

**TUTORIAL**

# Impact of your results: Beyond the relative risk

Bob Siegerink PhD<sup>1,2</sup>   | Jessica L. Rohmann MScPH<sup>1,3</sup>  

<sup>1</sup>Center for Stroke Research Berlin, Charité Universitätsmedizin-Berlin, Berlin, Germany

<sup>2</sup>Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

<sup>3</sup>Institute of Public Health, Charité Universitätsmedizin-Berlin, Berlin, Germany

**Correspondence**

Bob Siegerink, Charité Universitätsmedizin-Berlin, Center for Stroke Research, Charitéplatz 1, 10117 Berlin, Germany.  
Email: bob.siegerink@charite.de

**Abstract**

Universally, reporting guidelines emphasize the importance of using point estimates that indicate the strength of an effect. A single statement of the presence (or absence) of “statistical significance” and/or a *P* value alone do not provide sufficient information. Instead, an estimate of relative risk with a corresponding confidence interval should be routinely provided. Unfortunately, the context of the reported relative risk is often omitted, thereby hampering the readers’ understanding of the impact of the results. Additionally, commonly used binary outcomes might not be sensitive enough to fully convey the clinical relevance of an intervention or risk factor. This tutorial underlines the role of the context of results presented in clinical research papers. It also provides suggestions of meaningful ways to illustrate the impact of your own results by going beyond the relative risk.

**KEYWORDS**

epidemiology, ordinal outcome, relative risk, statistics

**Essentials**

- Relative risk estimates do not always convey the necessary context for a meaningful interpretation of the data.
- Next to a description of the absolute risk, measures like the number needed to treat (NNT) and population attributable fraction (PAF) help contextualize the obtained relative risks.
- In some cases, ordinal outcomes can be used instead of binary outcomes to prevent the loss of relevant information.

## 1 | THE CONTEXT OF A RELATIVE RISK

The abbreviation “RR” generally refers to relative risk, and should thus be interpreted as a ratio of two estimates of risk in epidemiological studies, be it a ratio of odds (odds ratio), risks (risk ratio), rates (rate ratio), or hazards (hazard ratio). Knowing which *type* of comparison is being made is essential for interpretation. As such, the specific effect size measure used should always be explicitly mentioned. Regardless of type, the relative risk has multiple practical applications in clinical research, as it not only provides an answer to the question “is there an association?” but also “how large is the association?”, thereby supplying information about the

magnitude and direction of the effect (e.g, harmful or protective). This is because the relative risk is calculated by dividing the risk of the outcome in the group with the exposure of interest by the risk of the outcome observed in the unexposed group. However, as with all ratio measurements, the absolute size of the numerator and denominator are lost<sup>1</sup> (see Box 1). This loss of information impedes interpretation, as it becomes difficult to understand the true impact of exposure. In this paper, we suggest multiple ways you, as both an author and critical reader of scientific publications, can put relative risks into meaningful context. In order to do so, we first need to establish a typology of domains in which relative risks are being used.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. *Research and Practice in Thrombosis and Haemostasis* published by Wiley Periodicals, Inc on behalf of International Society on Thrombosis and Haemostasis.

Generally speaking, all clinical research can be divided into four categories using typical questions a patient might ask a doctor when confronted with symptoms: “What is wrong with me?”, “What caused my disease?”, “Can you treat it?”, and “What will happen to me?”. You might recognize the four domains of clinical research implicit in these questions: diagnostic, etiologic, intervention (treatment), and prediction research. Even though the relative risk is traditionally used in all four areas, the context and interpretation of this effect measure is markedly different in each framework. Research questions rooted in *Treatment & Etiology* are focused on understanding the effect of a single factor. The focus differs for questions of *Diagnosis & Prediction*, where the emphasis lies on distinguishing specific groups within the study population and often involves combining information from (multiple) factor(s). This tutorial will focus on the former category of research questions.

## 2 | QUANTIFYING THE IMPACT

In both intervention and etiologic research domains, it is pivotal to isolate the effect of interest from other factors by selecting a suitable study design and sample, identifying and measuring potential confounding factors, as well as choosing appropriate statistical methods. In clinical trials, this means using tools such as randomization and concealment of treatment allocation to minimize bias and isolate the treatment effect. In observational studies, confounding control is needed to single out the effect of a causal factor using strategies such as stratification or regression adjustment. Recently, more advanced methods such as propensity score matching and inverse probability of treatment weighting have been popularized. Ultimately, it is often the relative risk that provides insight into the strength of an effect, be it a treatment effect or an estimate of the causal effect of a particular risk factor. However, as the ratio of two risk measures omits critical information about the absolute risk, one cannot easily judge the external validity of a study nor understand the contribution of a single factor to the total risk with this metric alone.

## 3 | ABSOLUTE RISK

The *internal validity* of a study refers to the extent to which the study minimizes potential biases, which could render the results invalid. In contrast, and perhaps more interesting in this context, *external validity* of a study refers to the extent to which the individual study results can be generalized to other populations. To make this assessment, a thorough understanding of the study population characteristics, customarily presented in Table 1 of scientific articles, is important. Next to this, the overall absolute risk of the outcome of interest plays an integral role in the external validity of a study. For example, meaningful differences in absolute risks between the study population and another population could indicate a difference in the prevalence of comorbidities or a difference in how healthcare is organized. This can have effects on both absolute and relative risk measures; when the absolute risk in a population is *high*, or the exposure is highly

prevalent, relative risk estimates are generally *smaller*. Therefore, without knowing information about the absolute risk *and* the prevalence of the exposure, it is difficult to determine whether the generalization of results from one population to another is appropriate.

## 4 | “NUMBER NEEDED TO”

Some of the information needed to assess the *contribution of a single factor to the total risk* in interventional studies is provided by the *risk difference (RD)*, a useful absolute measure that allows for ready calculation of the *number needed to treat (NNT)*. Computed by taking the inverse of the absolute risk difference, the NNT provides an idea of efficacy (or effectiveness) by providing an answer to the question, “How many patients do we need to treat in order to prevent the occurrence of a single event?” The usefulness of the NNT in the context of a randomized controlled trial is particularly convenient, since the NNT can be easily calculated from the crude, unadjusted analyses in well-conducted trials. On the other hand, in observational research, we should not simply calculate NNT from crude data due to bias and confounding. Certain regression-based approaches such as Poisson and Cox proportional hazards regression are capable of controlling for confounding and thus come to a more valid estimate NNT.<sup>2,3</sup>

The NNT concept can also be extended to apply to harmful exposures (i.e., *number needed to harm, NNH*; see Box 1) as well as potential biomarkers informing treatment decisions (i.e., *number needed to screen*). Essential for the interpretation of NNT and derived measures is the time frame to which they are applied. In some fields, there is little confusion surrounding this aspect, as standard practice dictates a fixed time period. For example, most stroke studies involve an assessment of functional outcome at the time point of 3 months post-stroke, making the interpretation and comparison of reported NNT findings unambiguous across different studies.<sup>4</sup> However, in the context of chronic disease treatment and variable follow-up time, failure to explicitly mention the time frame could result in a misleading representation of the NNT and inaccurate comparisons.

## 5 | POPULATION ATTRIBUTABLE FRACTION

A more common approach to illustrate the *contribution of a single factor to the total risk* in observational research settings is to show to what extent the risk factor of interest contributes to the *total risk* in the entire study population. The population attributable fraction (PAF; sometimes PAR, population attributable risk) provides this information;<sup>5</sup> see sample PAF calculation in Boxes 1 and 2. The PAF can be seen as the proportion of a population’s incidence of a given disease that can be accounted for by exposure to the risk factor of interest, and is distinct from the attributable fraction which is simply the risk difference divided by the risk among the exposed. When interpreting this measure, it is often colloquially said that the removal of the exposure (hypothetically or literally) should decrease the total incidence of disease by the

### Box 1 Practical example: putting your results into perspective

The data displayed below show the results of two fictitious cohort studies examining the role of exposure *X* on thrombotic disease *D* in young women. For simplicity's sake, we will assume an unbiased study, i.e., no selection bias, information bias, and confounding present.

	Developed disease <i>D</i>	No disease <i>D</i>	Total
Exposed to <i>X</i>	50	950	1000
Unexposed	25	2975	3000
Total	75	3925	4000

First, we will use the table to calculate the risk among young women exposed and not exposed to *X* and compare these risks using a risk ratio.

#### Risk of disease *D* (cumulative incidence):

Among the exposed	$50/1000 = 0.05$
Among the unexposed	$25/3000 = 0.0083$

#### Risk ratio (type of relative risk, RR):

RR	risk of disease among the exposed/risk among the unexposed	$0.05/0.0083 = 6.0$
----	--	---------------------

A RR of 6.0 indicates a six-fold increased risk for disease *D* among the young women exposed to *X* compared to the unexposed. An interesting find, but to grasp the impact of this observation, it is important to compare the risks on the absolute scale starting off with a simple overall absolute risk. Subsequently, we can calculate the risk difference

#### Absolute risk of disease:

Absolute risk	Number of disease cases/Total population	$75/4000 = 0.019$
---------------	--	-------------------

#### Risk difference (RD):

RD	Risk of disease among the exposed - Risk among the unexposed	$0.05 - 0.0083 = 0.042$
----	--	-------------------------

Both measures on the absolute scale are not very high. The low absolute risk and the risk difference put the 6-fold increase in risk in some perspective. But how can we make these numbers even more understandable in a clinically relevant context? We can calculate the number needed to harm, which tells us how many young women on average would need to be exposed to *X* for one additional woman to develop disease *D*.

#### Number needed to harm (NNH):

NNH	Inverse of the risk difference = $1/RD$	$1/0.042 = 24$
-----	---	----------------

This means that, on average, for every 24 young women exposed to *X*, 1 will develop disease *D* (24 "need to be exposed" for 1 to be "harmed"). Since we know diseases are multicausal, it may be useful to know what proportion of all cases of disease in our study population can be attributed to exposure *X*. To do this, we first need to calculate two pieces of information:

1. The proportion of exposed cases of disease:

Proportion exposed cases	Number of exposed participants with disease/Total participants with disease	$50/75 = 0.667$ (=67%)
--------------------------	---	------------------------

2. The attributable fraction among the exposed (AF):

AF	$RD/Risk\ among\ the\ exposed$	$0.042/0.05 = 0.84$ (=84%)
----	--------------------------------	----------------------------

Now, we simply multiply these two numbers together to get the population attributable fraction (PAF) as follows.

#### Population attributable fraction (PAF):

PAF	Proportion of exposed cases $\times$ Attributable fraction among exposed	$0.667 \times 0.840 = 0.56$ (=56%)
-----	--	------------------------------------

This example illustrates three valuable lessons: First, it shows that in order to interpret a high relative risk, additional context is needed. When the absolute risk for a particular disease is low, a high relative risk is less concerning. The risk difference is therefore often more intuitive and can readily be interpreted as "excess risk". Second, calculating the NNH (or a related measure) provides further useful information that can aid in decision making. In the case of NNH, lower numbers are more concerning. Third, the PAF answers a different contextual question; specifically, what fraction of all cases can be attributed to the exposure? Taking the prevalence of the exposure in the population into account, the PAF tells us that more than half of the total cases of disease *D* could be attributed to exposure *X*.

## Box 2 Multiple measures paint the full picture

For a complete understanding of the impact of your results, it is important to consider multiple measures as discussed in this paper. Below, we provide six different variations on the cohort study studying the effect of exposure X on disease D in 4000 young women, as introduced in Box 1. Notice how looking at one measure alone could lead you to erroneously believe that two very different scenarios produce the same result.

### Original example

	D+	D-		AR	0.02
X+	50	950	1000	RR	6
X-	25	2975	3000	NNT	24
	75	3925	4000	PAF	0.56

### Six variations

	D+	D-		AR	0.003		D+	D-		AR	0.19
X+	10	1799	1809	RR	6	X+	720	2762	3482	RR	6
X-	2	2189	2191	NNT	217	X-	18	500	518	NNT	6
	12	3988	4000	PAF	0.70		738	3262	4000	PAF	0.81
	D+	D-		AR	0.22		D+	D-		AR	0.024
X+	505	1615	2120	RR	1.2	X+	89	1906	1995	RR	18
X-	370	1510	1880	NNT	24	X-	5	2000	2005	NNT	24
	875	3125	4000	PAF	0.10		94	3906	4000	PAF	0.89
	D+	D-		AR	0.75		D+	D-		AR	0.028
X+	2500	1	2501	RR	3	X+	90	1950	2040	RR	3.5
X-	500	999	1499	NNT	1.5	X-	25	1935	1960	NNT	32
	3000	1000	4000	PAF	0.56		115	3885	4000	PAF	0.56

PAF.<sup>6</sup> For this reason, the PAF is considered an appropriate tool to evaluate the impact of individual risk factors on a population level. Importantly, the PAF is dependent not only on the strength of the effect of the exposure on the outcome, but also on the absolute risk of outcome as well as the prevalence of the exposure in the population; two additional valuable pieces of contextual information (see Box 2). For these reasons, PAFs calculated using data from well-conducted cohort and case-control studies provide meaningful information beyond the relative risk. When calculating and reporting PAFs, two major issues must be considered: as PAFs are often obtained from observational data, the PAFs are dependent on the extent of bias in the data. Additionally, many thrombotic diseases are multicausal, meaning there are (theoretically) multiple ways to prevent the onset of a single event.<sup>7</sup> Therefore, it follows that the sum of PAFs for *all* risk factors is not capped at 1; in fact, the sum of all PAFs for a particular outcome often exceeds this number.

## 6 | BEYOND THE BINARY OUTCOME

To understand the impact of a treatment or biomarker, it can be useful to look beyond a dichotomous outcome and instead make use of a clinically relevant measurement with multiple levels. Many research fields regularly use such approaches to assess outcomes with multiple levels, especially when they are ordinal in nature (e.g. outcomes after birth, pain management, rheumatology and surgery).<sup>8-11</sup> This approach is distinct from the practice of using so-called “surrogate outcomes”, in which binary outcomes are also replaced, for example, by a biomarker.<sup>12</sup> This practice is debated, primarily as the clinical relevance is often unclear. If one wants to create or adopt ordinal outcomes that are not surrogate outcomes, the clinical relevance of the outcome categories needs to be a guiding principle.

A well-known example from stroke research is the modified Rankin Scale (mRS), which has seven ordered steps of patient functionality, ranging from “no impairment” (mRS = 0) to “dead” (mRS = 6)

and is traditionally measured three months after the initial stroke.<sup>13</sup> The scale is interview-based, and as such, is not as objective as some binary outcomes. However, the scale has a good inter-observer agreement, especially after adequate training.<sup>14</sup> Specific statistical analysis techniques, such as proportional ordinal logistic regression, make full use of the ordered structure of this outcome measurement by quantifying the odds of a person ending up in a better category of outcome (e.g. one mRS step lower) if exposed to treatment.

Some researchers unnecessarily dichotomize ordinal scales in order to calculate absolute risks and subsequently a NNT. For example, mRS is often dichotomised to model the risk of dependency a after stroke (using a cut-off at mRS 2/3). Although this dichotomous approach may lead to answers that are seemingly easy to interpret, this practice does not answer the arguably more important global question of whether patients are *overall* better off with treatment. Additionally, collapsing ordered outcome scores into a dichotomous variable results in a loss of important information about severity by reducing the resolution of the outcome measure. Fortunately, it is possible to translate ordinal effect measures into NNT estimates using modern methods.<sup>15,16</sup>

Currently, the broader field of clinical thrombosis and hemostasis does not (yet) have a well-developed and widely used ordinal outcome akin the mRS in stroke. But, for example, even adding a simple distinction between fatal/non-fatal outcome (no disease, non-fatal disease, fatal disease) provides meaningful added information. Categorizing venous thromboembolism by location (no venous thromboembolism, superficial venous thrombosis, deep vein thrombosis, pulmonary embolism) or using ordinal bleeding severity scores in hemophilia studies are just some ways researchers in the field could look beyond binary outcomes.

## 7 | SUMMARY

Relative risk estimates provide an answer to the research question at hand; however, due to their relative nature, they do not convey the necessary context for meaningful interpretation. Ultimately, it is the duty of the authors of a publication to go beyond the relative risk and explicitly provide the context needed to help readers understand the impact of their results. Measures like the NNT and PAF add additional dimensions to the effect estimate (i.e. absolute risk and prevalence of the exposure), and as such, help contextualize the obtained results. When possible and of (clinical) relevance, authors may consider using ordinal instead of binary outcome measurements and corresponding analyses techniques to prevent the loss of important information.

## RELATIONSHIP DISCLOSURE

The authors have nothing to disclose.

## AUTHOR CONTRIBUTIONS

BS: concepts, draft manuscript, JLR concepts and critical revision of manuscript.

## ORCID

Bob Siegerink  <http://orcid.org/0000-0002-8454-9142>

Jessica L. Rohmann  <http://orcid.org/0000-0003-2420-5716>

## REFERENCES

1. Siegerink B, Maino A, Algra A, Rosendaal FR. Hypercoagulability and the risk of myocardial infarction and ischemic stroke in young women. *J Thromb Haemost*. 2015;13:1568–75.
2. Laubender RP, Bender R. Estimating adjusted risk difference (RD) and number needed to treat (NNT) measures in the Cox regression model. *Stat Med*. 2010;29:851–9.
3. Austin PC, Laupacis A. A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *Int J Biostat*. 2011;7:6.
4. Ebinger M, Harmel P, Nolte CH, Grittner U, Siegerink B, Audebert HJ. Berlin prehospital or usual delivery of acute stroke care—study protocol. *Int J Stroke*. 2017;12:653–8.
5. Mansournia MA, Altman DG. Population attributable fraction. *BMJ*. 2018;360:k757.
6. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008.
7. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. 1999;353:1167–73.
8. LaValley MP, Felson DT. Statistical presentation and analysis of ordered categorical outcome data in rheumatology journals. *Arthritis Rheum*. 2002;47:255–9.
9. Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *J Clin Epidemiol*. 1997;50:45–55.
10. Ozragat-Baslanti T, Blanc P, Thottakkara P, et al. Preoperative assessment of the risk for multiple complications after surgery. *Surgery*. 2016;160:463–72.
11. Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *Int J Epidemiol*. 1997;26:1323–33.
12. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med*. 2012;31:2973–84.
13. Farrell B, Godwin J, Richards S, Warlow C. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. *J Neurol Neurosurg Psychiatry*. 1991;54:1044–54.
14. Quinn TJ, Lees KR, Hardemark H-G, Dawson J, Walters MR. Initial experience of a digital training resource for modified Rankin scale assessment in clinical trials. *Stroke*. 2007;38:2257–61.
15. Bath PMW, Lees KR, Schellinger PD, et al. Statistical analysis of the primary outcome in acute stroke trials. *Stroke*. 2012;43:1171–8.
16. Rahlfs VW, Zimmermann H, Lees KR. Effect size measures and their relationships in stroke studies. *Stroke*. 2014;45:627–33.

**How to cite this article:** Siegerink B, Rohmann JL. Impact of your results: Beyond the relative risk. *Res Pract Thromb Haemost*. 2018;2:653–657. <https://doi.org/10.1002/rth2.12148>