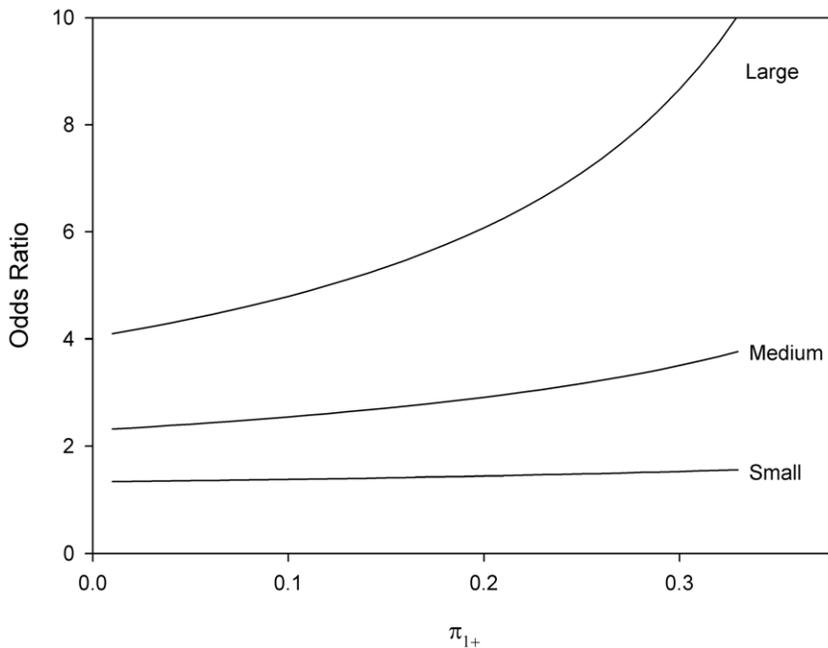
**Figure 2. Odds ratios and marginal probability by small, medium and large effect sizes for 1:1 allocation.**  
doi:10.1371/journal.pone.0058777.g002

$$r = \frac{\sum_{\ell=1}^n (X_{\ell} - \bar{X})(Y_{\ell} - \bar{Y})}{\sqrt{\sum_{\ell=1}^n (X_{\ell} - \bar{X})^2 \sum_{\ell=1}^n (Y_{\ell} - \bar{Y})^2}}$$

where  $\bar{X}$  and  $\bar{Y}$  are the sample means of the  $X_{\ell}$  and  $Y_{\ell}$  respectively. Although used primarily as a measure of linear association, Pearson's correlation coefficient can be applied to

binary variables and is often given the notation  $\phi$ . For the  $2 \times 2$  table case, we get

$$\sum_{\ell=1}^n (X_{\ell} - \bar{X})(Y_{\ell} - \bar{Y}) = n_{11} - \frac{n_{1+}n_{+1}}{n}$$



**Figure 3. Odds ratios and marginal probability by small, medium and large effect sizes for 1:2 allocation.**  
doi:10.1371/journal.pone.0058777.g003

**Table 2.** Sample sizes calculated for small, medium and large effect sizes for 1:1 allocation, 80% power and  $\alpha=0.05$ .

Odds Ratio	$\pi_{1+}$								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.22	8168	4688	3646	3254	3188	3386	3948	5282	9576
1.86	724	436	354	330	338	374	454	632	1188
3.00	200	128	110	108	116	134	170	246	480

doi:10.1371/journal.pone.0058777.t002

$$\sum_{\ell=1}^n (X_{\ell} - \bar{X})^2 = \frac{n_{0+}n_{1+}}{n}$$

$$\sum_{\ell=1}^n (Y_{\ell} - \bar{Y})^2 = \frac{n_{+0}n_{+1}}{n}$$

$$\omega = \phi = \text{sgn}(\pi_{11} - \pi_{1+}\pi_{+1}) \sqrt{\sum_{i=0}^1 \sum_{j=0}^1 \frac{(\pi_{ij} - \pi_{i+}\pi_{+j})^2}{\pi_{i+}\pi_{+j}}}$$

Simple arithmetic demonstrates the equivalence of  $\omega$  with  $\phi$ . The *sgn* function is used to give the appropriate sign since the chi-square statistic is inherently non-directional.

So, Pearson’s correlation coefficient for binary random variables  $X$  and  $Y$  is

$$r = \phi = \frac{\pi_{11} - \pi_{1+}\pi_{+1}}{\sqrt{\pi_{1+}\pi_{+1}(1 - \pi_{1+})(1 - \pi_{+1})}}$$

Since  $\pi_{11} = \pi_{1+}\pi_{+1}$  under the hypothesis of independence,  $\phi$  can be interpreted as measuring the departure from independence between  $X$  and  $Y$ . Note that Cramér’s  $\phi$  is equivalent to this equation for the  $2 \times 2$  table case [11] as well as the square root of Goodman and Kruskal’s  $\tau$  [13].

For the analysis of contingency tables, in general (not just the  $2 \times 2$  table case) the effect size formula for  $K$  total cells is

$$\omega = \sqrt{\sum_{k=1}^K \frac{(P_{1k} - P_{0k})^2}{P_{0k}}}$$

where  $P_{0k}$  and  $P_{1k}$  are cell probabilities under the null and alternative hypotheses respectively. Note that  $\omega$  is related to the usual chi-square statistic  $\chi^2$  by  $\omega = \sqrt{\chi^2/n}$  and is sometimes called the contingency coefficient. Using this formula, Cohen [3] recommends  $\omega=0.1, 0.3$  and  $0.5$  for small, medium and large effect sizes. Making note that  $P_{1k}$  is the probability of each cell ( $\pi_{ij}$ ) and  $P_{0k}$  is the cell probability under an independence assumption (so that  $P_{0k} = \pi_{i+}\pi_{+j}$ ), we can then write the effect size formula for the  $2 \times 2$  table as follows

**The relationship of  $\phi$  to the odds ratio**

The odds ratio for the association between  $X$  and  $Y$  is  $\pi_{11}\pi_{00}/(\pi_{10}\pi_{01})$ . When the marginal probabilities are held constant and the cell probability  $\pi_{11}$  is known, the remaining cell probabilities can be written as

$$\pi_{01} = \pi_{+1} - \pi_{11}$$

$$\pi_{10} = \pi_{1+} - \pi_{11}$$

$$\pi_{00} = \pi_{11} - \pi_{+1} + \pi_{0+}$$

Therefore, when the marginal probabilities are fixed, the odds ratio can be computed directly from  $\pi_{11}$ , which can then be expressed as

$$OR = \frac{\pi_{11}^2 - \pi_{11}(\pi_{1+} + \pi_{+1}) + \pi_{11}}{\pi_{11}^2 - \pi_{11}(\pi_{1+} + \pi_{+1}) + \pi_{1+}\pi_{+1}}$$

It is clear from the above formula that the odds ratio will be greater than one (or less than one) precisely when the joint probability  $\pi_{11}$  is greater (or less) than expected under an assumption of independence, i.e.,  $\pi_{11} > \pi_{1+}\pi_{+1}$ . Additionally, the formula for  $\phi$  can be rearranged to solve for  $\pi_{11}$ , i.e.,

$$\pi_{11} = \pi_{1+}\pi_{+1} + \phi \sqrt{\pi_{1+}\pi_{+1}(1 - \pi_{1+})(1 - \pi_{+1})}$$

Although mathematically unattractive, it is clear the odds ratio can then be computed from  $\phi$ ,  $\pi_{1+}$ , and  $\pi_{+1}$ . Note that when  $\phi=0$  (i.e., no correlation), we get  $\pi_{11} = \pi_{1+}\pi_{+1}$  (i.e.,  $X$  and  $Y$  are independent) and the odds ratio is  $OR=1$ . When  $\phi \neq 0$ , the term  $\phi \sqrt{\pi_{1+}\pi_{+1}(1 - \pi_{1+})(1 - \pi_{+1})}$  is then a measure of the departure from independence.

**Table 3.** Observed proportion of helmet use and safe passing manoeuvres from Walker (2007).

	No Helmet	Helmet	Total
Safe	0.491	0.462	0.953
Unsafe	0.021	0.026	0.047
Total	0.512	0.488	

doi:10.1371/journal.pone.0058777.t003

## Maximum $\phi$ and Modified Effect Sizes

When the marginal probabilities are fixed constants,  $\phi$  is an increasing linear function of  $\pi_{11}$ . Further,  $\pi_{11}$  is bounded by

$$\max(0, \pi_{+1} - \pi_{0+}) \leq \pi_{11} \leq \min(\pi_{1+}, \pi_{+1})$$

These bounds are due to all cell probabilities being non-negative and the relationship of  $\pi_{11}$  with the other cell probabilities given above. As a result,  $\phi$  is bounded as well and attains its maximum when  $\pi_{11} = \min(\pi_{1+}, \pi_{+1})$ . Using the upper bound of the above inequality, it can be shown that

$$\phi_{\max} = \max_{\pi_{11}} \phi = \sqrt{\frac{\pi_{1+}(1 - \pi_{+1})}{\pi_{+1}(1 - \pi_{1+})}}$$

where  $\pi_{1+} < \pi_{+1}$  to ensure  $\phi_{\max} \leq 1$ . It is clear from the formula for  $\phi_{\max}$  that the full range of correlation coefficients, i.e.,  $-1 \leq \phi \leq 1$ , is attainable only when the marginal probabilities are equal, i.e.,  $\pi_{1+} = \pi_{+1}$  or  $\pi_{1+} = (1 - \pi_{+1})$ . This has an intuitive appeal as perfect correlation for two binary variables is only possible when two cell probabilities are zero. For example, when all observations are in either the (0,1) or (1,0) cells,  $\hat{\phi} = -1$ . However, it would appear highly unlikely both marginal probabilities will be equal in practice. For example, in a 1:1 case-control study with mortality as the primary outcome, half of all patients would need to die for perfect correlation to be possible. On the other hand, if 10% of all patients die, the maximum correlation possible is  $\phi_{\max} = 1/3$  which is near a medium recommended effect size. So, in this situation, all estimates of  $\phi$ , computed from observed proportions, are bounded by

$$-1 < -1/3 \leq \hat{\phi} \leq 1/3 < 1$$

Importantly, odds ratios are not bounded with possible values of  $[1, \infty)$  as  $\phi$  varies on the interval  $[0, \phi_{\max}]$ . In fact, as  $\phi$  approaches  $\phi_{\max}$ , the *OR* increases without bound. Figure 1 demonstrates this relationship. Importantly, this indicates  $\phi$  has serious limitations as a measure of association and that these limitations are not applicable to the odds ratio.

## Effect Sizes Relative to $\phi_{\max}$

In many practical instances, the marginal probabilities are not equal, making the full range of values for  $\phi$  impossible with the potential of making Cohen's recommended effect sizes unusable for 2x2 tables. Although not equivalent to perfect correlation,  $\phi_{\max}$  can be interpreted as the maximum possible correlation given the marginal probabilities. In fact,  $\phi/\phi_{\max}$  has been proposed as an association measure with the interpretation as the proportion of observed correlation relative to the maximum attainable with fixed marginal probabilities [7], although the researcher is cautioned when the marginal probabilities diverge [6]. Note that  $\phi$  is not equivalent to Cohen's similarity/agreement measure  $\kappa$ . However,  $\kappa$  suffers from the same boundary problems as  $\phi$  and the two are equivalent when scaled to their maximum values, i.e.,  $\phi/\phi_{\max} = \kappa/\kappa_{\max}$ , making the two measures similar [6].

## Recommended effect sizes in terms of the odds ratio

As an alternative to Cohen's recommendations, increments of  $\phi_{\max}$  can be related to the odds ratio, say  $\alpha\phi_{\max}$ , where  $\alpha \in (0, 1)$ . Note that values of  $\alpha = 0.1, 0.3$  or  $0.5$  coincide with Cohen's usual recommendations when  $\phi_{\max} = 1$ . The relationship between  $\alpha\phi_{\max}$  and the odds ratio can be simplified by choosing marginal probabilities for commonly used participant allocations. As an example, Figures 2 and 3 demonstrate the relationship between  $\pi_{1+}$  and odds ratios for  $0.1\phi_{\max}$ ,  $0.3\phi_{\max}$  and  $0.5\phi_{\max}$  for 1:1 and 1:2 allocations respectively. Note that the minimal odds ratios, and therefore most conservative when used to compute sample size, occur when  $\pi_{1+}$  tends to 0. Although the odds ratio does not exist when  $\pi_{1+} = 0$ , the limit exists and is

$$OR_{\min} = \lim_{\pi_{1+} \rightarrow 0^+} OR = 1 + \frac{\alpha}{(1 - \alpha)\pi_{+1}}$$

Additionally, the maximal odds ratio, and therefore most anti-conservative, occurs when the marginal probabilities are equal, as expected. Below is the maximum attainable odds ratio for equal marginals  $\pi_{1+} = \pi_{+1} = \pi$  for increments  $\alpha$  of  $\phi_{\max}$ ,

$$OR_{\max} = \frac{(\pi - \alpha(1 - \pi))(1 - \pi(1 - \alpha))}{\pi(1 - \pi)(1 - \alpha)^2}$$

It is important to note that when  $0 < \pi_{+1} \leq 0.5$ , as is often true for case-control studies where cases are harder to identify or enrol than controls, the minimal odds ratio will be smallest for evenly allocated studies, i.e.,  $\pi_{+1} = 0.5$ . Further, it is generally recommended to use 1:1 allocation as it is the most statistically efficient ratio, i.e., maximum power for a fixed overall sample size. So, odds ratios of 1.22, 1.86 and 3.00 can be used as small, medium and large effect sizes without assumptions regarding marginal probabilities. Sample sizes computed using these odds ratios for 1:1 allocation are given in Table 2 for 80% power and 5% level of significance. A SAS macro that will compute sample sizes from given marginal probabilities for small, medium and large odds ratios has been provided as a supplementary file.

Interestingly, Haddock et al. [12] as a rule of thumb consider odds ratios greater than 3 large effect sizes, although there is no clear justification given. In a situation where an allocation ratio other than 1:1 is used, recommended odds ratios can be computed directly using the above formula. These results are also applicable for other values of  $\pi_{+1} > 0.5$  through its complement  $0 < \pi_{+0} \leq 0.5$ . This is equivalent to swapping the columns (or rows) and the researcher should be aware the recommended odds ratio effect sizes are now the reciprocals of those above, i.e., 0.82, 0.54 and 0.33 for small, medium and large respectively.

This approach can also be applied to the relative risk and risk difference. If  $X$  is taken as the grouping variable and  $Y$  as the outcome, the relative risk is  $\pi_{11}(1 - \pi_{+1})/(\pi_{10}\pi_{+1})$ . Simple substitution of  $\alpha\phi_{\max}$  and the marginal probabilities  $\pi_{1+}$  and  $\pi_{+1}$  results in a relative risk identical to  $OR_{\min}$  for  $\pi_{1+} > 0$ , i.e.,

$$RR = 1 + \frac{\alpha}{(1 - \alpha)\pi_{+1}}$$

Therefore, recommendations can also be derived for relative risk and are identical to those given for the odds ratio above. This

result is expected as the odds ratio converges to the relative risk as the incidence rate approaches 0.

Instead of comparing the risk between two groups as a ratio, it is sometimes useful to compare their differences [14]. Again taking  $X$  as the grouping variable and  $Y$  as the outcome, the risk difference can be written as

$$RD = \frac{\pi_{11}}{\pi_{+1}} - \frac{\pi_{10}}{\pi_{+0}} = \frac{\pi_{11} - \pi_{1+}\pi_{+1}}{\pi_{+1}(1 - \pi_{+1})}$$

where  $\pi_{1+} \leq \pi_{+1}$  to ensure  $\phi_{\max} \leq 1$  as above. It is clear from the numerator in this representation that  $RD$  is a measure of the departure from independence, i.e.,  $\pi_{11} = \pi_{1+}\pi_{+1}$ . Simple substitution of  $\alpha\phi_{\max}$  into  $RD$  yields

$$RD_{\alpha} = \alpha \frac{\pi_{1+}}{\pi_{+1}}$$

where the subscript  $\alpha$  is used to distinguish between risk difference formulae. This formula can be simplified somewhat for 1:1 allocations, i.e.,  $RD_{\alpha}(\pi_{+1}=0.5) = 2\alpha\pi_{1+}$ ; however, a general result independent of the marginal probabilities is clearly not possible in this instance as  $0 \leq \pi_{1+} \leq 0.5$  and therefore  $0 \leq RD_{\alpha}(\pi_{+1}=0.5) \leq \alpha$ .

Alternatively, the  $OR_{\min}$  formula can be solved for  $\alpha$  and compared to previously given odds ratio recommendations. In terms of  $OR_{\min}$  and  $\pi_{+1}$ , we get

$$\alpha = \frac{\pi_{+1}(OR_{\min} - 1)}{1 + \pi_{+1}(OR_{\min} - 1)}$$

When the allocation ratio is 1:1, this formula simplifies to  $\alpha(\pi_{+1}=0.5) = (OR_{\min} - 1)/(OR_{\min} + 1)$  which has a form identical to Yule's  $Q$  [15]. So, Ferguson's [10] odds ratio recommendations of 2.0, 3.0 and 4.0 therefore correspond to proportions of maximum correlation of  $\alpha = 0.33, 0.5$  and  $0.6$ . This suggests Ferguson's recommendations have the potential to be anti-conservative from a sample size viewpoint.

## Example

This paper was motivated by a reanalysis of passing distances for motor vehicles overtaking a bicyclist [16]. One of the primary results of this study was a significant association between helmet wearing and less overtaking distance, supporting a theory of risk perception for motor vehicle drivers directed towards bicyclists. Prior to collecting data, Walker [16] reported computing a sample size of  $n = 2259$  overtaking manoeuvres based on a  $2 \times 5$  fixed effects factorial ANOVA for a small effect size  $f = 0.1$ , 5% level of significance and 98% power. The factors for this study were helmet wearing (2 levels) and bicycle position relative to the kerb (5 levels). It has been noted, however, that passing distances are often recommended and sometimes legislated to one metre or more [17]. So, passing manoeuvres of at least a metre are considered safe and less than a metre unsafe, with the implication that large differences in passing distance are unimportant beyond one metre in terms of bicycle safety. When compared with helmet wearing, safe/unsafe passing distances can be analysed using a  $2 \times 2$  table. Since Walker's study was powered at an unusually high level with subsequent increased probability of a type I error, bootstrap standard errors were estimated for more reasonable values for power of 80%, 85% and 90%. Operationally defined small,

medium and large effect sizes were also used since a meaningful difference in overtaking distance is unknown.

The relevant observed data from Walker [16] is given in Table 3. The observed marginal proportions here are  $\pi_{+1} = 0.488$  for helmet wearing and  $\pi_{1+} = 0.047$  for unsafe passing manoeuvres. Using the marginal probabilities, the maximum attainable effect size is  $\phi_{\max} \approx 0.227$  and the estimated correlation is  $\hat{\phi} = 0.028$ . A consequence is the effect size for the association between helmet wearing and safe passing distance is, at best, much less than a small effect size by Cohen's index. The corresponding small, medium and large odds ratio effect sizes using increments of  $\alpha\phi_{\max}$  are 1.24, 1.94 and 3.21 for  $\alpha = 0.1, 0.3$  and  $0.5$ . Note that these values are not much greater than the minimal recommended odds ratios mentioned in the previous section, further suggesting the association between safe/unsafe passing distance and helmet wearing is, at best, a small effect size. In fact, the unadjusted odds ratio is  $OR = 1.3$  and non-significant by the chi-square test ( $p = 0.182$ ). Conversely, sample sizes for a future study can be computed from the observed probabilities using G\*Power for logistic regression with a single binomially distributed predictor for  $\alpha = 0.05$  and 80% power [18] resulting in 16237, 1409 and 383 observations for small, medium and large odds ratios. To put these sample size computations into perspective, a future study would need to extend the sampling period by a factor greater than seven to detect a significant association between helmet wearing and safe/unsafe overtaking distance given a small effect size and identical marginal probabilities.

## Discussion

We present a demonstration that many contingency table correlation measures are equivalent for the  $2 \times 2$  case and their use is limited due to constraints created by fixed marginal probabilities. The odds ratio, which is a function of these measures for fixed marginal probabilities, is not problematic, is regularly used in statistical analyses and has a direct application to logistic regression. Recommended odds ratios have been proposed from Cohen's small, medium and large effect sizes for  $\phi$  relative to the maximum attainable correlation  $\phi_{\max}$ . Further, minimal odds ratios can be computed with only knowledge of participant allocation.

The use of effect size recommendations should be avoided in situations in which clinical or practical differences are known. However, they can help the researcher balance between overly large or overly small sample size calculations when such information is unknown. In these situations, conservative estimates for odds ratio effect sizes can be derived from only the allocation ratio leading to a general result and, when a 1:1 allocation is chosen for optimal power, odds ratios of 1.22, 1.86 and 3.00 correspond to small, medium and large effect sizes.

## Supporting Information

**File S1 SAS Macro to compute sample sizes from marginal probabilities for small, medium and large odds ratios.**

(SAS)

## Acknowledgments

The authors would like to thank Warren May, David Warton and Jakub Stoklosa for their help in the preparation of this manuscript.

## Author Contributions

Analyzed the data: JO MLB. Wrote the paper: JO MLB.

## References

- Lewis J (1999) Statistical principles for clinical trials (ICH E9): an introductory note on an inter-national guideline. *Statistics in Medicine* 18: 1903–1942.
- Halpern S, Karlawish J, Berlin J (2002) The continuing unethical conduct of underpowered clinical trials. *JAMA: The Journal of the American Medical Association* 288: 358–362.
- Cohen J (1992) A power primer. *Psychological Bulletin* 112: 155–159.
- Lenth R (2001) Some practical guidelines for effective sample size determination. *The American Statistician* 55: 187–193.
- King M (2011) A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research* 11: 171–184.
- Davenport Jr E, El-Sanhurry N (1991) Phi/phi<sub>max</sub>: review and synthesis. *Educational and Psychological Measurement* 51: 821–828.
- Ferguson G (1941) The factorial interpretation of test difficulty. *Psychometrika* 6: 323–329.
- Guilford J (1965) The minimal phi coefficient and the maximal phi. *Educational and Psychological Measurement* 25: 3–8.
- Breaugh J (2003) Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management* 29: 79–97.
- Ferguson C (2009) An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice* 40: 532–538.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Haddock C, Rindskopf D, Shadish W (1998) Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods* 3: 339–353.
- Agresti A (2002) *Categorical Data Analysis*. New York: Wiley-interscience.
- Greenberg R, Daniels S, Flanders W, Eley J, Boring J (1996) *Medical Epidemiology*. Appleton & Lange.
- Liebetrau A (1983) *Measures of Association*, volume 32. Sage Publications, Incorporated.
- Walker I (2007) Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis & Prevention* 39: 417–425.
- Olivier J (2013) Bicycle helmet wearing is not associated with close overtaking: A re-analysis of Walker 2007. Submitted.
- Faul F, Erdfelder E, Buchner A, Lang A (2009) Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41: 1149–1160.