

Statistical Thinking Part 2: Relative Risk, Absolute Risk, and Number Needed to Treat

Robert A. Calder, MD, MS; Jayshil J. Patel, MD

Clearly communicating the risks and benefits of (and alternatives to) a treatment to patients is one of the most important tasks of clinicians. The data informing the task of communication can take many forms. For example, small differences in risks can appear very impressive and provide a false perception of benefit. Therefore, clinicians should have the skills to critically evaluate the data presented within research articles.

When reading studies, clinicians will invariably come across terms that describe effect size, such as relative risk, relative risk reduction, absolute risk difference, odds ratio, and hazard ratio (Table). Both patients and clinicians may be confused by the differences between percentage decreases in these terms. In part 2 of this limited series on statistical thinking, we present a story of risk that sets the stage for defining statistical terms that describe effect size and the number needed to treat, an intuitive term that helps clinicians and patients better express absolute effect size.

• • •

Author Affiliations: Medical College of Wisconsin, Milwaukee, Wisconsin (Calder); Division of Pulmonary and Critical Care Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin (Patel).

Corresponding Author: Robert A. Calder MD, Adjunct Assistant Professor, Medical College of Wisconsin, Milwaukee, WI; email rcalder@mcw.edu.

A Story About Risk

Suppose one of your favorite patients, Mrs Smith, a 70-year-old retired algebra teacher, asks you to prescribe a new medicine she just read about in an online newspaper. The article reported a 50% reduction in heart attacks with this medicine. Mrs Smith noted that this result came from a large study of over 15 000 patients, and she astutely noted that she had many of the personal characteristics of the people in this 2-year study and, therefore, the results should apply to her. Furthermore, the authors described the results as “highly statistically significant,” and the study was published in a prestigious medical journal. Based on this information alone, would you prescribe the new medicine for Mrs Smith?

Having read the article, suppose you respond to Mrs Smith by stating that “there is a 99% chance that you will not benefit from this new treatment in the next 2 years” (reflecting the length of the study). Stated differently, Mrs Smith has a 1% chance of benefiting from this new treatment in the next 2 years. How could a “highly statistically significant” 50% relative risk reduction (RRR) benefit only 1 person in 100 (1%) over 2 years?

When Mrs Smith requested this new treatment, she thought her risk of a heart attack was going to be reduced in half, eg, from a 100% chance to a 50% chance. Indeed, that is a 50% RRR. However, let’s assume she was more optimistic and gauged her risk of heart attack to be reduced from 40% to 20%, which would also

represent a 50% RRR. In reality, suppose that in the study she referenced the absolute risk of heart attack was reduced from 2% (in the placebo group) to 1% (in the treatment group). That also represents a 50% RRR. However, the absolute risk difference (ARD) was only 1% in this 2-year study. Given a 1% decrease in absolute risk, 100 people like Mrs Smith would have to be treated over a 2-year period to prevent, on average, 1 heart attack. Her next question is likely to be, “what will this new medicine cost?” “I may not be the one who benefits!” Indeed, chances are she will not benefit, and we haven’t even considered the possible side-effects of or alternatives to this new treatment.

Stories such as the one presented here are not uncommon.¹ Studies presented in the lay media report treatment “x” reduces the risk of outcome “y” without reporting the probability of the outcome or precise statistics that describe the effect size, which creates a perception of benefit for lay individuals and confusion for clinicians. A fundamental understanding of statistical terms to describe effect size (relative risk, RRR, ARD, and number needed to treat) is paramount for understanding the results of a study and communicating with patients. In the “Critical Thinking in Medicine” thread within the Fusion Curriculum at the Medical College of Wisconsin, we teach medical students how to have a conversation with their patients by first explaining these terms using practical, story-like formats and analogies.

Relative Risk and Relative Risk Reduction

Relative risk (RR) is the ratio of 2 probabilities. For example, suppose 1 group of people has a 10% risk of a heart attack in the next 5 years and another group has a 5% risk. The RR of a heart attack for the second group, compared to the first, is 5% divided by 10% or 0.5. RRR is the proportional amount that a risk is decreased in 1 group versus another. Numerically, it is the control group event rate (CER) minus the experimental group event rate (EER) divided by the CER. In the example above, the RRR is calculated $(10\% - 5\%) / 10\%$ or 50%. RR is a very useful concept, especially for identifying risk factors for disease. A problem occurs, however, if RR and RRR are presented without their corresponding absolute risks. Another problem arises when the benefits of some treatment are presented in terms of RRRs and the risks of the treatment are presented as absolute risks.¹ In this way, the same thing can be stated in very different ways. For example, in the study that Mrs Smith read, the authors stated that the treatment caused a 50% reduction in heart attacks (which is actually the RRR), but suppose they stated that only 1% of the subjects in the study experienced an adverse event. In that case, the same proportion of people avoided a heart attack as experienced an adverse event (1% in each case).

To summarize, the RR is a ratio of the probability of an event occurring in a treatment versus a control population, and the RRR is the proportional amount that a risk is decreased in 1 group compared to another.

Absolute Risk

Absolute risk is a probability—and a very complex topic that is the subject of a future article. Probability is generally interpreted in 2 ways: as a long-run frequency (the “frequentist” view) or as a degree of belief that can be modified with additional data (the “Bayesian view”).

When viewed as a long-run frequency, probability is the proportion of times that some event occurs over many repetitions of some process, carried out under similar conditions. For example, if a coin lands on heads 50 times out of 100 flips, we would say that there is a 50% prob-

| Statistical Term | Definition | Calculation |
|---------------------------------------|--|--|
| Absolute Risk (AR) | The probability of an event occurring in a population | Number of events in a population divided by the total number of individuals at risk for the event in a population |
| Relative Risk (RR) | The ratio of the risk of an event between 2 groups | The ratio of 2 probabilities, for example, experimental group event rate (EER) divided by control group event rate (CER) |
| Relative Risk Reduction (RRR) | The proportional amount that a risk is decreased in 1 group compared to another | Control group event rate (CER) minus experimental group event rate (EER) divided by CER |
| Absolute Risk Difference (ARD) | The measure of the absolute effect size of an intervention | Control group event rate (CER) minus experimental group event rate (EER) |
| Number Needed to Treat/Harm (NNT/NNH) | Measures the effectiveness of an intervention by identifying the number of, patients, on average, that would need to be treated with an intervention for patient to benefit (NNT) or be harmed (NNH) | The NNT is the reciprocal of the ARR (1/ARR) and the NNH is the reciprocal of the ARI (1/ARI) |
| Odds | The probability of an event compared to the probability of it not occurring | Probability of an event (P) divided by the probability of it not occurring (1 - P) |
| Odds Ratio (OR) | The odds (likelihood) of an event occurring in 1 group (eg, exposed) compared to another group (eg, nonexposed) | Odds of an event occurring in an exposed group ($P_{\text{exposed}} / 1 - P_{\text{exposed}}$) divided by the odds of it occurring in the nonexposed group ($P_{\text{nonexposed}} / 1 - P_{\text{nonexposed}}$) |
| Hazard Ratio (HR) | A measure of effect in a time-to-event survival analysis | The hazard rate in the treatment group divided by the hazard rate in the control group |

Abbreviations: ARR, absolute risk reduction; ARI, absolute risk increase.

ability of the coin landing on heads. There are a few problems here, however. What number (of flips) constitutes “many repetitions?” Is 100 enough? How about 500? Another problem with this definition is “similar conditions.” What does that mean exactly? If all conditions, such as which side of the coin is facing up when flipped, exactly how vigorously the coin was flipped, the precise wind conditions, etc, were exactly the same, we could predict with certainty whether it would land heads or tails. Therefore, a coin flip appears to be random because we cannot measure or control the many variables that determine whether it lands heads or tails.

Another problem with the frequentist view is that many important processes do not occur “many times under similar conditions.” Your favorite football team is only going to play this year’s season once, not 100 times under similar conditions. Accordingly, what does it mean to state that there is a 10% chance of your team playing in the championship game?

To address some of these concerns with the frequentist view, another conception of probability has been advanced, referred to as the

“Bayesian view,” and named after the English clergyman, Thomas Bayes, who studied probability in the late 18th century.² In the Bayesian view, probability is a degree of belief that is modifiable with additional data. For example, before the start of the football season, we may feel that our team has a 10% chance of getting to the championship game, based on factors such as the results of the previous year and the current team makeup. As the year progresses and the team wins and loses games, we will probably revise our estimation of how likely the team will be in the championship game based on the new data. The revision of probability based on new data is the heart of the Bayesian view, which is especially useful when determining whether a patient has some disease. After taking a careful history, you develop some idea of the likelihood of a particular disease and then, after doing a physical exam and perhaps getting various lab tests, you revise your estimate of the likelihood of disease based on this additional data. In fact, our brain operates under a Bayesian framework on a day-to-day basis. For example, before the school

year begins, you have some idea of how well you will do in a particular course. As the year progresses and you see the results of various exams and quizzes, you then revise your impression (and perhaps your study habits) of how well you'll do.

Most statistical tests in the medical literature today are based on the frequentist view of probability. In part, this is because our first conception of probability was based on predicting games of chance. When outcomes are equally likely, the frequentist view works well. Also, the frequentist view is generally computationally much easier to understand and use. However, in time, with improvements in computer technology and artificial intelligence, the Bayesian approach is becoming more prevalent. We will rely heavily on the Bayesian approach in the next article in this series regarding interpreting laboratory tests.

Absolute risk, or probability, is a proportion, and a proportion is different than a rate. A rate has time in the denominator. For example, miles per hour or heart attacks per 100 patient-years are rates. Unfortunately in medicine, many terms that are really proportions are called "rates." For example, the "attack rate" is the number of patients who contract a given disease out of the total population at risk. Obviously, this is a proportion and not a rate, since it is a number between 0 and 1 and time is not in the denominator. It is important to keep the distinction between rates and proportions in mind because a ratio of rates is different from a ratio of proportions (probabilities). A proportion ranges from 0 to 1, whereas a rate ranges from 0 to infinity. So, a ratio of 2 proportions is different from a ratio of 2 rates, unless the rates and proportions are very small (as discussed below).

To summarize, the absolute risk is a probability—or the number times that some event occurs over many repetitions of some process—carried out under similar conditions and is measured as a proportion and, therefore, ranges between 0 and 1 (or 0% and 100%).

Number Needed to Treat and Absolute Risk Difference

In addition to RR and RRR, the effect of a therapy also can be expressed by the number

of patients needed to treat (NNT) to prevent some event (or "cause" some good event). Conversely, the number needed to harm (NNH) is the number of people who would have to be treated over some time period to cause 1 "bad" event. This intuitive concept first appeared in medical literature in 1988,³ which is very surprising given the simplicity of this measure.

The ARD is the difference in risk of some specific outcome between control and experimental groups. When the experimental group experiences greater harm compared to the control, the ARD is also known as the absolute risk increase (ARI). When the experimental group receives more benefit compared to the control group, the ARD is also known as the absolute risk reduction (ARR).

The NNT and NNH (NNT/NNH) is calculated as the reciprocal of the ARD (1/ARD) in the treatment groups, and the calculation always implies a certain follow-up time. In the example above with absolute risks of 10% and 5% heart attacks, the ARD is 5%. (This is also referred to as the "attributable risk," but we see no need to introduce extra terminology when teaching this concept for the first time.) The reciprocal of 5% or 0.05 is 1/0.05 (or 100%/5%), which equals 20. When calculating the NNT/NNH, it is crucial to state the time interval involved; after all, in the long run we all experience the same fate so we must state a time interval reflecting the study involved.

To make this concept clearer, suppose we have a treatment that cures everyone who receives it, and those who don't receive it all die. In that case, the risk difference is 100%, and 100%/100% equals 1. We would only have to treat 1 person to cure the disease. Moreover, if we have a treatment that cures 50% (and everyone not receiving the treatment dies), we would have to treat 2 people, on average, to cure 1 (100%/50%=2). Similarly, if we have a treatment that cures 25%, we would need to treat 4, on average, to cure 1 person (100%/25%=4).

Another way to remember how to calculate NNT/NNH is with a basketball analogy. If my free throw percentage is 50%, on average, I am going to have to go the free throw line twice to make 1 free throw. If my percentage is 25%,

I am going to have to go the line 4 times, on average, to make 1 free throw.

To summarize, the ARD is the difference in risk of some specific outcome between control and experimental groups, and the NNT/NNH is calculated as the reciprocal of the ARD (1/ARD) and is a practical way to communicate the risks and benefits of an intervention.

Number Needed to Treat and Baseline Risk

Assuming a constant RRR, the NNT is inversely proportional to the baseline risk (the number of adverse events in the control group). As the baseline risk increases, the NNT is reduced (implying fewer patients would need to be treated with the therapy for 1 patient to benefit). Consider an example of patients with coronary heart disease (CHD). Suppose we have a group of CHD patients who have a 10-year risk of another heart attack of 20% (the absolute risk). If we provide a lipid lowering treatment to this group that decreases their risk by 30% at 10 years (the RRR), we will have decreased their absolute risk of a heart attack from 20% to 14% (20% x 0.3 = 6%; where 20% is the absolute risk and 0.3 is the RRR), and the absolute risk decreases from 20% to 14% (20% - 6%, or an ARR of 6%). The reciprocal of this 6% ARR is about 17 (100/6 = 16.67). In this high-risk population, 17 patients would need to be treated for 10 years (the risk period in this example) to prevent, on average, 1 heart attack.

Now suppose we have another group of people who have a 10-year risk of a heart attack of 1% (the absolute risk). If we provide them the same treatment that decreases their risk by 30% (the RRR), we would reduce their absolute risk to 0.7% (1% x 0.3 = 0.3%; where 1% is the absolute risk and 0.3 is the RRR of the treatment; 1% - 0.3% = 0.7%) and an ARR of 0.3% (1% - 0.7%). Therefore, the reciprocal of 0.3% ARR is 333 (100/0.3 = 333). In this low-risk population, we would have to treat 333 people for 10 years to prevent, on average, 1 heart attack.

To summarize, when the baseline risk is high, the NNT is low. When the baseline risk is low, the NNT is high because we are taking the reciprocals of absolute risk differences.

Additional Risk Ratios

Odds Ratio

Odds is less intuitive than probability. Recall that the probability measures how likely an event will occur divided by the total number of possible outcomes. Consider the following example: if some event occurs in 50% of a population, then the probability of it occurring is $50\%/100\%=50\%$. The odds FOR an event occurring is the ratio of the probability of the event occurring divided by the probability of it not occurring. For example, if the probability of some event is 50%, the probability that it will occur is 50% and the probability that it will not occur is also 50%. The ratio of these is $50\%/50\%$, which reduces to 1:1 or even odds. If some event has a 75% chance of occurring, there is a 25% chance that it will not occur, and this ratio— $75\%/25\%$, which reduces to 3:1—is the equivalent odds to a 75% probability. When the probability of some event is 5%, the probability that it will not occur is 95%, giving odds of 5:95 or 1:19. Notice when the event is relatively rare—in this case 5%, the probability and the odds are quite similar.

The odds ratio (OR) is a ratio of 2 odds, just as the RR is the ratio of 2 probabilities. When odds and probability are fairly similar—as they are when the risks are low, such as above—the OR and the RR are very similar. However, when the probability of an event is much higher, the OR and the RR can be quite different. For example, if the probability of an event is 75%, as noted above, the odds are 3:1. If the probability of this event in another group is 50%, then the RR for the event in 1 group versus the other is $75\%/50\%$ or 1.5. For this same comparison, the OR would be 3:1/1:1 or 3. Thus, in this example, the OR is twice the RR. Only when the risks are low (under about 10%) are the odds and probability reasonably comparable and, therefore, the OR and the RR are nearly equal.

Hazard Ratio

The hazard ratio (HR) is a measure of effect in a time-to-event survival analysis (to be covered in a subsequent article). In brief, a survival analysis is used when the outcome of interest is the time between the start of a study to when the event of interest (eg, heart attack) occurs. The hazard rate is the instantaneous rate of failure

at some given time, given that the person has “survived” up to that time.

Mathematically, the HR is the ratio of hazard rates and is calculated by the hazard rate in the treatment group divided by the hazard rate in the control group. It is frequently more informative to compare the rates of the occurrence of 2 events rather than the cumulative number of events in each group at the end of the study (which the RR does). For example, if all subjects in the treatment group of a 4-year study experience the event of interest in the last year of the study (with no events in the first 3 years), their survival experience would be very different from a control group that experienced the event at a constant rate throughout the study. If, at the end of the study, the same number of events occurred in each equal-sized group, we would much rather be in the group that experienced events only in the last year of the study. Those 3 years of event-free survival would be very important! Survival analysis allows this kind comparison using HRs. In this example, the RR of the event would be 1.0 because the same proportion of patients experienced the event over the course of the study. However, the HR would be very different and would reflect the different hazard rates in each group.

When risks are below approximately 10%, the RR, OR, and HR are all comparable. For example, the event rates in cardiovascular studies are often less than 10%, and the RR, OR, and HR are similar. On the contrary, the event rates can be much higher in oncology studies, causing the RRs to be much different than the HRs and the ORs.

To summarize, the OR is the odds of an event occurring in an exposed group compared to the odds of it occurring in a nonexposed group (a ratio of 2 odds). The HR is measure of effect in a time-to-event survival analysis and informative to compare the rates of the occurrence of 2 events during a study.

Conclusion

Statistics of effect are measures used to describe, for example, the strength of a therapy in a study. Common measures include RR, RRR, ARD, and NNT. Reporting RRR without absolute risk may be misleading (eg, large RRR may have little clinical meaning if the absolute

risks are small [and the NNT is large]). The NNT/NNH incorporate baseline risk and are practical ways to express the effectiveness of a treatment to facilitate clinical decision-making. Probabilities in the frequentist approach are proportions (and not rates) and range from 0 to 1 (0% to 100%). As the baseline risk increases, the NNT decreases. The RR is a ratio of probabilities, the OR is a ratio of odds, and the HR is a ratio of hazard rates. Each of these measures have different uses, meanings, and limitations; and in future articles, we will expound on the appropriate uses of these measures of effect. In part 3 of this series, we will utilize the Bayesian approach and demonstrate how to interpret diagnostic tests with probabilistic thinking.

Financial Disclosures: None declared.

Funding/Support: None declared.

REFERENCES

1. Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. *BMJ*. 2010;341:c4830. doi:10.1136/bmj.c4830
2. Stigler SM. *The History of Statistics*. Belknap Press; 1986:88,97-98.
3. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728-1733. doi:10.1056/NEJM198806303182605

WMJ

Call for Artwork

The *WMJ* invites original artwork that illustrates the art and science of medicine to feature on our covers.

Learn more at
www.wmjonline.org

advancing the art & science of medicine in the midwest

WMJ

WMJ (ISSN 1098-1861) is published through a collaboration between The Medical College of Wisconsin and The University of Wisconsin School of Medicine and Public Health. The mission of *WMJ* is to provide an opportunity to publish original research, case reports, review articles, and essays about current medical and public health issues.

© 2024 Board of Regents of the University of Wisconsin System and The Medical College of Wisconsin, Inc.

Visit www.wmjonline.org to learn more.