## Some helpful definitions

| Measures of disease frequency | | | |
|---|---|---|---|
| **Measure** | **Calculation** | **Definition** | **Units** |
| Cumulative Incidence | $\dfrac{Subjects\ developing\ disease\ during\ time\ period}{Number\ of\ subjects\ followed\ for\ the\ time\ period}$ | Expresses the probability that an event will occur in a given population, or in a given individual, over a period of time. Also referred to as risk. | % |
| Incidence Rate | $\dfrac{Subjects\ developing\ disease\ during\ time\ period}{Total\ time\ at\ risk\ for\ all\ the\ subjects\ followed}$ | Expresses the rate at which the outcome occurs in a population over time. | person-time |
| Prevalence | $\dfrac{Subjects\ with\ disease\ at\ given\ time}{Number\ of\ subjects\ in\ population\ at\ same\ time}$ | The proportion of the population that is diseased at a given point in time. | % |

| Measures of association or effect | | | |
|---|---|---|---|
| **Measure** | **Calculation** | **Definition** | **Null** |
| Relative Risk | $\dfrac{Risk\ in\ group\ 1}{Risk\ in\ group\ 2}$ | Ratio of risks | 1 |
| Incidence Rate Ratio | $\dfrac{Incidence\ rate\ in\ group\ 1}{Incidence\ rate\ in\ group\ 2}$ | Ratio of rates | 1 |
| Hazard Ratio | $\dfrac{Hazard\ rate\ in\ group\ 1}{Hazard\ in\ group\ 2}$ | Calculated from Cox proportional hazard regression for time to event outcomes | 1 |
| Odds Ratio | $\dfrac{Odds\ in\ group\ 1}{Odds\ in\ group\ 2}$ | Odds are calculated as probability/1-probability. They approximate risk and are used primarily in case-control studies or with logistic regression. | 1 |
| Absolute Risk Difference | $Risk\ in\ group\ 1 - Risk\ in\ group\ 2$ | The difference in the risk (probability of event) or rate between two groups | 0 |
| Number needed to treat or harm | $\dfrac{1}{Absolute\ Risk\ Difference}$ | The average number of patients who need to be treated to prevent (or cause) one additional event during the follow-up period | ∞ |

Let's consider a hypothetical study where we are interested in examining the comparative-efficacy of two anticoagulants in a population of 10 patients with atrial fibrillation following a catheter ablation. At the time of the ablation, we randomized patients to either Drug 1 or Drug 2 in a 1:1 fashion. Figure 1 provides a graphical representation of the follow-up period. The white and black circles provide the baseline history of stroke for each patient. The solid lines are the amount of time each patient was followed, also referred to as time at risk, before being censored due to a stroke or last available follow up. The red x indicates the occurrence of a stroke, and the color of the follow-up line indicates the randomization assignment.

**Measures of disease frequency (or outcome frequency)**

**Incidence vs prevalence**

Before we can compare the two medications, we first need to describe the frequency of the events in the two groups. Incidence describes the frequency at which people transition from being at risk of disease to being diseased (or whatever the outcome may be). In other words, incidence deals with new outcomes occurring in a defined population over a defined period of time. Recurrent events may or may not be considered depending on the type of outcome and the study objectives. Conversely, prevalence
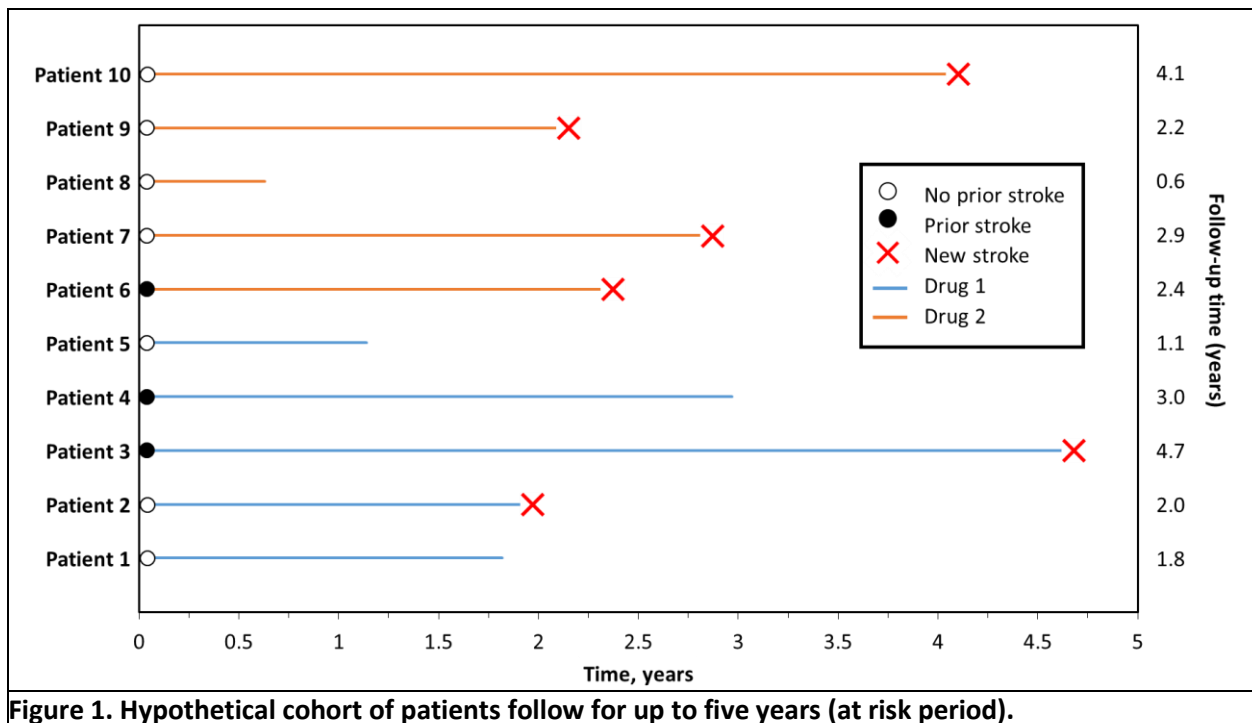


**Figure 1. Hypothetical cohort of patients follow for up to five years (at risk period).**

describes the frequency or proportion of the population that is in the diseased state at a specified point in time. This can be a snap shot in time (i.e., point prevalence) or over a period of time (i.e., period prevalence). It is easier to keep the concepts separate when we consider the study designs used to calculate them. Incidence is most commonly assessed through cohort study designs where we are following patients over time to ascertain outcomes. Prevalence is most commonly assessed through cross-sectional study designs where we are determining who currently has the disease. We do not assess disease frequency using case-control designs as they require us to arbitrarily define the ratio of cases to controls.

We see that six patients experienced a stroke during this time (black circle). We can describe the incidence as a probability (cumulative incidence) or as a rate (incidence rate). The cumulative incidence

is calculated by dividing the total number of outcomes by the total population followed (6/10 = 60%). We interpret this as there is a 60% of likelihood of developing a stroke in the 5 year period following a catheter ablation.

> **What is the problem with using and interpreting a cumulative incidence in this current study?** For one, not all patients were followed for the entire 5 year planned follow-up period. Patient 8 was only followed for 0.6 years, and it is quite possible that this patient would have experienced a stroke had they been followed for the entire 5 year period. The average duration of follow-up time among patients with a stroke vs those without a stroke was 3.2 and 1.8 years, respectively. The estimated risk of experiencing a stroke is therefore, likely underestimated in the population.

Another approach to assessing disease frequency is to express incidence as a rate of events rather than a probability of the event. The benefit of using a rate is that it allows for variable follow-up time between patients without providing biased estimates. Incidence rates are calculated by changing the denominator from the total population to the total time at risk in the population. We refer to this as person-time. Again, consider figure 1. The solid lines represents the time that a patient is still being followed and considered at risk of experiencing a stroke. We can calculate the person-time by summing the time at risk for each patient. Patient 1 was followed for 1.8 years, patient 2 for 2.0 years, and so on. Overall, there were 24.7 years of combined follow-up between the 10 patients, or 24.7 person-years. This results in an incidence rate of 0.24 events/person-years (6 strokes/24.7 person-years). The interpretation of a decimal with a rate is not very intuitive, but we can make this number more understandable by multiplying by 100 to come up with 24 events/100 person-years. This can be interpreted as, if we were to follow 100 patients for 1 year (or any combination of patients and time that add up to 100), we would anticipate that 24 patients would experience a stroke.

Now, let's describe the prevalence. Prevalence commonly refers to the proportion of the population with the disease at a given point of time (e.g., what proportion of the United States had lung cancer on January 1, 2016?). With an outcome that is an acute event or transient in nature, like stroke, this definition doesn't make much sense (e.g., what proportion of the United States was having a stroke on January 1, 2016?). For our purposes, we will define prevalence according to a history of the event; a patient who experiences a stroke never returns to a non-diseased state. What is the point prevalence, more commonly referred to simply as prevalence, at the time of randomization? Three of the 10 patients had a prior history of stroke at time of ablation, so the prevalence is 30% (3/10). We can also describe the period prevalence during the 5 year follow-up period, though this is not commonly reported.

**Measures of association (or measures of effect)**

Once we know how frequently an outcome is occurring, we can use these values to make between group comparisons. In a randomized clinical trial, we would consider this the causal relationship between the treatment groups. In retrospective or observational research, this would be the measure of association between the treatment groups. In either case, between group comparisons are expressed in either absolute or relative terms.

**Absolute (differences) vs relative (ratios) measures**

| Relative Risk | Compared to the reference group, the exposure is associated with: |
|---|---|
| < 1 | A lower risk of the outcome |
| = 1 | No difference in risk of the outcome |
| > 1 | A higher risk of the outcome |

In most medical research, relative measures of effect are the more commonly used tool. Relative risk is the ratio of the risk in group 1 to the risk of group 2. Our interest is to understand how

| Absolute Risk Difference | Compared to the reference group, the exposure is associated with: |
|---|---|
| < 0 | A lower risk of the outcome |
| = 0 | No difference in risk of the outcome |
| > 0 | A higher risk of the outcome |

this value deviates or not from 1, a null difference. An incident rate ratio is very similar, but compares the incident rates in each group rather than the risks. Since we have already determined that risk is likely not the best measure in our hypothetical study, let's calculate an incident rate ratio to compare the two treatment groups.

$$\frac{\dfrac{2 \; strokes}{12.6 \; person-years}}{\dfrac{4 \; strokes}{12.6 \; person-years}} = \frac{0.16}{0.33} = 0.48$$

Drug 1 is associated with a 52% reduction in stroke rate as compared with drug 2.

Another way to compare risk between the groups is the absolute risk difference, also called the absolute risk reduction. Absolute differences refer to subtracting risk or rates rather than dividing and represent another approach to comparing treatments. A null difference in an absolute terms is 0. Two out of 5 patients experienced a stoke on Drug 1 (40%) and 4 out of 5 patients experienced a stroke on Drug 2 (80%). The absolute difference in stroke risk is 40%.

**When should we use absolute vs relative measures of effect?** This answer largely depends on the desired messaging. For example, when addressing population health questions, an absolute measure is usually preferred. Whereas a relative measure is useful when we are interested in understanding to what extent the exposure accounts for the occurrence of the outcome. We need to always consider baseline risk. A relative effect may be large when the exposure accounts for nearly all of the outcome occurrence, but may still have little public health relevance if the outcome is rare. Personally, I prefer to report absolute values of the measure of effect for the population as a whole, and within each treatment arm, but use relative measures to describe the between group comparisons. (See Rothman 2012 for a more detailed discussion).

**Number Needed to Treat (NNT) and Number Needed to Harm (NNH)**

Absolute differences allow us to calculate the Number Needed to Treat (NNT), a valuable parameter that makes results more readily interpretable by clinicians and patients alike. The NNT is simply the inverse of the ARD, or 1/ARD. It represents the number of patients who need to receive the study intervention in order to avoid (NNT) or cause (NNH) one additional outcome over the length of the study.

In our hypothetical study, 1/0.40 = 2.5. NNT is typically rounded up to the nearest whole number since we can't treat a partial person. We would need to treat 3 additional patients with drug 1 as compared to drug 2 in order to prevent 1 additional stroke during the 5 year follow-up.

**Odds Ratios**

In the example we have been using a cohort design. What if we had been using a case-control design where we identified cases of stroke and randomly matched them with non-stroke controls? How would we calculate risk?
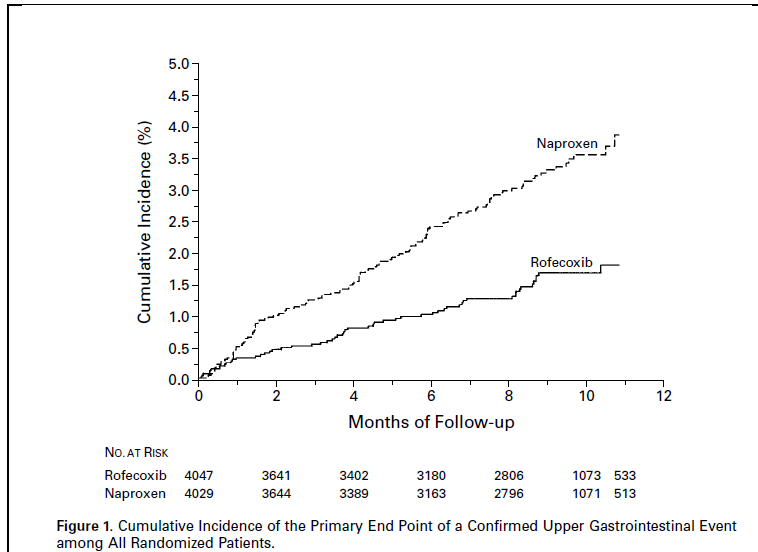
Figure 1. Cumulative Incidence of the Primary End Point of a Confirmed Upper Gastrointestinal Event among All Randomized Patients.

**Figure 2. Example of a graph of incidence over time from the VIGOR study**

The answer is we can't, because we have arbitrarily set the ratio of cases to controls. In a 1:1 match, the proportion of the study with the outcome is always 50%; this provides no information on the risk of the outcome in the underlying population. While we can't determine the measure of disease frequency using this design, we can still calculate a measure of association called an odds ratio. The odds ratio is an approximation of the relative risk. While the relative risk provides information regarding the risk of developing the outcome in a given population, the odds ratio tell us the odds of exposure given the occurrence of the outcome. Odds are calculated as probability/1-probability, and the odds ratio is the ratio of the odds among case compared to controls. The odds ratio closely approximates the relative risk when the outcome is rare. However, the magnitude of effect begins to be overestimated as the baseline risk increases and the odds ratio increases.

**Hazard Ratios**

Hazard ratios are very commonly reported in both clinical trials and observational research. They are the appropriate measure when the outcome is dichotomous and time-to-event is important; this is true in many if not most cohort studies. They are calculated from a Cox proportional hazard regression models which allow for both unadjusted results (such as in randomized trials) or adjusted results (as are required in non-randomized studies). An important aspect of the hazard ratio, is that is allows for variable follow-up among patients. The hazard ratio frequently accompanies a time-to-event graph (see Figure 2).

**Confidence Intervals**

So far, everything we have calculated has been a point estimate. The next logical question is whether this observed difference is statistically significant. What is the likelihood that the observed difference is due to chance? The p-value is one common way to determine this. However, the p-value provides precious little information outside of the statistical significance. Confidence intervals also provide information regarding statistical significance, but importantly also describe precision. See figure 2. Values to the >1, red dashed line, represent greater risk in the exposed group as opposed to the referent group. Conversely, values <1 represent a protective effect in the exposed group. A value is considered statistically significant if the 95% confidence interval does not cross 1 and corresponds to a p-value of <0.05. Now look at the size or width of the confidence intervals. This is what is meant by precision. The narrower the interval, the more confident we are in the magnitude of the estimate. The two values in the middle both cross 1 and are therefore considered non-statistically significant. However, the first value has a much narrower confidence interval, and so we are confident that the true value is likely

close to null. The second value is much wider and the true value may very well be substantially less that 1, but was underpowered to achieve a significant difference.
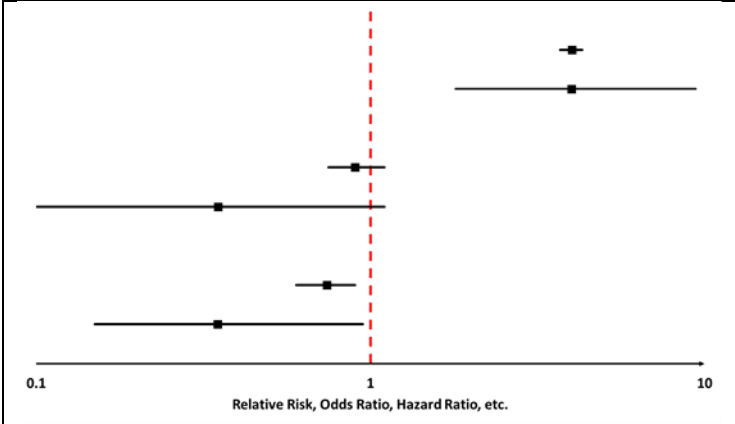


**Figure 3. Interpreting confidence intervals**

**References:**

Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med. 2000 Nov 23;343(21):1520-8

Koepsell, T. D., & Weiss, N. S. (2003). Epidemiologic methods: Studying the occurrence of illness. Oxford: Oxford University Press.

Rothman, K. J. (2012). Epidemiology: An Introduction, (2nd ed.). New York, NY: Oxford University Press.

**EXERCISES:**

**Question 1: Which of the following study designs can NOT be used to assess disease frequency in a population?**

A) Cohort

B) Case-control

C) Cross-sectional

D) All of the above can be used

**Question 2: What type of outcome variable must you be dealing with to calculate a relative risk?**

A. Continuous

B. Ordinal

C. Dichotomous

D. Categorical

**Question 3: Which of the following are appropriate measures of effect when the time of follow-up varies between patients? Select all that apply:**

A. Relative Risk

B. Incidence Rate Ratio

C. Odds Ratio

D. Hazard Ratio

**Results from the VIGOR study**

| Outcome | Rofecoxib | | | Naproxen | | |
| --- | --- | --- | --- | --- | --- | --- |
| | N total | N with Event | Person-years at risk | N total | N with Event | Person-years at risk |
| GI event | 4047 | 56 | 2667 | 4029 | 121 | 2689 |
| Myocardial infarction | 4047 | 17 | 2315 | 4029 | 4 | 2316 |

**Using the results from the VIGOR study to answer questions 4-6:**

4. **Using the data from the VIGOR study, calculate the <u>relative risk</u> for each outcome**

    a. GI event (relative to naproxen): _____

    b. Myocardial infarction (relative to naproxen): _____

5. **Using the data from the VIGOR study, calculate the <u>incidence rate ratio</u> for each outcome**

    a. GI event (relative to naproxen): _____

    b. Myocardial infarction (relative to naproxen): _____

6. **Using the data from the VIGOR study, calculate the <u>NNT and NNH</u> for each outcome using the ARD**

    a. GI event (relative to naproxen): _____

    b. Myocardial infarction (relative to naproxen): _____

**Question 7: Read through the abstract from the VIGOR study and pay particularly close attention to the results section. Think about the following questions:**

- What were the measures of disease frequency used?

- How were the between group differences measured (relative or absolute measures of effect)?

- What was the reference group?

- Were the measures used and reference group the same for all outcomes reported? If not, why do you think this was done?

**ABSTRACT**

**Background**

Each year, clinical upper gastrointestinal events occur in 2 to 4 percent of patients who are taking nonselective nonsteroidal anti-inflammatory drugs (NSAIDs). We assessed whether rofecoxib, a selective inhibitor of cyclooxygenase-2, would be associated with a lower incidence of clinically important upper gastrointestinal events than is the nonselective NSAID naproxen among patients with rheumatoid arthritis.

**Methods**

We randomly assigned 8076 patients who were at least 50 years of age (or at least 40 years of age and receiving long-term glucocorticoid therapy) and who had rheumatoid arthritis to receive either 50 mg of rofecoxib daily or 500 mg of naproxen twice daily. The primary end point was confirmed clinical upper gastrointestinal events (gastroduodenal perforation or obstruction, upper gastrointestinal bleeding, and symptomatic gastroduodenal ulcers).

**Results**

Rofecoxib and naproxen had similar efficacy against rheumatoid arthritis. During a median follow-up of 9.0 months, 2.1 confirmed gastrointestinal events per 100 patient-years occurred with rofecoxib, as compared with 4.5 per 100 patient-years with naproxen (relative risk, 0.5; 95 percent confidence interval, 0.3 to 0.6; P<0.001). The respective rates of complicated confirmed events (perforation, obstruction, and severe upper gastrointestinal bleeding) were 0.6 per 100 patient-years and 1.4 per 100 patient-years (relative risk, 0.4; 95 percent confidence interval, 0.2 to 0.8; P=0.005). The incidence of myocardial infarction was lower among patients in the naproxen group than among those in the rofecoxib group (0.1 percent vs. 0.4 percent; relative risk, 0.2; 95 percent confidence interval, 0.1 to 0.7); the overall mortality rate and the rate of death from cardiovascular causes were similar in the two groups.

**Conclusions**

In patients with rheumatoid arthritis, treatment with rofecoxib, a selective inhibitor of cyclooxygenase-2, is associated with significantly fewer clinically important upper gastrointestinal events than treatment with naproxen, a nonselective inhibitor.

(N Engl J Med 2000;343:1520-8.)

7. Draft an abstract for your study following the relevant abstract submission guidelines for you (e.g., Mountain States Conference, etc.). It is ok if you don't have results yet. Still attempt to draft all sections using blanks in the results section as place holders that can be filled in later. For example:

"During a median follow-up of xx.x months, x.x confirmed gastrointestinal events per 100 patient-years occurred with rofecoxib, as compared with x.x per 100 patient-years with naproxen (relative risk, x.xx; 95 percent confidence interval, x.xx to x.xx; P<x.xxx)."

| Question | Answer | Explanation |
|---|---|---|
| 1 | B | In case-control studies we artificially select the ratio of cases to controls, and do not have a true measure of the overall population. While we can still use odds ratios to compare exposures, we cannot assess disease frequency. |
| 2 | C | Odds ratios, relative risks, and hazard ratios all require the outcome to be dichotomous (yes/no). However, the exposure variable is not confined to these same. |
| 3 | B, D | Both incident rate ratios and hazard ratios utilize rates rather than risk, and allow for variable follow-up between groups. Relative risk and odds ratios can still be calculated if the outcome is dichotomous, however, they are not provide biased estimates when there is variable follow-up. |
| 4 | 0.47, 4.20 | (56/4047)/(121/4029) = 0.014/0.030=0.47 <br> (17/4047)/(4/4029)=0.0042/0.0100=4.20 |
| 5 | 0.47, 4.29 | (56/2667)/(121/2689)=2.1/4.5=0.47 <br> (17/2689)/(4/2316)=7.3/1.7=4.29 |
| 6 | NNT=63, NNH=313 | Calculated from the ARD (using the incident rate ratio would give different numbers). NNT = 1 / (0.014 - 0.030). NNH = 1 / (0.0042 - 0.0010). |
| 7 | | The authors changed from using incidence rates to using cumulative incidence for the safety outcome. They also changed the referent group from naproxen to rofecoxib. Given the short follow up (average of 9 months), the cumulative incidence looks smaller than the rate. Also, by using rofecoxib as the referent group it changes the interpretation from "rofecoxib causes MI" to "naproxen prevents MI." |