Research Series

Interpretation of Absolute Measures of Disease Risk in Comparative Research

William H. Replogle, PhD; William D. Johnson, PhD

When comparing two groups, one receiving an experimental intervention and the other a placebo or nothing, researchers often wish to assess the disparity in risk of experiencing an event of interest, such as onset of disease. Relative risk, relative risk reduction, and odds ratio are often used to measure the association between potential benefit or harm and the intervention. However, these summary measures reflect relative disparities and are perhaps less useful in clinical practice than measures of absolute benefit or harm. We demonstrate that relative risk reduction is unaffected by the risk of an event in the control group and hence may either overestimate or underestimate the treatment effect. Absolute risk reduction accounts for the baseline control group event rate and is a more realistic quantification of treatment effect than relative measures. Number needed to treat (NNT) estimates the therapeutic effort needed to prevent one additional adverse event. NNT incorporates both relative risk reduction and the event rate without treatment. For a given relative risk reduction, we demonstrate the NNT will increase as the control event rate decreases. Thus, NNT has more-obvious implications for clinical decision making than risk estimates expressed in relative terms.

(Fam Med 2007;39(6):432-5.)

In the Swedish mammography screening trials, 247,010 female participants were randomized to receive an invitation to participate in either a mammography breast screening intervention group or a no-invitation control group. The primary endpoint was breast cancer death, and the median follow-up period was 15.8 years. The authors reported that the intervention resulted in a 21% relative risk reduction in breast cancer mortality.[1] Others reported that this study resulted in an absolute risk reduction in breast cancer mortality of 0.0011%[2] and that 1,000 women would have to be screened to prevent one additional breast cancer death.[3] Thus, the effect of a treatment or risk factor may be considered to be large or small depending on the risk measure used to summarize the results, even when these measures are derived from the same underlying data.[4]

In this paper, we describe the most commonly used and misinterpreted measures of risk and discuss the proper interpretation of each using numerical examples. Specifically, we present relative and absolute risk reduction, attributable risk, and number needed to treat (NNT).

From the Department of Family Medicine (Dr Replogle) and the Department of Preventive Medicine (Dr Johnson), University of Mississippi Medical Center.

Relative Measures of Risk

Relative Risk

The DIAL trial reported the effectiveness of a telephone intervention among outpatients with stable chronic heart failure.[5] In this randomized controlled trial, intervention patients received standard cardiac services in addition to frequent telephone calls from a nurse experienced in the management of chronic heart failure. Control patients received only standard cardiac services. One primary endpoint was hospital admission for worsening heart failure.

From Table 1, the proportion of the control group who experienced at least one cardiovascular admission, termed the control event rate (CER), was 30.1%. The proportion of the intervention group who experienced at least one cardiovascular admission, termed the experimental event rate (EER), was 24.1%.

In cohort studies such as this, relative risk (RR) is often used to summarize the association between the intervention and the subsequent outcome. In the DIAL trial example, this resulted in an RR (RR=EER/CER) of 0.241/0.301=0.80. The RR=0.80 indicates that the

Table 1

DIAL Trial

		Event Cardiovascular Admission		Total
		Positive	Negative	
Exposure	Intervention Group	a 183 (24.1%)	b 577 (75.9%)	760
Non-exposure	Control Group	c 228 (30.1%)	d 530 (69.9%)	758

EER was 80% of the CER. If EER and CER had been equal, RR (EER/CER) would equal 1.0. Since the 95% confidence interval (CI) (0.678–0.946, $P<.05$) for the RR value did not include the value of 1, as expected if the true risks are equal, we can conclude that the difference in the event rates of the two groups is statistically significant.

Relative Risk Reduction

Risk difference between two groups is often used to summarize the magnitude of the treatment or exposure effect and can be expressed in either relative or absolute terms. Relative risk reduction (RRR) is defined as the difference in the two event rates expressed as a proportion of the event rate in the unexposed group, (CER– EER)/CER or simply (1–RR). As shown in Table 2, in the DIAL trial, RRR was 20% ([30.1%–24.1%]/30.1%) or (100%–80%). Therefore, we conclude that the cardiovascular admission rate in the intervention group was 20% lower than the cardiovascular admission rate in the control group.

Absolute Measures of Risk

Absolute Risk Reduction

Contrasted with RRR, absolute risk reduction (ARR) is defined as the absolute value of the arithmetic difference in the event rates of the two groups, ARR = |CER–EER|. When the event rate in the intervention group is greater than the event rate in the control group, the absolute measure of risk difference is sometimes referred to as absolute risk increase (ARI). In the DIAL trial, ARR was |30.1%–24.1%| = 6%, (1.5%–10.5%, $P<.05$). Since the 95% CI did not include the value of zero, as expected if the true risks are equal, we conclude that the 6% absolute difference in cardiovascular admission rates between the intervention group and the control group is statistically significant (Table 2).

Attributable Risk

ARR is sometimes referred to as attributable risk (AR) and is generally interpreted as the difference in morbidity or mortality in the intervention group relative to that experienced in the control group. From the DIAL trial data, AR is calculated as |CER–EER| = 0.06 or 6%. Accordingly, there were six fewer cardiovascular admissions per 100 patients who received the telephone

Table 2

Measures of Effect Size

	Event Rate: Cardiovascular Admission (Mean follow-up 16 months)		***Relative Risk Reduction (RRR)***	***Absolute Risk Reduction (ARR)***	***Number Needed to Treat (NNT)***
	Control Event Rate (CER)	*Experimental or Exposure Event Rate (EER)*	*|CER–EER|/CER or (1–RR)*	*|CER–EER|*	*1/ARR*
DIAL Trial	30.1%	24.1%	|30.1%–24.1%| / 30.1%=20%	|30.1%–24.1%| = 6%	1/0.06 = 16.66 or 17
Hypothetical trial with lower CER	5%	4%	|5%–4%| / 5% = 20%	|5%–4%| = 1%	1/0.01 = 100
Hypothetical trial with very low CER	1%	0.8%	|1%–0.8%| / 1% = 20%	|1%–0.8%| = 0.2%	1/0.2 = 500

CER—Control event rate or the risk of the event without treatment or exposure
EER—Experimental or exposure event rate or the risk of the event given treatment or exposure
Adapted from Straus SE, Richardson WS, Glasziou P, Haynes RB. *Evidence-based Medicine: How to Practice and Teach EBM*. Edinburgh: Elsevier; 2005.

intervention compared to 100 patients who received only standard cardiac services during the 16-month mean study period.

Baseline probability is the probability of the outcome (eg, a cardiovascular admission) in the absence of the intervention (eg, in the DIAL trial, the telephone intervention) or, equivalently, the CER. Under many conditions, RRR tends to be stable across a range of risk levels or baseline probabilities. That is, an intervention may produce approximately the same percent reduction or increase in the event rate (RRR) when applied to groups that are at high, medium, or low risk for the outcome. In these circumstances, RRR is deceptively attractive as it fails to account for the magnitude of disease risk in the absence of therapy (CER). When the CER is relatively low or high, RRR will overestimate or underestimate, respectively, the effect of a treatment.[6] For example, the intervention that produced a RRR of 20% in the DIAL trial may produce a similar relative risk reduction when applied to a population that has a much lower baseline probability (CER) of a cardiovascular admission.

Assume that the baseline probability of a cardiovascular admission in a population of younger patients with newly diagnosed heart failure is 5% as opposed to the CER of 30.1% in the DIAL trial. Among this younger, newly diagnosed group with a baseline probability of 5%, if the intervention again produced an RRR of 20%, the probability of an admission in the intervention group (EER) would be 20% less than the CER of 5%. The EER, therefore, would be 4%, and the ARR would be (5%–4%) 1% (Table 2). Hence, given a relative risk reduction of 20%, a CER=30.1% versus a CER=5% would result in an absolute risk reduction (ARR) of 6% versus 1%, respectively. If the CER were only 1%, a RRR of 20% would produce an ARR of 0.2%. Given the scenarios of Table 2 where RRR is identical across patient populations with varying levels of baseline risk, the absolute benefit (ARR) of the intervention decreases as the event rate in the control group decreases.

In the Swedish mammography trial, the RRR was in fact 21%. However, the risk of a breast cancer death during the study period without the benefit of the mammography screening (CER) was 0.005%. The 21% RRR implies the event rate was reduced in the intervention group to 0.0039%.[2] Therefore, given the relatively low baseline event rate of breast cancer death without the mammography intervention (CER), the absolute risk reduction of breast cancer death was 0.005%–0.0039%=0.0011%.[2]

RRR is unaffected by the risk of an event without treatment (CER) and does not discriminate well between large and small treatment effects.[7] RRR will always be greater in magnitude than ARR, and treatment effects reported as RRR are perceived as more dramatic than when reported as ARR.[8,9] ARR does account for the CER and, in this context, it is a more useful measure of treatment effect than RRR.[7]

Number Needed to Treat

ARR is useful in describing the amount of benefit or harm derived from an intervention or exposure. However, these concepts can be difficult to apply to individual patients. Laupacis et al[6] proposed the number needed to treat (NNT), a measure perhaps easier to interpret and more intuitive than relative risk estimates.[7] NNT is simply the inverse of ARR (ie, NNT=1/ARR or NNT=1/ |CER–EER|) and represents the number of patients needed to treat for a duration equal to the study period to prevent one additional adverse event (NNT to benefit=NNT:B) or produce one additional adverse event (NNT to harm=NNT:H).

For example, in a hypothetical study of aspirin users and nonusers, a primary end point may be ischemic stroke, and a secondary end point may be a gastrointestinal bleed. We might find that aspirin users experience a decrease in the event rate of ischemic stroke but also experience an increase in the event rate of gastrointestinal bleeds. Consequently, for ischemic stroke, NNT:B would represent the number of patients needed to treat with aspirin therapy to prevent one additional ischemic stroke whereas for gastrointestinal bleed, NNT:H represents the number to be treated with aspirin to produce one additional gastrointestinal bleed.

For the DIAL trial, we calculated ARR=6%, and NNT:B is calculated as |1/6%| 17 (9.6–65.0, $P<.05$) (Table 2). As a result, 17 chronic heart failure outpatients must be treated by standard cardiac care plus the telephone intervention (versus standard cardiac care alone) for a period equal to the mean DIAL trial follow-up period (16 months) to prevent one additional cardiovascular admission.

Earlier we noted that ARR, unlike RRR, preserves the baseline event probability and calculated ARR for a hypothetical group of younger, newly diagnosed heart failure patients to be 1% or 0.01. For this population, the NNT:B is 1/0.01=100. For both the population of chronic heart failure outpatients (DIAL trial) and the hypothetical population of younger patients with newly diagnosed heart failure, RRR=20%, suggesting that the intervention is equally efficacious in both patient populations. However, NNT indicates that when the baseline probability of a cardiovascular admission is 5% versus 30.1%, 100 patients versus 17 patients, respectively, will have to be treated to prevent one additional admission. Given a RRR of 20% and a CER of 1%, NNT becomes 500 (Table 2). Therefore, NNT as a measure of treatment effect incorporates both the relative reduction of the event rate and the risk of the event without treatment (CER).[6] For a given RRR, NNT will increase as the CER decreases. Thus, NNT has more obvious implications for clinical decision making

than risk estimates expressed in relative terms in that it estimates the therapeutic effort needed to prevent one additional event.

Summary

RRR is not affected by variations in baseline event rates whereas the expected absolute benefit (ARR) of treatment diminishes with the CER. As such, ARR is generally considered a more useful measure of treatment effect than RRR. NNT incorporates both relative risk reduction and the event rate without treatment and is useful in determining the estimated therapeutic effort needed to prevent one additional event.[7] When interpreting relative measures of risk, RR and RRR, one must consider the risk of the event without the benefit of treatment (CER).

Corresponding Author: Address correspondence to Dr Replogle, University of Mississippi Medical Center, Department of Family Medicine, 2500 North State Street, Jackson, MS 39216. 601-984-5415. Fax: 601-984-6835. wreplogle@familymed.umsmed.edu

References

1. Nyustrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002;359(9310):909-9.
2. Gulbrandsen P. Update on effects of screening mammography. Lancet 2002;360(9329):339.
3. Bonneux L. Update on effects of screening mammography. Lancet 2002;360(9329):338.
4. Barratt A, Wyer PC, Hatala R, et al. Tips for learning and teaching evidence-based medicine: introduction to the series. CMAJ 2004;171(4):347-8.
5. GESICA Investigators. Randomised trial of telephone intervention in chronic heart failure: DIAL trial. BMJ 2005;331(7514):425-9.
6. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988;318(26):1728-33.
7. Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Elsevier, 2005.
8. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med 1992;117(11):916-21.
9. Fahey T, Griffiths S, Peteres TJ. Evidence-based purchasing: understanding results of clinical trials and systematic reviews. BMJ 1995;311(7012):1056-9.