Draft version February 29, 2020
Typeset using LaTeX default style in AASTeX63

The relative risk is logically inconsistent

Ilija Barukčić[1]

[1] *Medical doctor*
Horandstrasse
Jever, 26441, Germany

(Received February 29, 2020; Revised February 29, 2020; Accepted February 29, 2020)

Submitted to viXra

ABSTRACT

Many different measures of association are used by medical literature, the relative risk is one of these measures. However, to judge whether results of studies are reliable, it is essential to use among other measures of association which are logically consistent. In this paper, we will present how to deal with one of the most commonly used measures of association, the relative risk. The conclusion is inescapable that the relative risk is logically inconsistent and should not be used any longer.

Keywords: Statistical methods, logical consistency — — measures of relationships — relative risk

1. INTRODUCTION

The relation between data actually obtained (the sample) and hypotheses is studied by a mathematical and conceptual discipline called statistics. In particular, the data of a sample can be biased which can be a source of incorrect conclusions with serious consequences.
In general, in almost all scientific research, empirical data or facts are investigated by specific statistical methods in order to evaluate some hypotheses of a particular kind [1]. However, the statistical methods, in turn, need to be at least logically consistent. Central to the correctness of statistical methods is this problem of logical consistency, which concerns the justification of any statistical method. In point of fact, even if statistics provide us with various methods and means to evaluate hypotheses it is insightful to consider that statistics may harbour a large variety of errors and logical fallacies too even if sometimes hidden behind highly abstract mathematical stuff. One of such commonly used statistical methods is the risk ratio or relative risk (RR) which is designed to detect or to measure the relation between an exposure to an event A_t and an outcome of an event B_t.
Despite the frequent use of RR, founded doubts regarding the correctness and logical consistency of RR are not automatically excluded. In any case, the issue is not how often RR is used, but whether RR is logically correct or not logically correct.

2. MATERIAL AND METHODS

From the beginning of statistics onward the same is interrelated with probability theory. However, what kinds of ‘things ’are probabilistic statements, or more generally under which circumstances are probability statements true or false?

2.1. *Material*

Corresponding author: Ilija Barukčić
Barukcic@t-online.de

[1] https://plato.stanford.edu/entries/statistics/

The subject of study in statistics is among other the relation between data and hypotheses. Summing up, it remains problematic to study anything with some definitions.

2.1.1. *Definitions*

Definition 2.1 (Independence)**.**

The independence of two events $\mathrm{A_t}$ and $\mathrm{B_t}$ regarded from the standpoint of a certain observer was defined by de Moivre on page 7 as "... therefore, those two Events being independent, the Probability of their both happening will be 1/13 * 1/13 = 1/169 "[2] and Kolmogoroff [3] and other, as

$$p(B_{\mathrm{t}}) \times p(A_{\mathrm{t}}) = p(a_{\mathrm{t}}) \tag{1}$$

where $\mathrm{p(A_t)}$ denotes the probability of an event $\mathrm{A_t}$ at the Bernoulli trial t and $\mathrm{p(B_t)}$ denotes the probability of another event $\mathrm{B_t}$ at the same Bernoulli trial t while $\mathrm{p(a_t)}$ denotes the joint probability of $\mathrm{p(A_t}$ AND $\mathrm{B_t)}$ at the same Bernoulli trial t.

Definition 2.2 (Dependence)**.**

The Dependence of two events $\mathrm{A_t}$ and $\mathrm{B_t}$ regarded from the standpoint of a certain observer is defined as

$$p(a_{\mathrm{t}}) = (p(B_{\mathrm{t}}) \times p(A_{\mathrm{t}}))^{1/2} \tag{2}$$

where $\mathrm{p(A_t)}$ denotes the probability of an event $\mathrm{A_t}$ at the Bernoulli trial t and $\mathrm{p(B_t)}$ denotes the probability of another event $\mathrm{B_t}$ at the same Bernoulli trial t while $\mathrm{p(a_t)}$ denotes the joint probability of $\mathrm{p(A_t}$ AND $\mathrm{B_t)}$ at the same Bernoulli trial t while the dependence of n events[4] follows as

$$p(a_{1,\mathrm{t}}, a_{2,\mathrm{t}}, ..., a_{\mathrm{n,t}}) = (p(A_{1,\mathrm{t}}) \times p(A_{2,\mathrm{t}}) \times ... \times p(A_{\mathrm{n,t}}))^{1/\mathrm{n}} \tag{3}$$

Definition 2.3 (Contingency table)**.**

The relationship between two Binomial or Bernoulli distributed random variables $\mathrm{A_t}$ and $\mathrm{B_t}$ at a certain Bernoulli trial (or period of time) t can be illustrated by a 2 by 2 table. Furthermore, a 2 by 2 contingency table is able to provide a basic picture of the interrelation between two binomial distributed random variables and is of use to analyse the relationships between them in detail. Karl Pearson was the first to use the term contingency table in his paper "On the Theory of Contingency and Its Relation to Association and Normal Correlation"[5].

Relative risk		Outcome YES	Outcome NO	Total
Exposed	YES	$p(a_t)$	$p(b_t)$	$p(A_t)$
	NO	$p(c_t)$	$p(d_t)$	$p(\underline{A}_t)$
	Total	$p(B_t)$	$p(\underline{B}_t)$	+1

[2] https://doi.org/10.3931/e-rara-10420

[3] https://doi.org/10.1007/978-3-642-49888-6

[4] Ilija Barukčić, Die Kausalität, Hamburg: Wissenschaftsverlag, 1989, pp. 57-59.

[5] https://archive.org/details/cu31924003064833/page/n2/mode/2up

where $p(a_t)$ denotes the joint probability of A_t and B_t, $p(b_t)$ denotes the joint probability of A_t and Not B_t, $p(c_t)$ denotes the joint probability of not A_t and B_t and $p(d_t)$ denotes the joint probability of not A_t and Not B_t.

Definition 2.4 (Basic relationships between probabilities of a 2 by 2 table)**.**

In general, it is

$$p(A_t) = p(a_t) + p(b_t) \tag{4}$$

and

$$p(NotA_t) = 1 - p(A_t) = p(c_t) + p(d_t) \tag{5}$$

and

$$p(B_t) = p(a_t) + p(c_t) \tag{6}$$

and

$$p(NotB_t) = 1 - p(B_t) = p(b_t) + p(d_t) \tag{7}$$

where $p(a_t)$ denotes the joint probability of A_t and B_t. In general, it is

$$p(a_t) + p(b_t) + p(c_t) + p(d_t) = +1 \tag{8}$$

Definition 2.5 (Relative risk)**.**

The degree of association between the two binomial variables can be assessed by a number of very different coefficients, the relative risk [6] is one of them. In this context, see also Sir Ronald Aylmer Fisher's (1890 - 1962) contribution in his publication "The Logic of Inductive Inference"[7] . In general, relative risk is defined as

$$RR(A_t, B_t) = \frac{\frac{p(a_t)}{p(A_t)}}{\frac{p(c_t)}{p(NotA_t)}} = \frac{p(a_t) \times p(NotA_t)}{p(A_t) \times p(c_t)} \tag{9}$$

That what scientist generally understand by relative risk is the ratio of a probability of an event occurring with an exposure versus the probability of an event occurring without an exposure. In other words, relative Risk = (Probability of event in exposed group) / (Probability of event in not exposed group). An $RR(A_t,B_t) = +1$ means that exposure does not affect the outcome or both are independent of each other while $RR(A_t,B_t)$ less than +1 means that the risk of the outcome is decreased by the exposure. In this context, an $RR(A_t,B_t)$ greater than +1 denotes that the risk of the outcome is increased by the exposure. Widely known problems with odds ratio [8] [9] and relative risk [10] are already documented [11] [12] in literature.

Definition 2.6 (Exclusion relationship)**.**

The exclusion relationship is defined as

$$p(A_t \mid B_t) = p(b_t) + p(c_t) + p(d_t) = +1 \tag{10}$$

Definition 2.7 (Conditio sine qua non relationship)**.**

The conditio sine qua non relationship is defined as

$$p(A_t \longleftarrow B_t) = p(a_t) + p(b_t) + p(d_t) = +1 \tag{11}$$

[6] https://www.ncbi.nlm.nih.gov/books/NBK430824/
[7] https://www.jstor.org/stable/pdf/2342435.pdf?seq=1
[8] https://www.ncbi.nlm.nih.gov/pubmed/9832001
[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6178613/
[10] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC522855/
[11] https://www.crcpress.com/Principles-of-Biostatistics-Second-Edition/Pagano-Gauvreau/p/book/9781138593145
[12] https://www.biometricsociety.org/wp-content/uploads/2018/07/IBS-IBC2012-Final-Programme.compressed.pdf

Remark. Since thousands of years, human mankind is familiar with the concept of a necessary condition. For example, we all know that air or gaseous oxygen is a necessary for (human) life. Without gaseous oxygen, there is no (human) life. However, the first documented mathematiziation of the concept of a necessary condition (**conditio sine qua non**) was published by Barukčić 1989 [13]. Conditions may be necessary without being sufficient and vice versa. Sufficient conditions need not to be necessary. However, there may exist conditions which are both, necessary and sufficient.

Definition 2.8 (Conditio per quam relationship)**.**

The conditio per quam relationship is defined [14] [15] [16] [17] [18] [19] [20] [21] as

$$p(A_{\text{t}} \longrightarrow B_{\text{t}}) = p(a_{\text{t}}) + p(c_{\text{t}}) + p(d_{\text{t}}) = +1 \tag{12}$$

Conditio per quam		Street is wet: YES	Street is wet: NO	
It is raining	YES	+1	+0	A_t
It is raining	NO	+1	+1	$\underline{A}_t$
		B_t	$\underline{B}_t$	

Remark. Chile's **Atacama desert** is a desert plateau covering about 1,000-km (600-mi) strip of land on the Pacific coast.In contrast to the equator where is rains very often, the Atacama desert is widely considered as world's driest nonpolar desert with an average rainfall of as little as 0.04 inches per year. However, a **conditio per quam** relationship between raining and a street which is wet can be investigated even under these circumstances.

Under conditions of the Atacama desert a thought experiment is performed and the following data were achieved. It rained seldom thus that the experimenter put 999 times by himself some water on the street where he performed measurements in order to study what happens if it is not raining. The realtive risk can be calculated as

The relative risk is calculate as

$$RR(A_{\text{t}}, B_{\text{t}}) = \frac{p(a_{\text{t}}) \times p(NotA_{\text{t}})}{p(A_{\text{t}}) \times p(c_{\text{t}})} = \frac{1000 \times 1000}{999 \times 1000} = 1.0010 \tag{13}$$

The relative risk can be calculated as RR = 1.0010 while the 95% CI is 0.9990 to 1.0030 and the P value is P = 0.3173. In other words, according to the relative risk, raining is not a risk factor of a wet street or raining and a wet street are independent of each other. However, such a result is far away from any possible reality. Therefore, what is becoming more and more visible is the complete collapse of the relative risk. Formally, even if relative risk is able to recognise a **conditio per quam relationship** in reality the same does not. Depending upon **study design and other factors**, the relative risk present us a false and completely misleading picture of objective reality. On the other there is no longer any doubt that it is really not necessary to hold onto relative risk.

[13] Ilija Barukčić, Die Kausalität, Hamburg: Wissenschaftsverlag, 1989
[14] https://aip.scitation.org/doi/abs/10.1063/1.3567453
[15] https://aip.scitation.org/doi/abs/10.1063/1.4773147
[16] https://www.scirp.org/journal/paperinformation.aspx?paperid=69478
[17] https://www.scirp.org/journal/paperinformation.aspx?paperid=67272
[18] http://www.ijapm.org/show-64-515-1.html
[19] https://www.sciencedirect.com/science/article/pii/S1875389211006626
[20] https://view.publitas.com/amph/rjr_2018_4_art-02/page/1
[21] http://jddtonline.info/index.php/jddt/article/view/3385

Conditio per quam (Atacama desert)		The street is wet		
		YES	NO	
It is raining	YES	1000	0	1000
	NO	999	1	1000
		1999	1	2000

2.1.2. *Axioms*

Axiom 1. Lex identitatis [22] [23] [24].

$$+1 = +1 \tag{14}$$

Axiom 2. Lex contradictionis[25] [26] [27].

$$+0 = +1 \tag{15}$$

2.2. *Methods*

2.2.1. *Proof methods*

Proof methods like a direct proof [28], proof by contradiction[29], modus ponens[30], modus inversus[31] [32] and other methods are of use to detect inconsistencies and inadequacies in scientific theories.

[22] https://www.scirp.org/journal/paperinformation.aspx?paperid=69478
[23] https://www.ncbi.nlm.nih.gov/nlmcatalog/101656626
[24] https://doi.org/10.22270/jddt.v9i2.2389
[25] https://www.ncbi.nlm.nih.gov/nlmcatalog/101656626
[26] https://doi.org/10.22270/jddt.v9i2.2389
[27] https://doi.org/10.22270/jddt.v10i1-s.3856
[28] http://www.ijmttjournal.org/Volume-65/Issue-7/IJMTT-V65I7P524.pdf
[29] https://aip.scitation.org/doi/abs/10.1063/1.3567453
[30] http://www.ijmttjournal.org/Volume-65/Issue-7/IJMTT-V65I7P524.pdf
[31] http://www.ijmttjournal.org/Volume-65/Issue-7/IJMTT-V65I7P524.pdf
[32] https://vixra.org/pdf/1911.0410v1.pdf

3. RESULTS

3.1. *Independence of A_t and B_t*

Theorem 1 (Independence of A_t and B_t).

Claim.
In general, under circumstances of independence of of A_t and B_t, it is

$$p(B_t) = \frac{p(a_t)}{p(A_t)} \tag{16}$$

Proof By Modus Ponens.
The premise of modus ponens [33] in the case of independence according to de Moivre [34] and Kolmogoroff [35] and other, is that

$$p(B_t) \times p(A_t) = p(a_t) \tag{17}$$

Dividing by $p(A_t)$, we obtain

$$\frac{p(B_t) \times p(A_t)}{p(A_t)} = \frac{p(a_t)}{p(A_t)} \tag{18}$$

At the end, the conclusion

$$p(B_t) = \frac{p(a_t)}{p(A_t)} \tag{19}$$

is true.
Quod erat demonstrandum.

3.2. *Independence of Not A_t and B_t*

Theorem 2 (Independence of not A_t and B_t).

Claim.
In general, under circumstances of independence, it is

$$p(B_t) = \frac{p(c_t)}{p(NotA_t)} \tag{20}$$

Proof By Modus Ponens.
The premise of modus ponens in the case of independence according to de Moivre [36] and Kolmogoroff [37] and other, is that

$$p(B_t) \times p(NotA_t) = p(c_t) \tag{21}$$

Dividing by $p(\text{Not } A_t)$, we obtain

$$\frac{p(B_t) \times p(NotA_t)}{p(NotA_t)} = \frac{p(c_t)}{p(NotA_t)} \tag{22}$$

At the end, the conclusion

$$p(B_t) = \frac{p(c_t)}{p(NotA_t)} \tag{23}$$

is true.
Quod erat demonstrandum.

[33] http://www.ijmttjournal.org/archive/ijmtt-v65i7p524
[34] https://doi.org/10.3931/e-rara-10420
[35] https://doi.org/10.1007/978-3-642-49888-6
[36] https://doi.org/10.3931/e-rara-10420
[37] https://doi.org/10.1007/978-3-642-49888-6

3.3. *Case $p(a_t) = 0$: The relative risk RR is defined*

Theorem 3 (Case $\mathbf{p(a_t) = 0}$: The relative risk RR is defined).

Claim.
In general, under circumstances $p(a_t) = 0$, the relative risk RR is determined as

$$RR(A_t, B_t) = \frac{\frac{p(a_t)}{p(A_t)}}{\frac{p(c_t)}{p(notA_t)}} = \frac{p(a_t) \times p(notA_t)}{p(A_t) \times p(c_t)} = 0. \tag{24}$$

Proof By Modus Ponens.
The premise of modus ponens is that the relative risk RR is true. Thus far, it is

$$RR(A_t, B_t) = \frac{p(a_t) \times p(notA_t)}{p(A_t) \times p(c_t)} \tag{25}$$

Under conditions where $p(a_t) = 0$, it is

$$RR(A_t, B_t) = \frac{0 \times p(notA_t)}{p(A_t) \times p(c_t)} \tag{26}$$

Under these circumstances the conclusion

$$RR(A_t, B_t) = 0 \tag{27}$$

is true.
Quod erat demonstrandum.

Remark. Theoretically, the relative risk has the potential to **detect an exclusion relationship**, but only if RR = 0. The following figure may illustrate the basic relationships again.

Relative risk		Outcome		Total
		YES	NO	
Exposed	YES	$p(a_t)$	$p(b_t)$	$p(A_t)$
	NO	$p(c_t)$	$p(d_t)$	$p(\underline{A}_t)$
	Total	$p(B_t)$	$p(\underline{B}_t)$	+1

3.4. *Case $p(b_t) = 0$: The relative risk RR is defined*

Theorem 4 (**Case $\mathbf{p(b_t) = 0}$: The relative risk RR is defined**).

Claim.
In general, under circumstances $p(b_t) = 0$, the relative risk RR is determined as

$$RR(A_t, B_t) = \frac{p(notA_t)}{p(c_t)} \tag{28}$$

Proof By Modus Ponens.
The premise of modus ponens is that the relative risk RR is true. Thus far, it is

$$RR(A_t, B_t) = \frac{p(a_t) \times p(notA_t)}{p(A_t) \times p(c_t)} \tag{29}$$

which is equivalent with

$$RR(A_t, B_t) = \frac{p(a_t) \times p(notA_t)}{(p(a_t) + p(b_t)) \times p(c_t)} \tag{30}$$

Under conditions where $p(b_t) = 0$, the equation before changes to

$$RR(A_t, B_t) = \frac{p(a_t) \times p(notA_t)}{(p(a_t) + 0) \times p(c_t)} \tag{31}$$

or to

$$RR(A_t, B_t) = \frac{p(a_t) \times p(notA_t)}{p(a_t) \times p(c_t)} \tag{32}$$

Under circumstances where $p(b_t) = 0$ the conclusion

$$RR(A_t, B_t) = \frac{p(notA_t)}{p(c_t)} \tag{33}$$

is true.
Quod erat demonstrandum.

Remark. Theoretically, the relative risk RR has the potential to detect a **conditio per quam** relationship, but only if **RR** $>$ **+1**. However, a significant and positive relative risk does not provide evidence of a conditio per quam relationship. Furthermore and depending especially on study design, an existing conditio per quam relationship must not be detected by the relative risk as proofed before. The following figure may illustrate the relationship again.

Conditio per quam		Street is wet		
		YES	NO	
It is raining	YES	+1	+0	A_t
	NO	+1	+1	$\underline{A}_t$
		B_t	$\underline{B}_t$	

3.5. *Case $p(c_t) = 0$: The relative risk RR is not defined*

Theorem 5 (Case $\mathbf{p(c_t) = 0}$: The relative risk RR is not defined).

Claim.
In general, under circumstances $p(c_t) = 0$, the relative risk RR is not defined due to

$$RR(A_t, B_t) = \frac{p(a_t) \times p(d_t)}{0} \tag{34}$$

Proof.
The premise of modus ponens is that the relative risk RR is true. Thus far, again it is

$$RR(A_t, B_t) = \frac{p(a_t) \times p(notA_t)}{p(A_t) \times p(c_t)} \tag{35}$$

which is equivalent with

$$RR(A_t, B_t) = \frac{p(a_t) \times (p(c_t) + p(d_t))}{(p(a_t) + p(b_t)) \times p(c_t)} \tag{36}$$

Under conditions where $p(c_t) = 0$, the equation before changes to

$$RR(A_t, B_t) = \frac{p(a_t) \times (0 + p(d_t))}{(p(a_t) + p(b_t)) \times 0} \tag{37}$$

or to

$$RR(A_t, B_t) = \frac{p(a_t) \times (0 + p(d_t))}{(p(a_t) + p(b_t)) \times 0} \tag{38}$$

or to

$$RR(A_t, B_t) = \frac{p(a_t) \times p(d_t)}{0} \tag{39}$$

However, today, the division by zero is not accepted. Therefore, the conclusion that

$$RR(A_t, B_t) = \frac{p(a_t) \times p(d_t)}{0} \tag{40}$$

the relative risk RR is not defined under circumstances where $p(c_t) = 0$ is true.
Quod erat demonstrandum.

Remark. Theoretically, a conditio sine qua non relationship is determined by the fact that $p(c_t) = 0$. However, under these circumstances the relative risk RR collapses into logical absurdity and cannot detect a necessary condition, a **conditio sine qua non** at all. The following figure may illustrate the relationship again.

Conditio sine qau non		Human being alive		
		YES	NO	
Oxygen	YES	$p(a_t)$	$p(b_t)$	$p(A_t)$
	NO	**0**	$p(d_t)$	$p(\underline{A}_t)$
		$p(B_t)$	$p(\underline{B}_t)$	+1

Under conditions of **a necessary and sufficient condition** is determined by $\mathbf{p(c_t) = 0}$ **AND** $\mathbf{p(b_t) = 0}$. However, even under these circumstances, the relative risk breaks together too, because

$$RR(A_t, B_t) = \frac{p(a_t) \times (0 + p(d_t))}{(p(a_t) + 0) \times 0} \tag{41}$$

4. DISCUSSION

The relative risk is a measure of association used in the statistical analysis of the data of different studies. Unfortunately, this publication has recognised the fundamental problems as associated with the relative risk. The relative risk depends to much on study design and can lead to contradictory and highly misleading results. The relative risk cannot recognise the conditio sine qua non relationship (theorem 5) and fails in principle on the conditio per quam relationship. The relative risk [38] is logically inconsistent, unreliable and highly dangerous, and will not be helpful either for decision makers, who will be unable to rely on the results achieved by the relative risk and to translate the same into effective interventions or action, or scientists, who will be unable to relate the relationship between two events to a causal mechanism.

5. CONCLUSION

There are many studies in clinical research published which rely on the relative risk. In this publication, we have investigated the interior logic of the relative risk. We cannot rely on the relative risk. The relative risk is logically inconsistent and completely useless, the relative risk is refuted. The hope that this will help clinicians an others when reading medical literature.

[38] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5841621/