CIRUGÍA ESPAÑOLA

www.elsevier.es/cirugia

Methodological letter

Concept of risk factor as an estimate of association and treatment effect: Measures and meaning☆

Concepto de factor de riesgo como medida de asociación y efecto: tipos y utilización

Jose M. Enríquez-Navascués, Ignacio Aguirre-Allende*

Servicio de Cirugía General y Digestiva, Hospital Universitario Donostia, Donostia, Gipuzkoa, Spain

The determination of risk factors is a typical example of clinical research. Although the concept of risk factor itself is not a precise and unanimous concept in the epidemiological literature[1]. Here we will not refer to how to assess ''surgical risk'' or the ''risk patient[2], but rather to risk factors as variables, and their use as statistical measures of frequency, correlation, association and effect. Risk is essentially an alternative word for probability[3], and because in medicine we are more concerned with adverse effects, a risk factor is seen as something negative, the antithesis of a protective factor. On the other hand, it is worth noting that risk factors are correlational and not necessarily causal.

Measures of the magnitude of association or risk effect

In many surgical studies the researcher is interested in measuring the degree of association between one or more independent factors (exposure, e.g. an intervention) and the occurrence of an adverse effect (dependent variable) (Table 1). This is done by calculating measures of association that quantify such a relationship[4]. Statistically it can be assessed by hypothesis testing, which is known as statistical significance and is represented by the ''p-value''. The problem is that the level of significance (''p'' value) does not inform about the magnitude of the effect and therefore whether it is clinically relevant or not, but simply whether it is statistically significant. It is therefore of interest to go beyond the dichotomous decision (significant/not significant association) and assess the strength of association between the event and the specified risk (groups).

1 Broadly speaking, and for binary qualitative variables, measures of the magnitude of association or effect can be divided into absolute measures of risk (based on differences) and relative measures of risk (based on ratios). For the calculation of measures of association, data are usually presented in contingency tables (2 × 2) (Table 1).

A Absolute measures of risk

A1. Absolute risk (AR): Defined as the probability of a disease or adverse event occurring in the study population. It is expressed as a percentage. In a cross-sectional study it would represent the prevalence, in a cohort, observational or experimental study, the cumulative incidence of risk, and in a case-control study, the prevalence of exposure.

A2. Absolute risk reduction (ARR): In a randomised double-blind clinical trial between placebo and an antiseptic to assess their efficacy in preventing wound infection, we found that with placebo 10% developed infection and with antiseptic 5%. These figures represent the AR for

☆ Please cite this article as: Enríquez-Navascués JM, Aguirre-Allende I. Concepto de factor de riesgo como medida de asociación y efecto: tipos y utilización. Cir Esp. 2022;100:652–655.

* *Corresponding author.*
Correo electrónico: josemaria.enriquez-navascues@osakidetza.eus (J.M. Enríquez-Navascués).

2173-5077/ © 2021 AEC. Published by Elsevier España, S.L.U. All rights reserved.

Table 1 – The contingency table 2 × 2 is the basis for the measurement or association calculations.

	SSI^+	SSI^-	
Placebo (n = 100)	10 (10%)	90	100
Antiseptic (n = 100)	5 (5%)	95	100
	15	185	100

The table represents the example from the text; randomised clinical trial of antiseptic administration vs. placebo and risk of SSI.
SSI: surgical site infection.

Table 2 – Importance of presenting the AR together with the RR and, eventually, the NNT.

	Placebo (n = 1.000)	Analgesic (n = 1.000)	ARR	RRR	NNT
% Rescue A1	200 (20%)	100 (10%)	10	50	10
% Rescue A2	20 (2%)	10 (1%)	1	50	100
% Rescue A3	2 (0,2%)	1 (0,1%)	0,1	50	1.000

In the example, patients are randomly treated with placebo (n = 1000) or 3 different analgesics (A1, A2, A3, with n = 1000 each) and the number of cases in each group requiring Rescue medication is assessed. We see that the RRR is significant and invariable between the different analgesics, however, the differences in ARR and NNT are substantial.
NNT: number needed to treat; AR: absolute risk; ARR: absolute risk reduction; RRR: reduction of relative risk.

each group, the ARR is the difference between them, in this case 5%.

Although absolute measures appear to be associated with a straightforward interpretation, these measures can be misleading when interpreting the effect of a treatment if the rate of the baseline or control outcome is not taken into account. For example, suppose a therapy doubles the probability of a successful outcome: if the success rate of the control group is low, say 1%, the experimental therapy will present a rate of 2%, a rather small increase in absolute terms, as opposed to, if the baseline success rate is 25% and the treatment success rate is 50%, a much larger increase in absolute terms.

A3. Number needed to treat (NNT): A measure of the benefit of a treatment that has a straightforward and simple interpretation. It represents the number of patients who must be treated to prevent an adverse outcome. The NNT is obtained as the reciprocal of the ARR, i.e. 100 divided by the difference of the ARR (in the example of trial 100: 5 = 20, 20 patients need to be treated to prevent one wound infection).

A4. Number Needed to Harm (NNH): It is calculated in the same way as the NNT, and indicates the level of safety of a treatment or intervention, only in this case it takes into account the adverse events of the treatment relative to the control group. Its interpretation is also straightforward: number of people who need to receive a treatment or intervention during a specific time to experience an adverse effect.

5. *Estimation of the population at risk:* In the clinical trial example above, being a prospective, randomised study, it is assumed that the enrolled population is representative of the general population.[5] After randomisation each group of patients should be similar in risk characteristics. The study was prospective and all patients enrolled in both groups were at equal risk of developing an infection, so the risk of infection for each group would have estimated the risk of the general population whether it was with antiseptic or placebo. Inference involves extrapolation of the results obtained in the study population to the general population. In this respect, it should be noted that probability in statistics can be established from two perspectives, the frequentist (more intuitive and practised) and the Bayesian (perhaps more recommendable and informative). We refer the reader to the bibliography[6,7].

B Relative risk measures

B1. Relative risk (RR): This is defined as the ratio or relationship between two absolute probabilities or risks (the concept of risk being equivalent to the epidemiological concept of incidence, as noted above).

The RR measures the strength of the association between exposure and adverse effect, and indicates the probability (the number of times it is more likely) that an adverse effect will develop in those who present a risk factor in relation to those who do not. It is therefore a ratio between patients exposed to a risk factor and patients not exposed to it. It is a ratio that can only take positive values. If the exposure is a risk factor then its value will be > 1 (RR > 1), while if the exposure is a protective factor its value will be < 1; if the RR value is neutral (RR = 1) it is understood that there is no association between exposure and adverse effect or disease.

When we measure the effect of a treatment, it seems strange to see improvement as a ''risk'', however, as we said the term is used in the sense of probability, and is easily interpreted as the number of times more likely that a patient will improve with one treatment compared to another.

The RR can only be calculated when we can estimate the population at risk. As the estimation of risk incidence is only feasible in prospective studies, the calculation of RR is restricted to these types of studies (prospective observational cohort studies and experimental studies such as randomised clinical trials).

In the trial example, the RR would be 0.5 (5%: 10%), i.e. the antiseptic-treated group has a risk 0.5 times that of the placebo group, i.e. half the risk reduction compared to the

placebo group. This may appear to be a very significant improvement, however, this data should be presented together with the ARRs, so in the example above the ARR is 5%, a modest benefit and could well be considered not very cost-effective. For this reason, studies should present ARs and ARRs in addition to the RR (Table 2).

On the other hand, the RR (as well as the OR and HR which will be discussed below) should be presented with confidence intervals (95 or 99%). This calculation indicates the direction of the effect, the statistical significance (if the interval does not encompass the value 1) and the precision of the interval, which is directly related to the sample size of the study.

B2. Odds ratio (OR): As the name indicates, it is the quotient between 2 ''odds'', the problem is that we do not have a good translation of ''odds'' in Spanish, although it is used in the betting world. An example: during the month of June (30 days) in San Sebastian it rains 12 days, while in Malaga it rains only 3 days; that is, the probability of rain in San Sebastian is 12/30 = .4 (40%) and in Malaga it is 3/30 = .1 (10%), with which the RR would be 0.4: 0.1 = 4. On the other hand, the odds of rain/no rain in San Sebastian is 12/18 = .66, and in Malaga 3/27 = .11, so the OR is 0.66/0.11 = 6.

If we replace the word rain with sick and non-rain with healthy, we see that the odds ratio is a ratio of sick to healthy. We also see here that the concept of RR (rain/rain or sick/sick) is close to that of incidence and, as mentioned above, is only used in prospective studies. If the risk of disease or event of interest is low or infrequent (e.g. <10%) the RR and OR values are very similar.

OR as a measure of effect is used in retrospective cohort studies (disease OR), cross-sectional studies (prevalence OR), case-control studies where incidence cannot be calculated because the study population is selected from individuals who have already developed the disease. When we want to calculate measures of effect adjusted for confounding variables, using logistic regression models, the results obtained are adjusted ORs (and not adjusted RRs).[2]

B3. Hazard ratio (HR): Both RR and OR are measures of association at a fixed point in time, and thus constitute a static view.[3] The HR, on the other hand, is a dynamic view, which takes into account the time it takes for the outcome to occur. The hazard ratio, or HR, is the relative risk of an event (e.g. disease progression) occurring over the entire duration of the study. It is therefore a measure of RR adapted to survival analysis (understood as a ''time-to-event'' variable). The interpretation of HR is similar to that of RR. An HR of 1 (null) indicates equal probability of an event in the 2 study groups in the following time interval. A HR > 1 or <1 indicates more or less risk, respectively, in the intervention group than in the control. The HR is an average of the instantaneous hazard rate ratios at each time point over the duration of the study, and is calculated using a (Cox) regression model.

Suppose we want to evaluate two treatments in a trial that looks at how many patients have metastases at 10 years. Of the 20 patients treated in each group, at 10 years and in both groups, 10 have metastases and 10 do not; this gives an OR of 1, there is no advantage of one treatment over the other. But if we look at what happened during the 10 years, there are differences: in a control group most of the metastases occurred in the first 2 years, while in the experimental group they occurred in the last years; therefore, it will have an HR < 1, favourable to the experimental group.

It is a mistake to consider that the HR reports the speed to the event; an HR = 2 does not mean that one group develops the event 2 times faster, but rather that it has a double risk of an event occurring in one group relative to another. An HR for overall survival of 0.75 indicates that there is a 25% reduction ($1-.75 \times 100 = 25\%$) in the instantaneous risk of progression or death in one group relative to the other (or can also be read as 0.75 deaths in one group for every 1 in the other).

The HR does not provide a survival duration data, but it is a good predictor of the true treatment effect in the entire patient population, with the advantage of using all available information, including patients who did not complete the trial for whatever reason, and is most useful when the risk we want to assess is not constant over time.

B4. Relative risk difference (RRD): In the study of bivariate factors, the RRD gives us an idea of the value of both variables. When a RR, OR or HR is >1, the following formula is applied, the result of which is a proportion: RRD = OR-1; thus, if we evaluate the cure obtained by experimental treatment ''A'' vs. control treatment ''B'', if the OR is 1.2, it tells us that treatment B cures 80% of the cures provided by A, or, in other words, that A provides 20% more than B. When the RR, OR or HR is <1, the formula will be: RRD = 1-R, whose interpretation is reciprocal to the previous one.

As noted above, and despite the RRD, RR, OR or HR figures alone cannot be considered high or low, and thus whether the clinical difference is relevant or not if we do not know the absolute proportions.

1 Adjusted effect measures: When more than two variants are involved, bivariate tests do not control for confounding effects. In these cases, multivariate statistics capable of separating (''adjusting'') the independent individual value of each variable must be performed. The most commonly used are binary logistic regression and Cox regression. They can be used as explanatory or predictive models. Another methodological letter is devoted to this topic.

2 Correlation coefficients: The intensity of association between continuous variables (e.g. anastomotic height and incontinence score, etc.) is studied by means of correlation coefficients (Pearson's for parametric data and Spearman's for non-parametric data). In addition to the intensity of the association, linear regression allows us to use one variable to predict another. Its development exceeds the content of this methodological letter.

REFERENCES

1. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? BMJ. 1999;319:1561–5.
2. Shaydakov ME, Tuma F. Operative Risk. 2021 May 4. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2021.
3. Escrig-Sos J. Sobre cómo analizar la credibilidad de un ensayo clínico o metaanálisis cuyo resultado principal se ofrezca en odds ratio, riesgo relativo o hazard ratio. Cir Esp. 2005;78:351–6. http://dx.doi.org/10.1016/S0009-739X(05)70953-2.
4. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. P T. 2008;33:700–11.
5. Sedgwick P. Estiating the population at risk. BMJ. 2012;345. http://dx.doi.org/10.1136/bmj.e6859.
6. Greenland S, Poole C. Living with p values: resurrecting a Bayesian perspective on frequentist statistics. Epidemiology. 2013;24:62–8. http://dx.doi.org/10.1097/EDE.0b013e3182785741.
7. Rendón-Macías ME, Riojas-Garza A, Contreras-Estrada D, Martínez-Ezquerro JD. Análisis bayesiano, Conceptos básicos y prácticos para su interpretación y uso. Rev Alerg Méx. 2018;65:285–98. http://dx.doi.org/10.29262/ram.v65i3.512.