2B2

THERAPY AND UNDERSTANDING THE RESULTS

Measures of Association

Roman Jaeschke, Gordon Guyatt, Alexandra Barratt, Stephen Walter, Deborah Cook, Finlay McAlister, and John Attia

Sharon Straus also made substantive contributions to this section

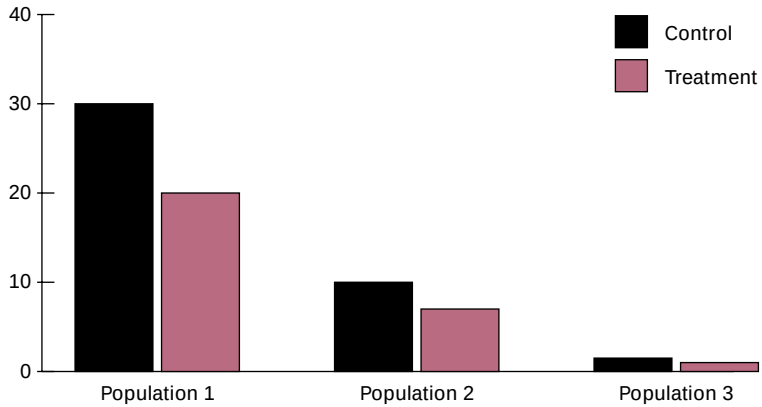
IN THIS SECTION

When clinicians consider the results of clinical trials, they are interested in the association between a treatment and an outcome. The study under consideration may or may not demonstrate an association between treatment and outcome; for example, it may or may not demonstrate a decrease in the risk of adverse events in patients receiving experimental treatment.

The focus of this section is on yes/no or dichotomous outcomes like death, stroke, or myocardial infarction. In their presentation of the results of studies addressing intervention effects on dichotomous outcomes, authors generally include the proportion of patients in each group who suffered an adverse event. As depicted in Figure 2B2-3, consider three different treatments that reduce mortality administered to three different populations. The first treatment, administered to a population with a 30% risk of dying, reduces the risk to 20%. The second treatment, administered to a population with a 10% risk of dying, reduces the risk to 6.7%. The third treatment reduces the risk of dying from 1% to 0.67%.

FIGURE 2B2-3

Constant Relative Risk With Varying Risk Differences



Although all three treatments reduce the risk of dying by a third, this piece of information is not adequate to fully capture the impact of treatment. Expressing the strength of the association as a *relative risk (RR)*, a *relative risk reduction (RRR)*, an *absolute risk reduction (ARR)* or *risk difference (RD)*, an *odds ratio (OR)*, or a *number needed to treat (NNT)* or *number need to harm (NNH)* conveys a variety of different information.

DICHOTOMOUS AND CONTINUOUS OUTCOMES

A study's primary analysis often is concerned with the proportion of patients who suffer a particular target outcome, endpoint, or event in the treatment and control groups. This is true whenever the outcome captures the presence or absence of negative events like stroke, myocardial infarction, cancer recurrence, or death. It is also true for positive events like ulcer healing or resolution of symptoms. Even if the outcome is not one of these dichotomous variables, investigators sometimes elect to present the results as if this were the case. For example, investigators may present endpoints such as the duration of exercise time before the development of chest pain, the number of episodes of angina per month, the change in pulmonary function, or the number of visits to the emergency department as the mean values in the two groups. Alternatively, they may transform these variables into dichotomous data by specifying a threshold or degree of change that constitutes an important improvement or deterioration and then examine the proportion of patients above and below this threshold. For example, in a study of the use of forced expiratory volume in 1 second (FEV_1) in the assessment of the efficacy of oral corticosteroids in patients with chronic stable airflow limitation, investigators defined an event as an improvement in FEV_1 over baseline of more than 20%.[1] In another study in patients with chronic lung disease, investigators examined the difference in the proportion of patients who achieved an important improvement in health-related quality of life.[2] The investigators' choice of the magnitude of change required to designate an improvement as "important" can affect the apparent effectiveness of the treatment (although less so for odds ratios, discussed later in this section, than for the other measures of association).

THE 2 X 2 TABLE

Table 2B2-3 depicts a 2 x 2 table that captures the information for a dichotomous outcome of a clinical trial. For instance, in a randomized trial, investigators compared mortality rates in patients with bleeding esophageal varices controlled either by endoscopic ligation or endoscopic sclerotherapy.[3] After a mean follow-up of 10 months, 18 of 64 participants assigned to ligation died, as did 29 of 65 patients assigned to sclerotherapy (Table 2B2-4).

TABLE 2B2-3

The 2 x 2 Table

		Outcome	
		Yes	No
Exposure	Yes	a	b
	No	c	d

Relative Risk (RR) $= \dfrac{a/(a + b)}{c/(c + d)}$

Relative Risk Reduction (RRR) $= \dfrac{c/(c + d) - a/(a + b)}{c/(c + d)}$

Absolute Risk Reduction (ARR) $= \dfrac{c}{c + d} - \dfrac{a}{a + b}$

Number Needed to Treat (NNT) $= \dfrac{1}{ARR}$

Odds Ratio (OR) $= \dfrac{a/b}{c/d} = \dfrac{ad}{cb}$

TABLE 2B2-4

Results From a Randomized Trial of Endoscopic Sclerotherapy as Compared With Endoscopic Ligation for Bleeding Esophageal Varices*

		Outcome		
		Death	**Survival**	**Total**
Exposure	Ligation	18	46	64
	Sclerotherapy	29	36	65

Relative Risk (RR) = 0.63

Relative Risk Reduction (RRR) = 0.37

Absolute Risk Reduction (ARR) = 0.165

Number Needed to Treat (NNT) = 6

Odds Ratio (OR) = 0.49

* Data from reference 3.

THE ABSOLUTE RISK

The simplest measure of association to understand is the *absolute risk*. The absolute risk of dying in the ligation group is 28% (18/64, or $a/a+b$), and the absolute risk of dying in the sclerotherapy group is 45% (29/65, or $c/c+d$). We often refer to the risk of the adverse outcome in the control group as the *baseline risk* or *control event rate*.

THE ABSOLUTE RISK REDUCTION

One can relate these two absolute risks by calculating the difference between them. We refer to this difference as the *absolute risk reduction* (ARR) or the risk difference (RD). Algebraically, the formula for calculating the ARR or RD is $[a/(a+c)]-[b/(b+d)]$ (see Table 2B2-3). This measure of effect tells us what proportion of patients are spared the adverse outcome if they receive the experimental therapy, rather than the control therapy. In our example, the ARR is 0.446 – 0.281, or 0.165 (ie, an ARR of 16.5%).

THE RELATIVE RISK

Another way to relate the absolute risks in the two groups is to take the ratio of the two; this is called the *relative risk* or *risk ratio (RR)*. The RR tells us the proportion of the original risk (in this case, the risk of death with sclerotherapy) that is still present when patients receive the experimental treatment (in this case, ligation). Looking at our 2 x 2 tables, the formula for this calculation is $[a/(a+c)]/[b/(b+d)]$ (see Table 2B2-3 and the Appendix). In our example, the RR of dying after receiving initial ligation versus sclerotherapy is 18/64 (the risk in the ligation group) divided by 29/65 (the risk in the sclerotherapy group), or 63%. In other words, we would say the risk of death with ligation is about two thirds of that with sclerotherapy.

THE RELATIVE RISK REDUCTION

Another measure used when assessing effectiveness of treatment is the *relative risk reduction (RRR)*. An estimate of the proportion of baseline risk that is removed by the therapy, it is calculated by dividing the absolute risk reduction by the absolute risk in the control group (see Table 2B2-3 and the Appendix). In our bleeding varices example, the RRR is 16.5% (the ARR) divided by 44.6% (the risk in the sclerotherapy group), or 0.37. One may also derive the RRR as (1.0 – RR). In the example, we have RRR = 1.0 – 0.63 = 0.37, or 37%. Using nontechnical language, we would say that ligation decreases the relative risk of death by 37% compared to sclerotherapy.

THE ODDS RATIO

Instead of looking at the risk of an event, we could estimate the odds of having vs not having an event. You might be most familiar with odds in the context of sporting events, when bookies or newspaper commentators quote the chances for and against a horse, a boxer, or a tennis player winning a particular event. When used in medicine, the *odds ratio (OR)* represents the proportion of patients with the target event divided by the proportion without the target event. In most instances in medical investigation, odds and risks are approximately equal—so much so that many authors calculate relative odds and then report the results as if they had calculated relative risks. The following discussion will inform clinicians who wish to understand what an odds ratio is and who wish to be alert to those circumstances when treating an odds ratio as a relative risk will be misleading.

To provide a numerical example: If 1/5 of the patients in a study suffer a stroke, the odds of their having a stroke is (1/5)/(4/5) or 0.20/0.80, or 0.25. It is easy to see that because the denominator is the same in both the top and bottom expressions, it is canceled out, leaving the number of patients with the event (1) divided by the number of patients without the event (4). To convert from odds to risk, divide the odds by 1 plus the odds. For instance, if the odds of a poor surgical outcome is 0.5, the risk is 0.5/1 + 0.5, or 0.33. Table 2B2-5 presents the relationship between risk and odds. Note that the greater the magnitude of the risk, the greater is the divergence between the risk and odds.

TABLE 2B2-5

Risks and Odds*

Risk	Odds
80%	4
60%	1.5
50%	1.0
40%	0.67
33%	0.50
25%	0.33
20%	0.25
10%	0.11
5%	0.053

* Risks are equal to odds / 1 + odds. Odds are equal to risk / 1 – risk.

In our example, the odds of dying in the ligation group are 18 (death) vs 46 (survival), or 18 to 46 or 18/46 (*a/b*), and the odds of dying in the sclerotherapy group are 29 to 36 (*c/d*). The formula for the ratio of these odds is (*a/c)/(b/d*) (see Table 2B2-3); in our example, this yields (18/46)/(29/36), or 0.49. If one were formulating a terminology parallel to risk (where we call a ratio of risks a relative risk), one would call the ratio of odds a *relative odds*. Epidemiologists, who have been averse to simplifying parallel terminology, have chosen *relative risk* as the preferred term for a ratio of risks and *odds ratio* for a ratio of odds.

Clinicians have a good intuitive understanding of risk and even of a ratio of risks. Gamblers have a good intuitive understanding of odds. No one (with the possible exception of certain statisticians) intuitively understands a ratio of odds.[4,5] Nevertheless, until recently the OR has been the predominant measure of association.[6] The reason is that the OR has a statistical advantage in that it is essentially independent of the arbitrary choice between a comparison of the risks of an event (such as death) or the corresponding nonevent (such as survival), which is not true of the RR.[7]

As clinicians, we would like to be able to substitute the RR—which we intuitively understand—for the OR—which we do not understand. Looking back at our 2 x 2 table (see Table 2B2-3), we see that the validity of this substitution requires that [*a/(a+b)]/[c/(c+d)*]—the RR—be more or less equal to (*a/b)/(c/d*)—the OR. For this to be the case, *a* must be much less than *b*, and *c* much less than *d*; in other words, the outcome must occur infrequently in both the treatment and the control groups. As we have noted, Table 2B2-5 demonstrates that as the risk falls, the odds and risk come closer together. For low event rates, common in most randomized trials, the OR and RR are very close. The RR and OR will also be closer together when the magnitude of the treatment effect is small (that is, OR and RR are close to 1.0) than when the treatment effect is large.

When event rates are high and effect sizes are large, there are ways of converting the OR to RR.[8,9] Fortunately, clinicians will rarely need to consult such tables. To see why, consider a meta-analysis of ligation vs sclerotherapy for esophageal varices,[10] which demonstrated a rebleeding rate of 47% with sclerotherapy—as high an event rate as one is likely to find in most trials. The OR associated with treatment with ligation was 0.52—a large effect. Despite the high event rate and large effect, the RR is 0.60, which is not very different from the OR. The two are close enough—and this is the crucial point—that choosing one measure or the other is unlikely to have an important influence on treatment decisions.

RELATIVE RISK AND ODDS RATIO VS ABSOLUTE RELATIVE RISK: WHY THE FUSS?

Having decided that distinguishing between OR and RR will seldom have major importance, introducing hypothetical changes to the 2 x 2 table (see Table 2B2-4) shows us why we must pay much more attention to distinguishing between the OR

and RR vs the ARR. Let us assume that the number of patients dying decreased by approximately 50% in both groups. We now have nine deaths among 64 patients in ligation group and 14 deaths among 65 patients in the sclerotherapy group. The risk of death in the ligation group decreases from 28% to 14%, and in the sclerotherapy group, it decreases from 44.6% to 22.3%. The RR becomes 14/22.3 or 0.63, the same as before. The OR becomes (9/55)/(14/51) or 0.60, moderately different from 0.49 and closer to the RR. The absolute risk reduction decreases quite dramatically from 16.5% to approximately 8%. Thus, the decrease in the proportion of those dying in both groups by a factor of two leaves the RR unchanged, results in a moderate increase in the OR, and reduces the ARR by a factor of 2. This (see Figure 2B2-3) shows how the same RR can be associated with quite different ARRs—and that although the RR does not reflect changes in the risk of an adverse event without treatment (or, as in this case, with the inferior treatment), the ARR can change markedly with changes in this baseline risk.

Thus, a RR of 0.67 may represent both a situation in which a treatment reduces the risk of dying from 1% to 0.67%, or from 30% to 20% (see Figure 2B2-3). Assume that the frequency of severe side effects associated with such a treatment were 10%—we might encounter this situation in offering chemotherapy to a patient with cancer, for instance. Under these circumstances we would probably not recommend the treatment to most patients if it reduced the probability of dying by 0.33% (from 1% to 0.67%), but we may well be willing to recommend this treatment if the probability of an adverse outcome drops from 30% to 20%. In the latter situation, 10 patients per 100 would benefit, whereas one would suffer adverse effects—a tradeoff that most would consider worthwhile.

The RRR behaves the same way as the RR and does not reflect the change in the underlying risk in the control population. In our example, the RRR will be of the same magnitude if the frequency of events decreases by approximately half in both groups: (22.3 – 14)/22.3, or 0.37.

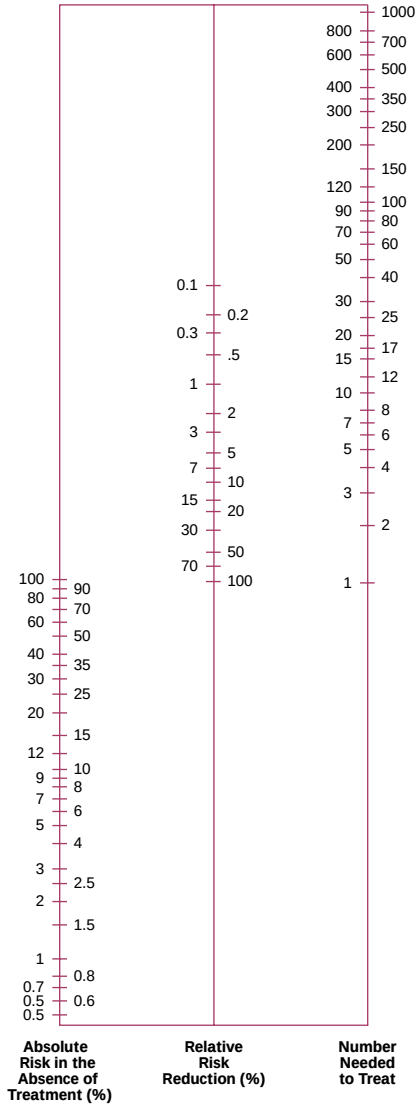
THE NUMBER NEEDED TO TREAT

One can also express the impact of treatment by the number of patients one would need to treat to prevent an adverse event, the *number needed to treat* (*NNT*).[11] Table 2B2-4 shows that the risk of dying in the ligation group is 28.1%, and in the sclerotherapy group, it is 44.6%. If these estimates are accurate, treating 100 patients with ligation rather than sclerotherapy will result in between 15 and 16 patients avoiding death (the ARR, the control event rate minus the intervention event rate). If treating 100 patients results in avoiding 16 events, how many patients do we need to treat to avoid one event? The answer, 100 divided by 16, or approximately 6 (that is, 100 divided by the risk difference expressed as a percentage), is the NNT. One can also arrive at this number by taking the reciprocal of the ARR expressed as a proportion; that is, one can calculate the NNT by the formula 1/ARR (see Table 2B2-3). You may see that both the NNT and the ARR

change with the difference in the underlying risk—which is not surprising, because the NNT is the reciprocal of the ARR. Given knowledge of the baseline risk and relative risk reduction, a nomogram presents a third way of arriving at the NNT (see Figure 2B2-4).[12]

FIGURE 2B2-4

Nomogram for Calculating the Number Needed to Treat



Reproduced with permission from the BMJ Publishing Group.

The NNT is inversely related to the proportion of patients in the control group who suffer an adverse event. If the risk of an adverse event doubles, we need treat only half as many patients to prevent an adverse event. If the risk decreases by a factor of 4, we will have to treat four times as many people. In our example, if the frequency of events (the baseline risk) decreases by a factor of 2 while the RRR remains constant, treating 100 patients with ligation would then result in avoiding eight events (22 – 14) and the NNT would double to 12.

The NNT is also inversely related to the RRR. A more effective treatment with twice the RRR will reduce the NNT by half. If the relative risk reduction with one treatment is only a quarter of that achieved by an alternative strategy, the NNT will be four times greater. Table 2B2-6 presents hypothetical data that illustrate these relationships.

TABLE 2B2-6

Relationship Between the Baseline Risk, the Relative Risk Reduction, and the Number Needed to Treat*

Control Event Rate	Intervention Event Rate	Relative Risk	Relative Risk Reduction	Risk Difference	NNT
0.02	0.01	50%	50%	0.01	100
0.4	0.2	50%	50%	0.2	5
0.04	0.02	50%	50%	0.02	50
0.04	0.03	75%	25%	0.01	100
0.4	0.3	75%	25%	0.1	10
0.01	0.005	50%	50%	0.005	200

* Relative risk is equal to the intervention event rate/control event rate; the relative risk reduction is equal to 1– relative risk; the risk difference is equal to control event rate – intervention event rate; the NNT is equal to 1 / risk difference.

Using ARR and its reciprocal, the NNT, incorporates the influence of the changing baseline risk. If all we know is the ARR or the NNT, however, we remain ignorant of the size of the baseline risk. For example, an ARR of 5% (and a corresponding NNT of 20) may represent reduction of the risk of death from 10% to 5% (a RRR of 50%) or from 50% to 45% (a RRR of 10%).

THE NUMBER NEEDED TO HARM

Clinicians can calculate the *number needed to harm* (NNH) in exactly the same way. If one expects that five of 100 patients will become fatigued when given a beta blocker, one will have to treat 20 patients to cause one to become tired, and the NNH is 20.

In this discussion we have not mentioned the problem that investigators may report odds ratios instead of relative risks. As we have mentioned, the best way of dealing with this situation when event rates are low is to assume the RR will be very close to the OR. The higher the risk, the less secure is the assumption. Tables 2B2-7 and 2B2-8 provide a guide for making an accurate estimate of the NNT and NNH when you know the patient's baseline risk and the investigator has provided only an odds ratio.

TABLE 2B2-7

Deriving the NNT From the Odds Ratio*

Control Event Rate	Therapeutic Intervention (OR)								
	0.5	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.05	41	46	52	59	69	83	104	139	209
0.1	21	24	27	31	36	43	54	73	110
0.2	11	13	14	17	20	24	30	40	61
0.3	8	9	10	12	14	18	22	30	46
0.4	7	8	9	10	12	15	19	26	40
0.5	6	7	8	9	11	14	18	25	38
0.7	6	7	9	10	13	16	20	28	44
0.9	12	15	18	22	27	34	46	64	101

* Adapted from reference 18.

The formula for determining the NNT is:

$$\text{NNT} = \frac{1 - CER(1 - OR)}{CER(1 - CER)(1 - OR)}$$

(CER = control event rate, OR = odds ratio)

TABLE 2B2-8

Deriving the NNH From the Odds Ratio*

Control Event Rate	Therapeutic Intervention (OR)								
	1.1	**1.2**	**1.3**	**1.4**	**1.5**	**2**	**2.5**	**3**	**3.5**
0.05	212	106	71	54	43	22	15	12	9
0.1	112	57	38	29	23	12	9	7	6
0.2	64	33	22	17	14	8	5	4	4
0.3	49	25	17	13	11	6	5	4	3
0.4	43	23	16	12	10	6	4	4	3
0.5	42	22	15	12	10	6	5	4	4
0.7	51	27	19	15	13	8	7	6	5
0.9	121	66	47	38	32	21	17	16	14

* Adapted from reference 18.

The formula for determining the NNH is:

$$\text{NNH} = \frac{CER(OR - 1) + 1}{CER(OR - 1)(1 - CER)}$$

(CER = control event rate, OR = odds ratio)

BACK TO THE 2 x 2 TABLE

Whichever way of expressing the magnitude of the treatment effect we choose, the 2 x 2 table reflects results at a given point in time. Therefore, our comments on RR, ARR, RRR, OR, and NNT or NNH must be qualified by imposing a time frame on them. For example, we have to say that using ligation rather than sclerotherapy resulted in absolute risk reduction of death of 17% and an NNT of 6 over a mean time of 10 months. The results could be different if the duration of observation was very short (if there was no time to develop an event) or very long (after all, if an outcome is death, after 100 years of follow-up, everybody will die).

CONFIDENCE INTERVALS

We have presented all of the measures of association of the treatment with ligation vs sclerotherapy as if they represented the true effect. The results of any experiment, however, represent only an estimate of the truth. The true effect of treatment may actually be somewhat greater—or less—than what we observed. The confidence interval tells us, within the bounds of plausibility, how much greater or smaller the true effect is likely to be (see Part 2B2, "Therapy and Understanding the Results, Confidence Intervals"). Statistical programs permit computation of confidence intervals for each of the measures of association we have discussed.

SURVIVAL DATA

As we pointed out, the analysis of a 2 x 2 table implies an examination of the data at a specific point in time. This analysis is satisfactory if we are looking for events that occur within relatively short periods of time and if all patients have the same duration of follow-up. In longer-term studies, however, we are interested not only in the total number of events, but in their timing as well. For instance, we may focus on whether therapy for patients with a uniformly fatal condition (severe congestive heart failure or unresectable lung cancer, for example) delays death.

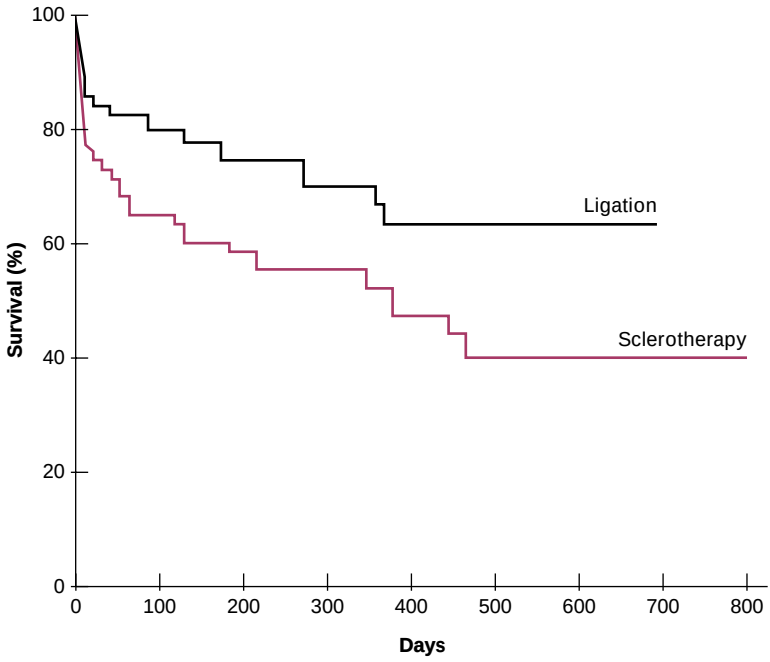
When the timing of events is important, investigators could present the results in the form of several 2 x 2 tables constructed at different points of time after the study began. For example, Table 2B2-4 represented the situation after a mean of 10 months of follow-up. Similar tables could be constructed describing the fate of all patients available for analysis after their enrollment in the trial for 1 week, 1 month, 3 months, or whatever time frame we choose to examine. The analysis of accumulated data that takes into account the timing of events is called *survival analysis*. Do not infer from the name, however, that the analysis is restricted to deaths; in fact, any dichotomous outcome will qualify.

The survival curve of a group of patients describes the status of patients at different time points after a defined starting point.[13] In Figure 2B2-5, we show the survival curve from the bleeding varices trial. Because the investigators followed approximately half of the patients for a longer time, the survival curve extends beyond the mean follow-up of 286 days. At some point, prediction becomes very imprecise because there are few patients available to estimate the probability of survival. Confidence intervals around the survival curves capture the precision of the estimate.

Even if the true relative risk, or relative risk reduction, is the same for each duration of follow-up, the play of chance will ensure that the point estimates differ. Ideally then, we would estimate the overall relative risk reduction by applying an average, weighted for the number of patients available, for the entire survival experience. Statistical methods allow just such an estimate. The weighted relative risk over the entire study is known as the *hazard ratio*.

FIGURE 2B2-5

Survival Curves for Ligation and Sclerotherapy



Reproduced with permission from the Massachusetts Medical Society.

Assuming the null hypothesis (ie, that there is no difference between two survival curves), we can generate a *P* value that informs us about the likelihood that chance explains the differences in results. Statistical techniques (most commonly, the *Cox regression model*) allow the results to be adjusted or corrected for differences in the two groups at baseline (see "Part 2B2, Therapy and Understanding the Results, Confidence Intervals"). If one group was older (and, thus, was at higher risk) or had less severe disease (and, thus, was at lower risk), the investigators might focus on an analysis that takes these differences into account. This, in effect, tells us what would have happened had the two groups had comparable risk factors for adverse outcome at the start of the trial.

Another way of reading survival curves is to plot the points at which a chosen percentage of the patients in each group have reached an endpoint. The difference between these points is a reflection of the delay in outcomes in the treatment group. For example, although ACE inhibitors may be associated with an up to 25% decrease in mortality in patients with postmyocardial infarction, this translates into an extra few months of life for patients in the treatment group, a result that may not appear as impressive.[14]

CASE-CONTROL STUDIES

Up to now, our examples have come from prospective randomized controlled trials. In these trials, we start with a group of patients who are exposed to an intervention and a group of patients who are not exposed to the intervention. The investigators follow the patients over time and record the frequency of events. The process is similar in observational studies termed *prospective cohort studies*, although in this study design the exposure or treatment is not controlled by the investigators. For randomized trials and prospective cohort studies we can calculate risks, absolute risk reductions, and relative risks.

In case-control studies, investigators choose or sample participants not according to whether they have been exposed to the treatment or risk factor, but on the basis of whether they have experienced a target outcome. Participants start the study with or without the event, rather than with or without the exposure or intervention. Investigators compare patients with the adverse outcome—be it stroke, myocardial infarction, or cancer—to controls who have not suffered the outcome. The usual question asked is if there are any factors that seem to be more commonly present in one of these groups than in the other group.

In one case-control study, investigators examined the question of whether sunbeds or sunlamps increase the risk of skin melanoma.[15] They identified 583 patients with melanoma and 608 controls. The control patients and the cases had similar distributions of age, sex, and residence. Table 2B2-9 presents the findings for the men who participated in this study.

TABLE 2B2-9

Results From a Case-Control Study Examining the Association of Cutaneous Melanoma and the Use of Sunbeds and Sunlamps*

	Exposure	Cases	Controls
Sunbeds or sunlamps	Yes	67	41
	No	210	242

* Data from reference 11.

If the information in Table 2B2-9 came from a prospective cohort study or randomized controlled trial, we could begin by calculating the risk of an event in the exposed and control groups. This would not make sense in the case-control study because the number of patients who did not have melanoma was chosen by the investigators. For calculation of relative risk, we need to know the population at risk, and a case-control study does not provide this information.

The OR provides the only sensible measure of association in a case-control study. One can ask whether the odds of having been exposed to sunbeds or sunlamps among people with melanoma were the same as the odds of exposure among the control patients. In the study, the odds of exposure were 67/210 in

the melanoma patients and 41/242 in the control patients. The OR is therefore (67/210)/(41/242), or 1.88 (95% CI, 1.20-2.98), suggesting an association between using sunbeds or sunlamps and developing melanoma. The fact that the confidence interval does not overlap or include 1.0 suggests that the association is unlikely to have resulted from chance.

Even if the association were not chance related, it does not necessarily mean that the sunbeds or sunlamps were the cause of melanoma. Potential explanations could include greater recollection of using these devices among people with melanoma (recall bias), longer sun exposure among these people, and different skin color; of these explanations, the investigators addressed many. To be confident that exposure to sunbeds or sunlamps was the cause of melanoma would require additional confirmatory studies.

WHICH MEASURE OF ASSOCIATION IS BEST?

As evidence-based practitioners, we must decide which measure of association deserves our focus. Does it matter? The answer is "yes." The same results, when presented in different ways, may lead to different treatment decisions.[16-20] For example, Forrow and colleagues[16] demonstrated that clinicians were less inclined to treat patients after presentation of trial results as the absolute change in the outcome compared with the relative change in the outcome. In a similar study, Naylor and colleagues[17] found that clinicians rated the effectiveness of an intervention lower when events were presented in absolute terms rather than using relative risk reduction. Moreover, effectiveness was rated lower when results were expressed in terms of NNT than when the same data were presented as relative or absolute risk reductions. The pharmaceutical industry's awareness of this phenomenon may be responsible for their propensity to present physicians with treatment-associated relative risk reductions.

Patients turn out to be as susceptible as clinicians to the mode in which results are communicated.[12, 21-23] In one study, when researchers presented patients with a hypothetical life-threatening illness, those patients were more likely to choose a treatment described in terms of relative risk reduction than in terms of the equivalent absolute risk reduction.[12]

Aware that they will perceive results differently depending on how they are presented, what are clinicians to do? We believe that the best option is to consider all of the data (either as a 2 x 2 table or as a survival analysis) and then consider both the relative and the absolute figures. As you examine the results, you will find that if you can calculate the ARR and its reciprocal, the NNT, in an individual patient, these will be most useful in deciding whether to institute treatment (see Part 2B3, "Therapy and Applying the Results, Example Numbers Needed to Treat"). The conscientious evidence-based practitioner will use all available information to formulate the likely risks and benefits for the individual patient (see Part 2B3, "Therapy, Applying Results to Individual Patients").

References

1. Callahan CM, Dittus RS, Katz BP. Oral corticosteroid therapy for patients with stable chronic obstructive pulmonary disease: a meta-analysis. *Ann Intern Med*. 1991;114:216-223.
2. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ*. 1998;316:690-693.
3. Stiegmann GV, Goff JS, Michaletz-Onody PA, et al. Endoscopic sclerotherapy as compared with endoscopic ligation for bleeding esophageal varices. *N Engl J Med*. 1992;326:1527-1532.
4. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol*. 1994;47:881-889.
5. Sackett DL. Down with odds ratios! *Evid Based Med.* 1996;1:164-166.
6. Laird NM, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care*. 1990;6:5-30.
7. Walter SD. Choice of effect measure for epidemiological data. *J Clin Epidemiol*. 2000;53:931-939.
8. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ*. 1998;316:989-991.
9. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. 1998;280:1690-1691.
10. Laine L, Cook D. Endoscopic ligation compared with sclerotherapy for treatment of esophageal variceal bleeding: a meta-analysis. *Ann Intern Med*. 1995;123:280-287.
11. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318:1728-1733.
12. Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. *BMJ*. 1996;312:426-429.
13. Coldman AJ, Elwood JM. Examining survival data. *CMAJ*. 1979;121:1065-1068, 1071.
14. Tan LB, Murphy R. Shifts in mortality curves: saving or extending lives? *Lancet*. 1999;354:1378-1381.
15. Walter SD, Marrett LD, From L, Hertzman C, Shannon HS, Roy P. The association of cutaneous malignant melanoma with the use of sunbeds and sunlamps. *Am J Epidemiol*. 1990;131:232-243.
16. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med*. 1992;92:121-124.
17. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med*. 1992;117:916-921.

18. Hux JE, Levinton CM, Naylor CD. Prescribing propensity: influence of life-expectancy gains and drug costs. *J Gen Intern Med*. 1994;9:195-201.
19. Redelmeier DA, Tversky A. Discrepancy between medical decisions for individual patients and for groups. *N Engl J Med*. 1990;322:1162-1164.
20. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet*. 1994;343:1209-1211.
21. Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. *J Gen Intern Med*. 1993;8:543-548.
22. McNeil BJ, Pauker SG, Sox HC Jr, Tversky A. On the elicitation of preferences for alternative therapies. *N Engl J Med*. 1982;306:1259-1262.
23. Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making*. 1995;15:152-157.