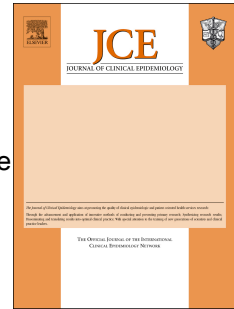


Journal Pre-proof

Questionable utility of the relative risk in clinical research: A call for change to practice

Suhail A. Doi, Luis Furuya-Kanamori, Chang Xu, Lifeng Lin, Tawanda Chivese, Lukman Thalib



PII: S0895-4356(20)31171-9

DOI: <https://doi.org/10.1016/j.jclinepi.2020.08.019>

Reference: JCE 10337

To appear in: *Journal of Clinical Epidemiology*

Received Date: 8 January 2020

Revised Date: 6 February 2020

Accepted Date: 28 August 2020

Please cite this article as: Doi SA, Furuya-Kanamori L, Xu C, Lin L, Chivese T, Thalib L, Questionable utility of the relative risk in clinical research: A call for change to practice, *Journal of Clinical Epidemiology* (2020), doi: <https://doi.org/10.1016/j.jclinepi.2020.08.019>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.

Controversy and debate series

Questionable utility of the relative risk in clinical research: A call for change to practice

Suhail A. Doi¹, Luis Furuya-Kanamori², Chang Xu³, Lifeng Lin⁴, Tawanda Chivese¹, and Lukman Thalib⁵

¹Department of Population Medicine, College of Medicine, QU Health, Qatar University, Doha, Qatar

²Research School of Population Health, College of Health and Medicine, Australian National University, Canberra, Australia

³Chinese Evidence Based Medicine Center, SiChuan University, Chengdu, China

⁴Department of Statistics, Florida State University, Tallahassee, FL, USA

⁵Department of Public Health, College of Health Sciences, QU Health, Qatar University, Doha, Qatar

Correspondence to

Suhail A. R. Doi

Department of Population Medicine

College of Medicine

Al Jamia Street, P.O. Box 2713, Doha, Qatar

Tel: +974 66001271

Email sardoj@gmx.net

Abstract

Objective: In clinical trials, the relative risk or risk ratio (RR) is a mainstay of reporting of the effect magnitude for an intervention. The RR is the ratio of the probability of an outcome in an intervention group to its probability in a control group. Thus the RR provides a measure of change in the likelihood of an event linked to a given intervention. This measure has been widely used because it is today considered a measure with 'portability' across varying outcome prevalence, especially when the outcome is rare. It turns out however that there is a much more important problem with this ratio, and this paper aims to demonstrate this problem.

Methods: We used mathematical derivation to determine if the RR is a measure of effect magnitude alone (i.e. a larger absolute value always indicating a stronger effect) or not. We also used the same derivation to determine its relationship to prevalence of an outcome. We confirm the derivation results with a follow-up analysis of 140,620 trials scraped from the Cochrane.

Results: We demonstrate that the RR varies for reasons other than magnitude of the effect because it is a ratio of two posterior probabilities, both of which are dependent on baseline prevalence of an outcome. Additionally we demonstrate that the RR shifts towards its null value with increasing outcome prevalence. The shift towards the null happens regardless of the strength of the association between intervention and outcome. The odds ratio (OR), the other commonly used ratio, measures solely the effect magnitude and has no relationship to prevalence of an outcome in a study nor does it overestimate the RR as is commonly thought.

Conclusions: The results demonstrate the need to a) end the primary use of the RR in clinical trials and meta-analyses as its direct interpretation is not meaningful; b) replace the RR by the OR; and c) only use the post-intervention risk recalculated from the OR for any expected level of baseline risk in absolute terms for purposes of interpretation such as the number needed to treat. These results will have far reaching implications such as reducing misleading results from clinical trials and meta-analyses and ushering in a new era in the reporting of such trials or meta-analyses in practice.

Keywords: binary effect measure; relative risk; odds ratio; risk difference; posterior probability

What is new?*Key findings*

- The RR does not reflect effect magnitude alone but rather is a ratio of two conditional probabilities that vary with outcome prevalence
- The OR is a true effect magnitude measure and represents the fold increase in odds of outcome from an unexposed state to an exposed state
- The risk difference and numbers needed to treat can easily be computed from the OR

What this study adds to what was known?

- The reason for the asymmetry of the RR is now clarified
- The reason why the RR and OR differ and why has now been clarified
- The utility of the RR as an effect magnitude measure has been questioned given that it is the ratio of two posterior probabilities, both of which are highly dependent on the prevalence of the outcome in a study
- The status of the OR as an effect measure has been clarified.

*What is the implication and what should change**now?*

- The RR should no longer be used in clinical trials, cohort studies and meta-analyses
- The OR is the effect measure of choice in analytic epidemiological studies
- The risk difference and numbers needed to treat should be derived from the OR

1. Introduction

In clinical trials, the relative risk or risk ratio (RR) is a mainstay of reporting of the effect magnitude for an intervention¹. The RR is the ratio of the probability of an outcome in an intervention group to the probability of an outcome in a control group, and thus the RR provides an increase or decrease in the likelihood of an event based on some intervention. Despite some concerns²⁻⁴, the RR has been widely used because, it is considered a measure with 'portability' across varying outcome prevalence, especially when the outcome is rare³. The RR is probably the oldest and most used of a variety of measures of association which have been developed⁵ to express in quantitative terms the relationship between an exposure (e.g. risk factor or intervention) and a binary outcome, but its first use in clinical research is more recent and attributed to Cornfield⁶⁻⁸.

The RR has also been approximated by the odds ratio (OR), an approximation that has been described much earlier in clinical research^{9,10}. The use of the OR to estimate the RR has been thought to be appropriate for studies of rare outcomes, however the RR and OR are not numerically equivalent when the outcome is common¹¹ and in the latter situation the RR may not even have a monotone relationship with the OR¹². It has also been thought that the OR overestimates the RR and thus may inappropriately affect clinical decision-making or policy development leading to unintentional errors in the analysis of treatments¹³.

By and large, the RR remains more popular today than the OR especially when reporting clinical trials and meta-analyses of clinical trial data. This is, in part, is because of the notion that researchers should preferably choose more interpretable measures of association based on a ratio or difference in risks to minimize confusion and misrepresentation of research^{2,14-17}. However, what has been ignored is the fact that the RR is asymmetric¹⁸ and this has led to some uncertainty over its constancy across

baseline risks^{2,3} and over interpretation and applicability of its results.¹⁹ Today, the debate continues over the merits of the OR versus the RR and their interpretation.^{12,20}

In this paper, we will show that the RR is not a measure of the magnitude of the intervention-outcome association alone because it has a stronger relationship with prevalence and therefore is not generalisable beyond the baseline risk of the population in which it is computed. More importantly, we will also demonstrate that the ORs accurately reflect the increase or decrease in odds attributed to an exposure/intervention over baseline odds in a control group and thus accurately estimate the effect magnitude and there is no overestimation of effect magnitude by the OR. We should point out that we will use outcome-prevalence and cumulative incidence (in the whole study) synonymously as the outcome prevalence is equal to the cumulative incidence proportion when we define (as we must) the outcome-prevalence at the end of the follow-up period.

2. Derivation from first principles

Bayes described the relationship between the posterior probability of an outcome conditional on an exposure based on the unconditional prior probability and an evidence factor. This can be restated in terms of the posterior odds of an outcome (D) conditional on an exposure (E) based on the population level prior odds and an evidence factor (likelihood ratio)^{21,22}. The latter can be given by the following expression:

$$\frac{P(D|E)}{P(\bar{D}|E)} = \frac{P(E|D)}{P(E|\bar{D})} \times \frac{P(D)}{P(\bar{D})} \quad [1]$$

where \bar{E} denotes non-exposure and \bar{D} denotes the absence of the outcome.

The evidence factor or likelihood ratio is a ratio of two exposure probabilities: the outcome group exposure probability (ϕ_1) and the no-outcome group exposure probability (ϕ_0) respectively. The ratio ϕ_1/ϕ_0 is called the exposure likelihood ratio but also equals the exposure-conditional odds ratio (OR_E) and represents the ratio of the exposure-conditional odds of outcome to the unconditional odds of outcome in the general population and is given by:

$$OR_E = \frac{P(D|E)}{P(\bar{D}|E)} \bigg/ \frac{P(D)}{P(\bar{D})} = \frac{\phi_1}{\phi_0} \quad [2]$$

Similarly the no-exposure likelihood ratio represents the conditional odds ratio ($OR_{\bar{E}}$) in the non-exposed population. This ratio is thus the non-exposure conditional odds of an outcome to the unconditional odds of the same outcome in the general population and is given by:

$$OR_{\bar{E}} = \frac{1-\phi_1}{1-\phi_0} \quad [3]$$

These two likelihood based OR's are different from that commonly computed in a case-control study, where the ratio of odds is that of the retrospective outcome conditional odds of exposure to the retrospective no-outcome conditional odds of exposure and this would be exactly the same (mathematically) as the retrospective exposure conditional odds of outcome to the retrospective no-exposure conditional odds of outcome. This classical odds ratio (OR_C) is therefore given by:

$$OR_C = \frac{\phi_1}{\phi_0} \bigg/ \frac{(1-\phi_1)}{(1-\phi_0)} = \frac{OR_E}{OR_{\bar{E}}} = \frac{\phi_1(1-\phi_0)}{\phi_0(1-\phi_1)} \quad [4]$$

and is equivalent to the diagnostic OR in diagnostic studies.²³ Both the OR_C and the OR_E are independent of prevalence but represent different OR's. The OR_C represents the ratio of two outcome odds (exposure-conditional vs no-exposure conditional) while the OR_E denotes the ratio of two

outcome odds (exposure-conditional vs unconditional) and so they only differ in the denominator odds.

The OR_C is therefore the ratio of the two likelihoods (or odds ratios) and represents the fold increase in odds from the no-exposure state to the exposure state. This is easy to see from the following expression:

$$OR_C = \frac{LR_E}{LR_{\bar{E}}} = \frac{OR_E}{OR_{\bar{E}}}$$

In prospective cohort studies we can compute prospective probabilities of outcome (known as risks) in the exposure group (r_1) and in the non-exposure group (r_0). The ratio of these risks is the RR popularly used in epidemiology today. However, the retrospective likelihood ratio and the prospective risk ratio are not equal and thus

$$\frac{r_1}{r_0} = \frac{P(D|E)}{P(D|\bar{E})} \neq \frac{\phi_1}{\phi_0} = \frac{P(E|D)}{P(E|\bar{D})}$$

Cornfield⁶ has shown that the retrospective proportions can be used to compute risks if prevalence of disease (δ) in the general population is taken into consideration as follows:

$$r_1 = P(D|E) = \frac{\phi_1 \delta}{\phi_1 \delta + \phi_0 (1 - \delta)} = \frac{\delta OR_E}{(\delta [OR_E - 1] + 1)} \quad [5]$$

$$r_0 = P(D|\bar{E}) = \frac{(1 - \phi_1) \delta}{(1 - \phi_0) - \delta(\phi_1 - \phi_0)} = \frac{\delta OR_{\bar{E}}}{(\delta [OR_{\bar{E}} - 1] + 1)} \quad [6]$$

It is quite easy to compute prospective conditional outcome risks at any value of the population level prevalence using the OR_E or the $OR_{\bar{E}}$ as depicted above and it follows that the ratio of expressions [5] and [6] is the RR given by

$$RR = \frac{OR_E(\delta OR_E - \delta + 1)}{OR_E(\delta OR_E - \delta + 1)} \quad [7]$$

It is clear therefore from expression [7] that the RR is *dependent* on the prevalence of an outcome and therefore is not ‘portable’ across prevalence levels and therefore only partially reflects the exposure–outcome association. In addition, the RR only approximates the OR_C when $RR_{\delta \downarrow 0}$ or $RR_{\delta(\phi_1 - \phi_0) \downarrow 0}$. In all other cases the numerical value of the OR_C will always differ from the RR. This can be easily demonstrated by computing the RR and OR at varying prevalence levels. As shown in Figure 1, when the OR is held constant and the prevalence of outcome increases, the RR shifts progressively towards the null. If we accept the OR as an effect magnitude measure that quantifies the fold increase in odds of an outcome from the non-exposure state to the exposure state, then as depicted in Figure 1, the RR does not reflect the exposure–outcome association if outcome prevalence is ignored.

It is quite easy to compute the prospective risk in those exposed at any value of the no-exposure risk (r_0) since we would simply need to replace OR_E with OR_C and δ with r_0 in expression [5] so that

$$r_1 = \frac{r_0 OR_C}{r_0 [OR_C - 1] + 1} \quad [8]$$

and dividing expression [8] by r_0 again gives the risk ratio as follows:

$$RR = \frac{OR_C}{r_0 [OR_C - 1] + 1} \quad [9]$$

This is exactly the expression suggested by Zhang and Yu²⁴ in 1998, though the use of this expression to generate confidence limits for the RR based on the limits of the OR by substitution has been criticized on the grounds that the proposed confidence interval for the RR would be too narrow presumably because of its failure to account for variability in the baseline risk.²⁵ This may, however, be a reflection of the RR

dependence on prevalence²⁶ and this is no longer an issue given that this derivation advocates using the OR and its limits to determine posterior probabilities for a fixed baseline risk of interest from which clinically useful limits for the absolute effect sizes^{27,28} are derived. These limits are necessarily a function only of the estimated variability in the OR.

Subtracting expression [8] from r_0 gives the risk difference^{27,28} (RD) for any baseline risk given by

$$RD = \frac{r_0(r_0 - 1)(OR_C - 1)}{r_0 OR_C - r_0 + 1} \quad [10]$$

and finally the reciprocal of expression [10] gives the number needed to treat^{27,28} (NNT) given by

$$NNT = \frac{r_0 OR_C - r_0 + 1}{r_0(r_0 - 1)(OR_C - 1)} \quad [11]$$

3. Implications of the derivation

Box 1. *This box depicts the main relationships that emerge from expressions [9] to [11] where r_0 is baseline risk and OR_C is the classical odds ratio. The confidence limits are determined from those of the classical odds ratio and the key expressions of interest are:*

$$RD = \frac{r_0(r_0 - 1)(OR_C - 1)}{r_0 OR_C - r_0 + 1}$$

$$NNT = \frac{r_0 OR_C - r_0 + 1}{r_0(r_0 - 1)(OR_C - 1)}$$

Just as the presence of a positive test result has a diagnostic likelihood associated with it, the presence of an exposure or an intervention also has an outcome likelihood associated with it. With diagnostic tests this outcome is the actual diagnosis while with exposures and interventions the outcome may be disease or disease states. Taking Bayes' theorem into consideration it becomes clear (expression [1]) that the ratio of both likelihoods (likelihood with and without the exposure) equals the ratio of two conditional

odds (expression [4]): the exposure conditional odds of outcome to no-exposure conditional odds of the outcome). From any baseline odds of non-exposure, this relationship can be used to generate the odds with exposure as well as the conditional risks and their ratio, RD and the NNT or harm (NNH). Box 1 depicts the important relationships that emerge based on the derivation above. The expression for the RR is not included in Box 1 and the reason this is left out is that exposure-outcome association magnitude cannot be ascertained directly with the RR because it actually reflects the ratio of outcome probability in the presence of exposure (akin to the positive predictive value; PPV) to outcome probability without exposure (akin to 1-negative predictive value; 1-NPV) and thus the $RR = \frac{PPV}{(1-NPV)^{29}}$, and both the PPV and 1-NPV increase with outcome prevalence in a study. Thus the RR represents a ratio of two posterior probabilities conditional on exposure and non-exposure respectively while the OR is a true measure of effect magnitude because it is a measure of the fold-increase in odds of the outcome from the no-exposure/no-intervention state to the exposure/intervention state and is constant for any given exposure/intervention. Table 1 & Figure 2 depict this issue using hypothetical data regarding an exposure (smoking 10 or more cigarettes per day). The hypothetical data are based on

the premise that 77% of lung cancer subjects smoked that much while only 58% of a group of non-lung cancer subjects smoked that much. As baseline population prevalence of cancer varies from 10% – 90% and the various likelihoods (OR_E and $OR_{\bar{E}}$) are held constant, the two risks and their ratio (RR) vary. The reason why this happens is clearly depicted in Table 1 and Figure 2 where the NNH and RR are highly dependent on prevalence of the outcome as mathematically demonstrated above. The OR does not have this issue because it also is a likelihood ratio and therefore is invariant with the prevalence of disease/outcome. The fact that it is a likelihood ratio also explains the observation by Hoppe et al³⁰ that the OR can also be interpreted in terms of a conditional RR given discordant pairs from unmatched data collected pair-wise. On the contrary, expressing both posterior probabilities as a ratio (i.e. the conventional RR) is meaningless. The absolute change from non-exposure to exposure is what is meaningful and can be expressed as a NNH/NNT and can be computed from the OR for any specific baseline non-exposure risk through the relationships depicted in Box 1.

4. Analysis of 140,620 trials scraped from Cochrane

Data from individual studies of meta-analyses published in the Cochrane Database of Systematic Reviews from 2003 (issue 1) to 2018 (issue 5) were extracted using the R package “RCurl”. The R code has been provided in a previous study which used this dataset to compare tests for publication bias.³¹ A total of 181,278 trials used in 18562 meta-analyses were scraped from Cochrane. We dropped meta-analyses containing trials with zero events in both arms. We also dropped studies with baseline or intervention risks exactly equal to zero or one and dropped meta-analyses with less than 5 studies (in this sequence). There were 140,620 trials left for analysis across 14,960 meta-analyses. The RR, OR, outcome prevalence and baseline risk in the control group were computed for each of these trials (regardless of meta-analysis they belonged to). The ORs were found to have no association with

prevalence and all the expressions derived above (and in box 1) were confirmed to be numerically equivalent to the directly computed RR, RD or NNT/NNH. The relationship between the RR and prevalence described mathematically above was confirmed and is depicted in Figure 3. Since the RR is a ratio of two conditional probabilities, it loses meaning given that the two conditional probabilities increase with prevalence and when the association between exposure and outcome is fixed (by generating percentiles of the logOR and analyzing by percentile) the RR is just a linear function of prevalence and nothing more (figure 3, panel C). A similar relationship is seen when prevalence is fixed by percentile and depicted in supplementary figure S1.

Given this finding, pooled effects from meta-analyses using the RR effect measure will tend to be increasingly misleading as prevalence of the outcome increases or varies across studies (in meta-analyses) and will also be expected to have decreasing heterogeneity with increasing prevalence. This was demonstrated by computing these meta-analyses using the fixed effect model in Stata and results show much less heterogeneity (between studies variance, tau²) when pooling using the RR compared to when pooling using the OR and this was expected given the range limitation of the RR as prevalence increased (Figure 4). In keeping with our observation, the use of the RR in meta-analysis has recently been flagged as a source of bias in RR based meta-analyses³².

5. Examples from the literature

In a meta-analysis of prostate cancer–specific mortality associated with androgen deprivation therapy (ADT) among patients with prostate cancer³³, the pooled OR=0.64 and RR=0.75 reported were dissimilar (Figure 5) because mortality varied across studies (2% - 59%). The implications are also different for when baseline mortality risk varies for those without ADT. For example, given a hypothetical baseline mortality risk of 10%, this would lead to different values for the post intervention risk difference and the

NNT would then be 58 (OR) vs 80 (RR) using the OR or the RR respectively. Of interest here is that Fagan's nomogram can also be used to go from control to post-intervention probability using the OR as the likelihood ratio.

In another meta-analysis of stroke or systemic embolism after high dose non-vitamin K oral anticoagulants (NOAC) vs vitamin K antagonists (VKA)³⁴, the authors report the RR for stroke or systemic embolism to be 0.79 (Figure 6). The meta-analysis using the OR returned numerically similar results because outcome incidence was similar across studies (4% - 6%). However, the interpretation of both measures differ. Using the OR, assuming varying baseline incidence of stroke or systemic embolism in the VKA group, we get varying results for the RD across use of the RR or the OR (Figure 6 panel C). The RR, therefore, gives inaccurate risk information regardless of whether it has or has not got a numerical similarity to the OR.

6. Conclusion

In this paper we discuss only two relative measures (OR & RR) and attempt to explain why one of them seems problematic and thus by default this leaves us with only one moving forwards. While many effect measures exist, the differences in objectives for the different measures are not raised in this paper as we focus solely on the debate between OR & RR as we show that all information previously derived from the RR can just as easily be derived from the OR. Of greater concern is that the RR may not be a valid measure of effect. The basic concepts underpinning this study are all well established but what this paper adds is that what we think is being measured is not actually what is being measured and this lends itself to conclusions that are not true.

A main thesis of this paper is that relative risk depends on prevalence, more so than on the strength of the exposure-outcome association that it is supposed to reflect. Thus, if the association is unchanged and we modify prevalence then the relative risk changes. If the strength of association is measured independently of the RR and kept constant then we can show that the RR is a linear function of prevalence. We can no longer accept the commonly argued for view that the relative risk is easier to understand. Once we realize that the RR depends more on prevalence than the exposure-outcome association, its interpretation becomes much more difficult to comprehend than the odds ratio.

It is well known that, for common events, large values of the risk ratio are impossible and this should have rung the alarm bells much earlier regarding whether the RR is more a measure of prevalence than a measure of effect. However this was not the main focus of the derivation outlined previously and the latter was aimed at demonstrating why the OR is a true measure of effect against which the RR can be compared. The derivations relating to the OR in this paper relate primarily to the odds ratio and then are linked to the relative risk through derivation of risk. The derivations do not require the assumption that the probability of exposure given outcome or the probability of exposure given no outcome, are constants, but rather that they are independent of outcome-prevalence. This is logical since they are derived from the outcome or no outcome groups separately. While test principles are applied in the derivation, a test being a consequence of the outcome or an exposure a cause of the outcome is not really important in this context because the probabilities are assessed at one point in time – in non-diagnostic studies this would be at the end of follow-up.

Clinical research has a substantial need for absolute measures and either the OR or RR can be used to compute these as we show in this paper. However the choice of binary effect measure in epidemiologic studies precedes the computation of such absolute effects. Creating a measure of impact is usually done by combining a relative measure with baseline risk information. However, this can only

be done when the relative measure has portability across different risk groups. This paper demonstrates this lack of portability with the RR and so both the direct interpretation and the translation to absolute measures of impact are flawed.

While it is a common perception that consistency of an effect measure in meta-analysis means that it is the effect measure of choice. This again is a misconception because If the RR does not measure effect magnitude and simply reflects prevalence of the outcome in the study, consistency is no longer meaningful. Consistency increases with the RR as prevalence increases simply because the RRs become more equal (as well as lose meaning as a measure of effect magnitude). It may be okay to meta-analyse studies with similar prevalence (though if the prevalence is high, interpretation is still questionable) but if this differs across studies (see the example) it makes little sense to meta-analyse such studies as the RRs cannot be interpreted as similar even if they are fortuitously consistent.

We reiterate that the RR does not fully reflect the magnitude of the effect in clinical trials. Rather, it reflects more of the baseline prevalence of an outcome, shifting towards 1 as prevalence increases. It is also a ratio of two conditional probabilities making it's interpretation questionable. On the other hand, the OR is a likelihood ratio whose magnitude reflects the fold increase in odds from that in the baseline group to that in the intervention group and reflects the magnitude of the intervention effect independent of prevalence. Interpretation of the OR is similar to that of the likelihood ratio in diagnostic studies. The OR is expected to be specific to any intervention and thus should be the binary effect measure in randomized trials and meta-analyses of such trials and the RD and NNT should preferably be derived from the OR and not the RR. To this end, we provide a conversion excel sheet in supplementary material (S2) for use by clinicians to move from baseline risk to post-intervention risk, just like is done for diagnostic tests with the likelihood ratio replaced by the OR from a clinical trial. What this paper raises about the RR dependence on prevalence should also not be conflated with the well known fact in

epidemiology that the OR approximates the RR when prevalence is small. The latter is about numerical equivalence thus rendering the OR interpretable as a RR. This has never been previously viewed as a problem inherent to the RR, which is what this study aims to explain. Therefore, while the RR and OR coincide numerically over a narrow range below baseline risks of 20-30%, this is no justification to continue to use this ratio. The risk relativism as it is currently viewed is an illusion.

Finally, we should point out that the basics used in this paper reflect well known material. What we are arguing for is a major change in the status quo regarding interpretation and usage based on these very same basic principles. While we appreciate the fact that some debates are put out there to stimulate thought and advance understanding, we are also seriously concerned about this issue because we believe the evidence we provide in support of change is compelling.

Funding statement

This work was made possible by Program Grant #NPRP10-0129-170274 to SAD, LT and LFK from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work, and are solely the responsibility of the authors. All authors had full access to all the data in the study and the corresponding author had final responsibility for the decision to submit for publication. LFK is also supported by an Australian National Health and Medical Research Council Fellowship (APP1158469)

Reference List

1. Pocock SJ, McMurray JJ, Collier TJ. Making sense of statistics in clinical trial reports: part 1 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol* 2015; 66:(22)2536-2549. 0735-1097.
2. Walter SD. Choice of effect measure for epidemiological data. *J Clin Epidemiol* 2000; 53:931-9.
3. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987; 125:(5)761-768. 1476-6256.
4. Rothwell PM. Can overall results of clinical trials be applied to all patients? *The Lancet* 1995; 345:(8965)1616-1619. 0140-6736.
5. Goodman LA, Kruskal WH. Measures of association for cross classifications. II: Further discussion and references. *J Am Stat Assoc* 1959; 54:(285)123-163. 0162-1459.
6. Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst* 1951; 11:1269-75.
7. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959; 22:(1)173-203. 1460-2105.
8. Greenland S. Cornfield, risk relativism, and research synthesis. *Stat Med* 2012; 31:2773-7.
9. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet* 1955; 19:(4)251-253. 0003-4800.
10. Haldane JBS. The estimation and significance of the logarithm of a ratio of frequencies. *Ann Hum Genet* 1956; 20:(4)309-311. 0003-4800.
11. Persoskie A, Ferrer RA. A Most Odd Ratio:: Interpreting and Describing Odds Ratios. *Am J Prev Med* 2017; 52:224-228.
12. Feng C, Wang B, Wang H. The relations among three popular indices of risks. *Stat Med* 2019; 1097-0258.
13. Balasubramanian H, Ananthan A, Rao S, Patole S. Odds ratio vs risk ratio in randomized controlled trials. *Postgrad Med* 2015; 127:359-67.
14. Tajeu GS, Sen B, Allison DB, Menachemi N. Misuse of odds ratios in obesity literature: an empirical analysis of published studies. *Obesity (Silver Spring)* 2012; 20:1726-31.
15. Lee J. Odds ratio or relative risk for cross-sectional data? *Int J Epidemiol* 1994; 23:201-3.
16. Bracken MB, Sinclair JC. When can odds ratios mislead? Avoidable systematic error in estimating treatment effects must not be tolerated. *BMJ* 1998; 317:1156; author reply 1156-7
17. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol* 1994; 47:881-9.
18. Eckermann S, Coory M, Willan AR. Indirect comparison: relative risk fallacies and odds solution. *J Clin Epidemiol* 2009; 62:1031-6.

19. Sonis J. Odds Ratios vs Risk Ratios. *JAMA* 2018; 320:2041
20. Furuya-Kanamori L, Doi SA. The outcome with higher baseline risk should be selected for relative risk in clinical studies: a proposal for change to practice. *J Clin Epidemiol* 2014; 67:364-7.
21. Johnson KM. Using Bayes rule in diagnostic testing: a graphical explanation. *Diagnosis* 2017; 4(3):159-167. 2194-802X.
22. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005; 365:1500-5.
23. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56(11):1129-1135. 0895-4356.
24. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998; 280:1690-1.
25. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003; 157:940-3.
26. Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *J Clin Epidemiol* 2007; 60(9):874-882. 0895-4356.
27. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995; 310:452-4.
28. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318:1728-33.
29. Murphy JR. The relationship of relative risk and positive predictive value in 2 X 2 tables. *Am J Epidemiol* 1983; 117:86-9.
30. Hoppe FM, Hoppe DJ, Walter SD. Odds ratios deconstructed: A new way to understand and explain odds ratios as conditional risk ratios. *J Clin Epidemiol* 2017; 82:87-93.
31. Furuya-Kanamori L, Xu C, Lin L, Doan T, Chu H, Thalib L, et al. P value-driven methods were underpowered to detect publication bias: analysis of Cochrane review meta-analyses. *J Clin Epidemiol* 2019; 118:86-92.
32. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Pitfalls of using the risk ratio in meta-analysis. *Res Synth Methods* 2019; 1759-2879.
33. Nguyen PL, Je Y, Schutz FA, Hoffman KE, Hu JC, Parekh A, et al. Association of androgen deprivation therapy with cardiovascular death in patients with prostate cancer: a meta-analysis of randomized trials. *JAMA* 2011; 306(21):2359-2366.
34. Ha JT, Neuen BL, Cheng LP, Jun M, Toyama T, Gallagher MP, et al. Benefits and Harms of Oral Anticoagulant Therapy in Chronic Kidney Disease. *Ann Intern Med* 2019;

Table 1 Tabulation of the data depicted in figure 2

δ	OR_E	$OR_{\bar{E}}$	OR_C	RR	r_1	r_0
0.1	1.33	0.55	2.42	2.24	0.1285	0.0574
0.2	1.33	0.55	2.42	2.07	0.2492	0.1204
0.3	1.33	0.55	2.42	1.91	0.3626	0.1901
0.4	1.33	0.55	2.42	1.76	0.4695	0.2674
0.5	1.33	0.55	2.42	1.61	0.5704	0.3538
0.6	1.33	0.55	2.42	1.48	0.6657	0.4510
0.7	1.33	0.55	2.42	1.35	0.7560	0.5610
0.8	1.33	0.55	2.42	1.23	0.8415	0.6866
0.9	1.33	0.55	2.42	1.11	0.9228	0.8313

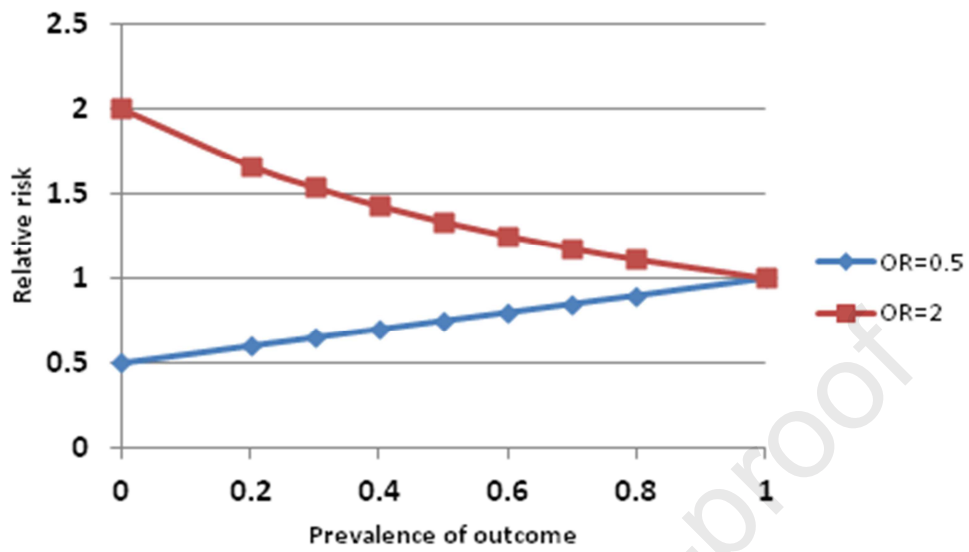


Figure 1. Relationship of the relative risk to outcome prevalence when the odds ratio (OR) is held constant

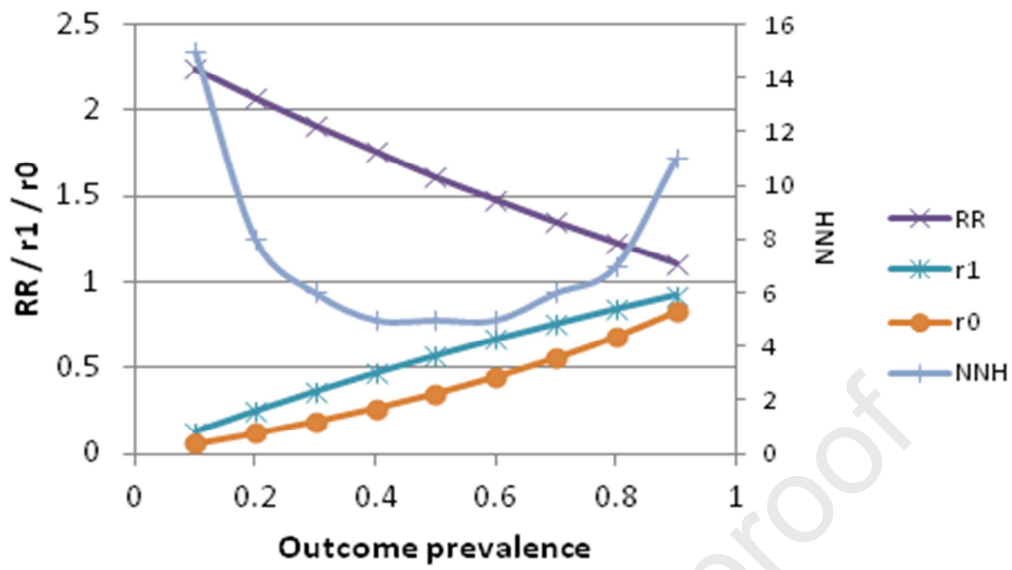


Figure 2. Relationship of the relative risk (RR), exposure risk (r1), no-exposure risk (r0), and number needed to harm (NNH) to outcome prevalence for the example in the text

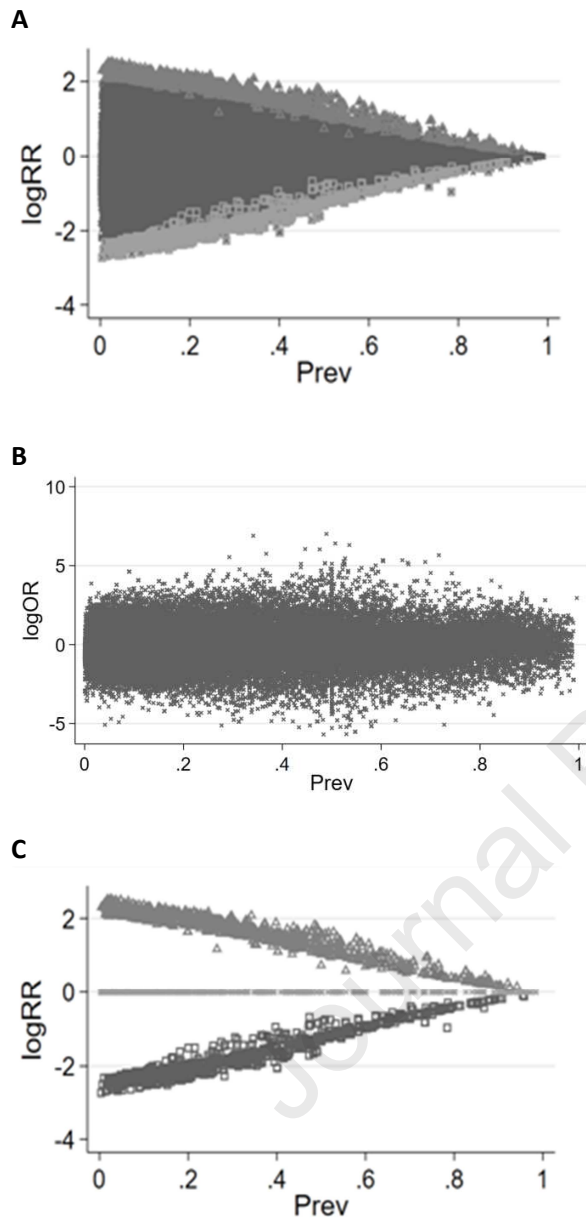


Figure 3. Relationship between the logRR (panel A), logOR (panel B) and outcome prevalence across all 140,620 trials. The trials with fixed ORs are depicted in panel C with a hollow square (percentile 2; OR range 0.06 to 0.1), with a X (percentile 55; OR of 1) and with a hollow triangle (percentile 99; OR range 8.3 to 13.2).

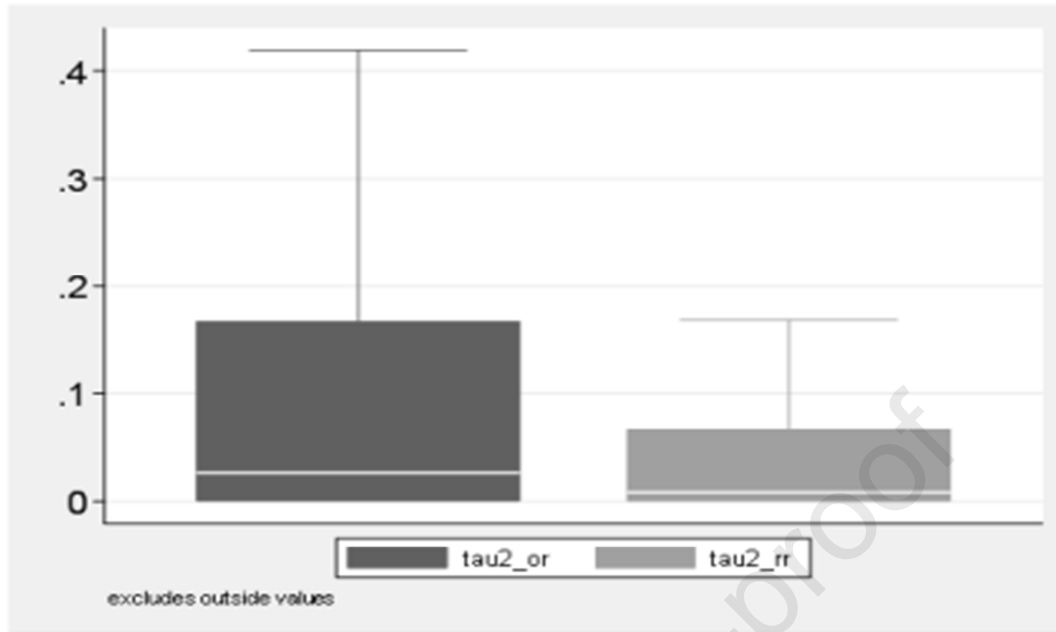
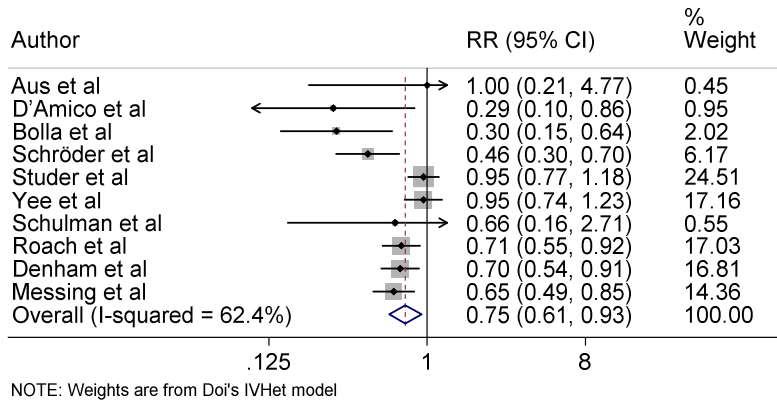
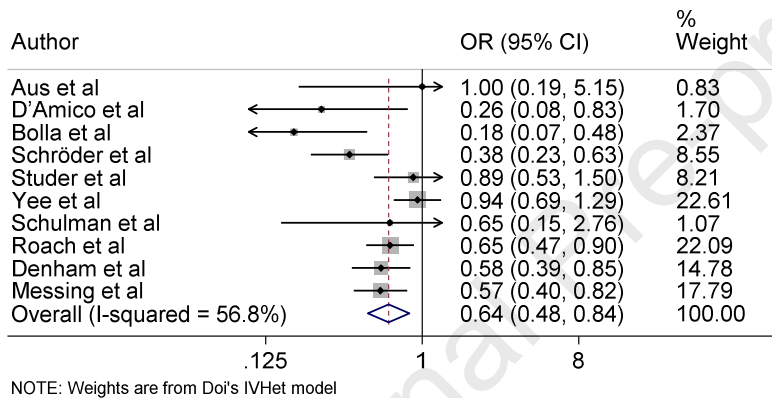


Figure 4. The RR based meta-analyses (right box plot) were less heterogeneous than the OR based ones (left box plot)

A



B



C

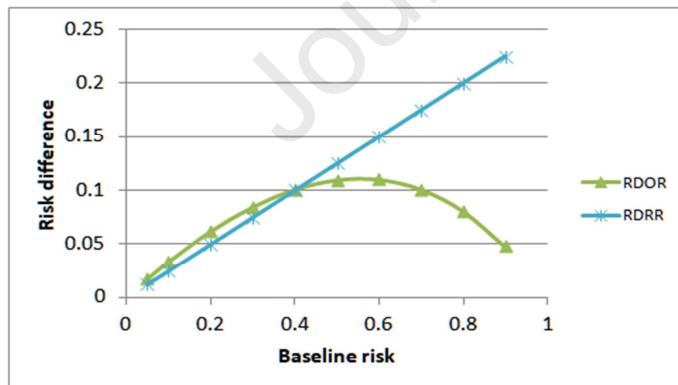
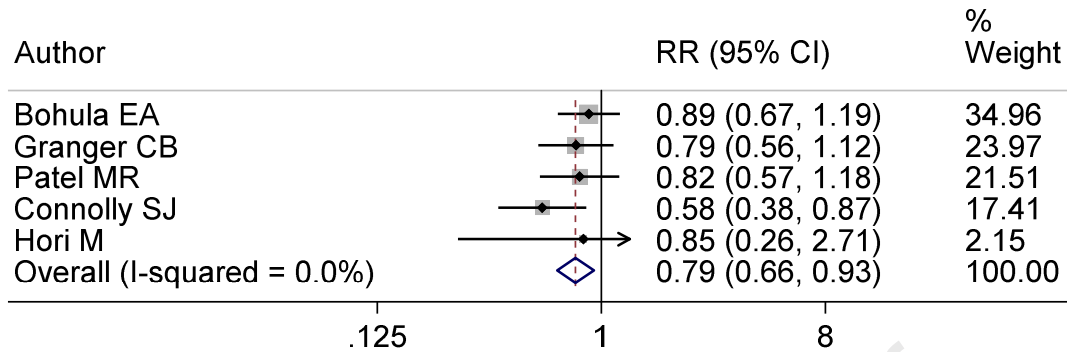


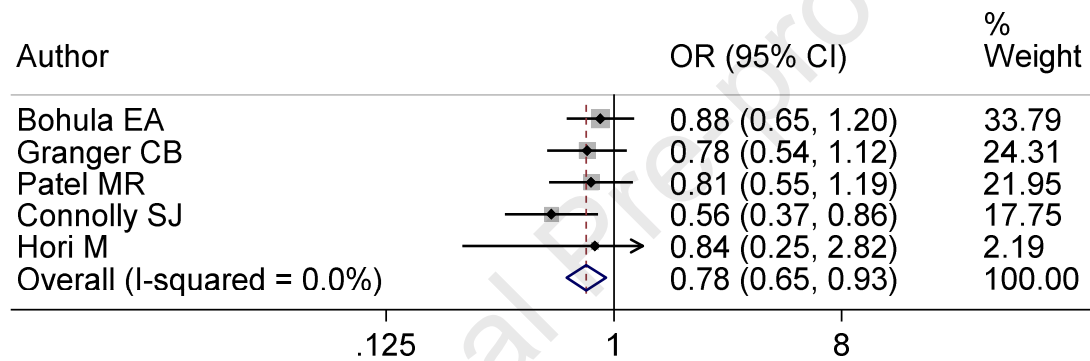
Figure 5. Meta-analysis of prostate cancer-specific mortality associated with ADT among patients with prostate cancer. The RR (panel A) and OR (panel B) differ because mortality across studies ranges between 2% – 59%. Panel C denotes the divergence in risk difference when computed from OR or RR.

A



NOTE: Weights are from Doi's IVHet model

B



NOTE: Weights are from Doi's IVHet model

C

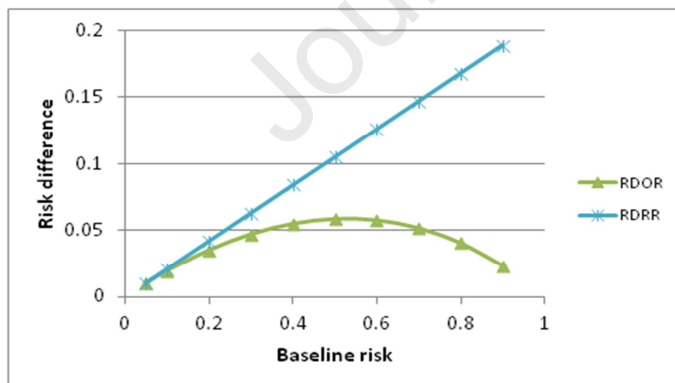


Figure 6. Meta-analysis of stroke or systemic embolism after high dose NOAC vs VKA. The RR (panel A) and OR (panel B) are similar because outcome incidence across studies ranges between 4% – 6%. Panel C denotes the divergence in risk difference when computed from OR or RR.

Supplementary material

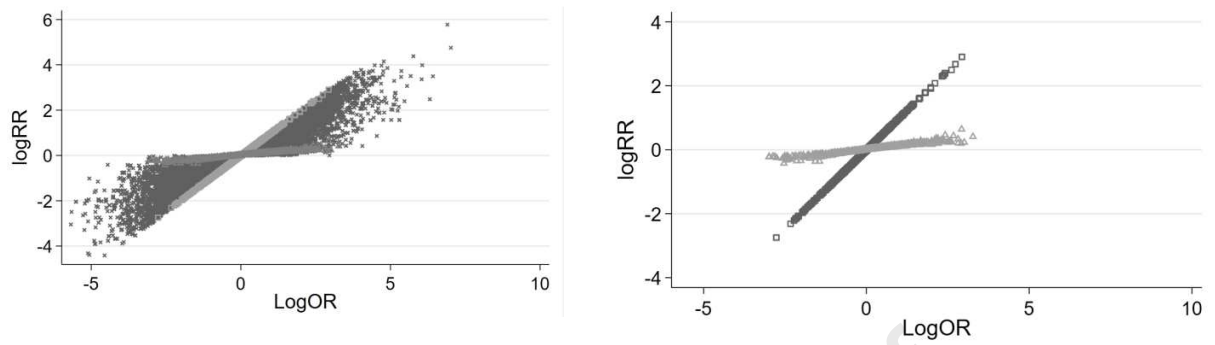


Figure S1. Relationship between LogRR and LogOR for all 140,620 trials (left panel) and for percentiles 2 and 99 of prevalence (right panel).

What is new?*Key findings*

- The RR does not reflect effect magnitude alone but rather is a ratio of two conditional probabilities that vary with outcome prevalence
- The OR is a true effect magnitude measure and represents the fold increase in odds of outcome from an unexposed state to an exposed state
- The risk difference and numbers needed to treat can easily be computed from the OR

What this study adds to what was known?

- The reason for the asymmetry of the RR is now clarified
- The reason why the RR and OR differ and why has now been clarified
- The utility of the RR as an effect magnitude measure has been questioned given that it is the ratio of two posterior probabilities, both of which are highly dependent on the prevalence of the outcome in a study
- The status of the OR as an effect measure has been clarified.

*What is the implication and what should change**now?*

- The RR should no longer be used in clinical trials, cohort studies and meta-analyses
- The OR is the effect measure of choice in analytic epidemiological studies
- The risk difference and numbers needed to treat should be derived from the OR

Author statement

Suhail Doi: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing – original draft; Writing – review & editing. **Luis Furuya-Kanamori:** Conceptualization; Formal analysis; Funding acquisition; Methodology; Writing – review & editing **Chang Xu:** Data curation; Software; Formal analysis; Methodology; Writing – review & editing **Lifeng Lin:** Data curation; Writing – review & editing **Tawanda Chivese:** Visualization; Writing – review & editing; **Lukman Thalib:** Conceptualization; Formal analysis; Funding acquisition; Methodology; Writing – review & editing

Journal Pre-proof

Declarations of interest: none

Journal Pre-proof