

7.3: Conditional probability and evidence-based medicine

Introduction

This section covers a lot of ground. We begin by addressing how probability of multiple events are calculated assuming each event is independent. The assumption of independence is then relaxed, and how to determine probability of an event happening given another event has already occurred, conditional probability is introduced. Use of **conditional probability** to interpret results of a clinical test are also introduced, along with the concept of “**evidence-based-medicine**,” or EBM.

Probability and independent events

Probability distributions are mathematical descriptions of the probabilities of how often different possible outcomes occur. We also introduced basic concepts related to working with the probabilities involving more than one event.

For review, for **independent events**, you multiply the individual chance that each event occurs to get the overall probability.

Example of multiple, independent events



Figure 7.3.1: Now that's a box full of kittens. Creative Commons License, source: <https://www.flickr.com/photos/83014408@N00/160490011>

What is the chance of five kittens in a litter of five to be of the same sex? In feral cat colonies, siblings in a litter share the same mother, but not necessarily the same father, **superfecundation**. Singleton births are independent events, thus the probability of the first kitten being female is 50%; the second kitten being female, also 50%; and so on. We can multiply the independent probabilities (hence, the **multiplicative rule**), to get our answer:

```
kittens <- c(0.5, 0.5, 0.5, 0.5, 0.5)
prod(kittens)
[1] 0.03125
```

Probabilistic risk analysis

Risk analysis is the use of information to identify hazards and to estimate the risk. A more serious example. Consider the 1986 *Challenger* Space Shuttle Disaster (Hastings 2003). Among the crew killed was Ellison Onizuka, the first Asian American to fly in space (Fig. 7.3.2, first on left back row). Onizuka was born and raised on Hawai`i and graduated from [Konawaena High School](#) in 1964.



Figure 7.3.2: STS-51-L crew: (front row) Michael J. Smith, Dick Scobee, Ronald McNair; (back row) Ellison Onizuka, Christa McAuliffe, Gregory Jarvis, Judith Resnik. Image by NASA – NASA Human Space Flight Gallery, Public Domain.

The shuttle was destroyed just 73 seconds after liftoff (Fig. 7.3.3).



Figure 7.3.3: Space shuttle *Challenger* launches from launchpad 39B Kennedy Space Center, FL, at the start of STS-51-L. Hundreds of shorebirds in flight.

This next section relies on material and analysis presented in the [Rogers Commission Report June 1986](#). NASA had estimated that the probability of one engine failure would be 1 in 100, or 0.01; two engine failures would mean the shuttle would be lost. Thus, the probability of two rockets failing at the same time was calculated as 0.01×0.01 , which is 0.0001 or 0.01%.

NASA had planned to fly the fleet of shuttles 100 times per year, which would translate to a shuttle failure once in 100 years. The Challenger launch on January 28, 1986, represented only the 25th flight of the shuttle fleet.

One difference on launch day was that the air temperature at Cape Canaveral was quite low for that time of year, as low as 22°F overnight.

Attention was pointed at the large O-rings in the boosters (engines). In all, there were six of these O-rings. Testing suggested that, at the colder air temperatures, the chance that one of the rings would fail was 0.023. Thus, the chance of success was only 0.977. Assuming independence, what is the chance that the shuttle would experience O-ring failure?

```
shuttle <- c(0.977, 0.977, 0.977, 0.977, 0.977, 0.977)
#probability of success then was
prod(shuttle)
[1] 0.869
```

```
#and therefore probability of failure was  
1 - prod(shuttle)  
[1] 0.1303042
```

Conditional probability of non-independent events

But in many other cases, independence of events cannot be assumed. The probability of an event given that another event has occurred is referred to as **conditional probability**. Conditional probability is used extensively to convey risk. We've touched on some of these examples already:

- the risk of subsequent coronary events given high cholesterol;
- the risk of lung cancer given that a person smokes tobacco;
- the risk of mortality from breast cancer given that regular mammography screening was conducted.

There are many, many examples in medicine, insurance, you name it. It is even an important concept that judges and lawyers need to be able to handle (e.g., Berry 2008).

A now famous example of conditional probability in the legal arena came from arguments over the chance that a husband or partner who beats his wife will subsequently murder her — this was an argument raised by the prosecution during pre-trial in the 1995 OJ Simpson trial (*The People of the State of California v. Orenthal James Simpson*), and successfully argued by O.J. Simpson's attorneys... the judge ruled in favor of the defense, and evidence of OJ Simpson's prior abuse were not included in trial. Gigerenzer (2002) and others have called this reverse **Prosecutor's Fallacy**, where the more typical scenario is that the prosecution provides a list of probabilities about characteristics of the defendant, leaving the jury to conclude that no one else could have possibly fit the description. In the OJ Simpson case, the argument went something like this. From the CDC we find that an estimated 1.3 million women are abused each year by their partner or former partner; each year about 1000 women are murdered. One thousand divided by 1.3 million is a small number, so even when there is abuse, the argument goes, 99% of the time there is not murder. The Simpson judge ruled in favor of the defense and much of the evidence of abuse was excluded.

Something is missing from the defense's argument. Nicole Simpson did not belong to a population of battered women — she belonged to the population of murdered women. When we ask, if a woman is murdered, what is the chance that she knew her murderer, we find that more than 55% knew their murderer — and of that 55%, 93% were killed by a current partner. The key is that Nicole Simpson (and Ron Goldman) was murdered and OJ Simpson was an ex-partner who had been guilty of assault against Nicole Simpson. Now, it goes from an impossibly small chance, to a much greater chance. Conditional probability, and specifically **Bayes' rule**, is used for these kinds of problems.

Diagnosis from testing

Let's turn our attention to medicine. A growing practice in medicine is to claim that decision-making in medicine should be based on approaches that give us the best decisions. A search of PubMed texts for “evidence based medicine” found more than 91,944 (13 October 2021, an increase of thirteen thousand since last I looked (10 October, 2019). **Evidence based medicine** (EBM) is the “conscientious, explicit, judicious and reasonable use of modern, best evidence in making decisions about the care of individual patients” (Masic et al 2008). By **evidence**, we may mean results from quantitative, **systematic reviews**, **meta-analysis**, of research on a topic of medical interest, e.g., [Cochrane Reviews](#).

Note:

Primary research refers to generating or collecting original data in pursuit of tests of hypotheses. Both systematic reviews and meta-analysis are **secondary research** or “research on research.” As opposed to a literature review, systematic review make explicit how studies were searched for and included; if enough reasonably similar quantitative data are obtained through this process, the reviewer can combine the data and conduct an analysis to assess whether a treatment is effective (De Vries 2018).

As you know, no **diagnostic test** is 100% foolproof. For many reasons, test results come back positive when the person truly does not have the condition — this is a **false positive** result. Correctly identifying individuals who do not have the condition, which ideally means having a 0% false positive rate, is called the **specificity** of a test. Think of specificity in this way — provide the test 100 **true negative** samples (e.g., 100 samples from people who do not have cancer) — how many times out of 100 does the test correctly return a “negative”? If 99 times out of 100, then the specificity rate for this test is 99%, which is pretty good. But the test results mean more if the condition/disease is common; for rare conditions, even 99% is not good enough. Incorrect assignments are

rare, we believe, in part because the tests are generally quite accurate. However, what we don't consider is that detection and diagnosis from tests also depend on how frequent the incidence of the condition is in the population. Paradoxically, the lower the base rate, the poorer diagnostic value even a sensitive test may have.

To summarize our jargon for interpreting a test or assay, so far we have

- **True positive (a)**, the person has a disease and the test correctly identifies the person as having the disease.
- **False positive (b)**, test result incorrectly identifies disease; the person does not have the disease, but the test classifies the person as having the disease.
- **False negative (c)**, test result incorrectly classifies a person as not having disease, but the person actually has the disease.
- **True negative (d)**, the person does not have the disease and the test correctly categorizes the person as not having the disease.
- **Sensitivity of test** is the proportion of persons who test positive and do have the disease (true positives): $Se = \frac{TP}{TP+FN}$
If a test has 75% sensitivity, then out of 100 individuals who do have the disease, 75 will test positive (TP = true positive).
- **Specificity of a test** refers to the rate that a test correctly classifies a person that does not have the disease: $Sp = \frac{TN}{TN+FP}$
If a test has 90% specificity, then out of 100 individuals who truly do not have the disease, 90 will test negative (TN = true negative).

A worked example. A 50-year-old male patient is in the doctor's office. The doctor is reviewing results from a diagnostic test, e.g., a FOBT — fecal occult blood test — a test used as a screening tool for colorectal cancer (CRC). The doctor knows that the test has a sensitivity of about 75% and specificity of about 90%. Prevalence of CRC in this age group is about 0.2%. Figure 7.3.4 shows our probability tree using our natural numbers approach.

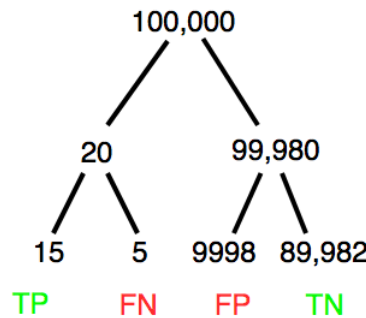


Figure 7.3.4: Probability tree for FOBT test. Desired test outcomes shown in green: TP stands for true positive and TN stands for true negative. Poor outcomes of a test shown in red: FN stands for false negative and FP stands for false positive.

The associated probabilities for the four possible outcomes of these kinds of tests (e.g., what's the probability of a person who has tested positive in screening tests actually having the disease?) are shown in Table 7.3.1.

Table 7.3.1. A 2×2 table of possible outcomes of a diagnostic test.

	Does person really have the disease?	
Test Result:	Yes	No
Positive	a TP	b FP
Negative	c FN	d TN

Bayes' rule is often given in probabilities,

$$p(D|\oplus) = \frac{p(D) \cdot p(\oplus|D)}{[p(D) \cdot p(\oplus) + p(ND) \cdot p(\oplus|ND)]}$$

=where truth is represented by either D (the person really does have the disease) or ND (the person really does not have the disease) and \oplus is the symbol for “exclusive or” and reads “not D ” in this example.

An easier way to see this is to use frequencies instead. Now, the formula is

$$p(\text{disease}|\text{positive}) = \frac{a}{a+b}$$

This is the simplified version of Bayes' rule, where a is the number of people who test positive and DO HAVE the disease and b is the number of people who test positive and DO NOT have the disease.

Standardized million

Where did the 100,000 come from? We discussed this in Chapter 2: it's a simple trick to adjust rates to the same population size. We use this to work with natural numbers instead of percentages or frequencies. You choose to correct to a standardized population based on the raw incidence rate. A rough rule of thumb:

Table 7.3.2. Relationship between standard population size and incidence rate.

Raw IR rate about	IR	Standard population
1/10	10%	1000
1/100	1%	10,000
1/1000	0.1%	100,000
1/10,000	0.01%	1,000,000
1/100,000	0.001%	10,000,000
1/1,000,000	0.0001%	100,000,000

The raw incident rate is simply the number of new cases divided by the total population.

Per capita rate

Yet another standard manipulation is to consider the average incidence per person, or per capita, rate. The Latin "per capita" translates to "by head" (Google translate), but in economics, epidemiology, and other fields it is used to reflect rates per person. Tuberculosis is a serious infectious disease of primarily the lungs. Incidence rates of tuberculosis in the United States have trended down since the mid-1900s: 52.6 per 100K in 1953 to 2.7 per 100K in 2019 (CDC). Corresponding per capita values are 5.26×10^{-4} and 2.7×10^{-5} , respectively. Divide the incidence rate by 100,000 to get the per-person rate.

Practice and introduce PPV and Youden's J

Let's break these problems down, and in doing so, introduce some terminology common to the field of "risk analysis" as it pertains to biology and epidemiology. Our first example considers the fecal occult blood test, FOBT, test. Blood in the stool may (or may not) indicate polyps or colon cancer. (Park et al 2010).

Table 7.3.3. A 2×2 table of possible outcomes of FOBT test.

	Person really has the disease		
	Yes	No	
Positive	15	9998	$PPV = \frac{15}{(15+9998)} = 0.15\%$
Negative	5	89,982	$NPV = \frac{89982}{(89982+5)} = 99.99\%$

The table shown above will appear again and again throughout the course, but in different forms.

We want to know, how good is the test, particularly if the goal is early detection? This is conveyed by the **PPV, positive predictive value** of the test. Unfortunately, the prevalence of a condition is also at play: the lower the prevalence, the lower the PPV must be, because most positive tests will be false when population prevalence is low.

Youden (1950) proposed a now widely adopted index that summarizes how effective a test is. **Youden's J** is the sum of specificity and sensitivity minus one.

$$J = Se + Sp - 1$$

where Se stands for sensitivity of the test and Sp stands for specificity of the test.

Software

R has several epidemiology packages, and with some effort, can save you time. Another option is to run your problems in [OpenEpi](#), a browser-based set of tools. OpenEpi is discussed with examples in the next section, [7.4](#).

Here, we illustrate some capabilities of the epiR package, expanded more also in the next section, [7.4](#). We'll use the example from Table 7.3.3.

R code:

```
library(epiR)
Table3 <- matrix(c(15, 5, 9998, 89982), nrow = 2, ncol = 2)
epi.tests(Table3)
```

R output:

```

      Outcome + Outcome -   Total
Test +         15      9998  10013
Test -          5     89982  89987
Total          20     99980 100000

Point estimates and 95% CIs:
-----
Apparent prevalence *           0.10 (0.10, 0.10)
True prevalence *              0.00 (0.00, 0.00)
Sensitivity *                  0.75 (0.51, 0.91)
Specificity *                  0.90 (0.90, 0.90)
Positive predictive value *      0.00 (0.00, 0.00)
Negative predictive value *     1.00 (1.00, 1.00)
Positive likelihood ratio       7.50 (5.82, 9.67)
Negative likelihood ratio       0.28 (0.13, 0.59)
False T+ proportion for true D- * 0.10 (0.10, 0.10)
False T- proportion for true D+ * 0.25 (0.09, 0.49)
False T+ proportion for T+ *    1.00 (1.00, 1.00)
False T- proportion for T- *    0.00 (0.00, 0.00)
Correctly classified proportion * 0.90 (0.90, 0.90)
-----
* Exact CIs
```

Oops! I wanted PPV, which by hand calculation was 0.15%, but R reported “0.00?” This is a **significant figure** reporting issue. The simplest solution is to submit `options(digits=6)` before the command, then save the output from `epi.tests()` to an object and use `summary()`. For example

```
options(digits=6)
myEpi <- epi.tests(Table3)
summary(myEpi)
```

And R returns

```

      statistic      est      lower      upper
1          ap 0.1001300000 0.0982761568 0.102007072
```

```

2      tp  0.0002000000 0.0001221693 0.000308867
3      se  0.7500000000 0.5089541283 0.913428531
4      sp  0.9000000000 0.8981238085 0.901852950
5      diag.ac 0.8999700000 0.8980937508 0.901823014
6      diag.or 27.0000000000 9.8110071871 74.304297826
7      nndx  1.5384615385 1.2265702376 2.456532054
8      youden 0.6500000000 0.4070779368 0.815281481
9      pv.pos 0.0014980525 0.0008386834 0.002469608
10     pv.neg 0.9999444364 0.9998703379 0.999981958
11     lr.pos  7.5000000000 5.8193604069 9.66010707
12     lr.neg  0.2777777778 0.1300251423 0.593427490
13     p.rout  0.8998700000 0.8979929278 0.901723843
14     p.rin  0.1001300000 0.0982761568 0.102007072
15     p.tpdn  0.1000000000 0.0981470498 0.101876192
16     p.tndp  0.2500000000 0.0865714691 0.491045872
17     p.dntp  0.9985019475 0.9975303919 0.999161317
18     p.dptn  0.0000555636 0.0000180416 0.000129662

```

There we go — `pv.pos` reported as 0.0014980525, which, after turning to a percent and rounding, we have 0.15%. Note also the additional statistics provided — a good rule of thumb — always try to save the output to an object, then view the object, e.g., with `summary()`. Refer to help pages for additional details of the output (`?epi.tests`).

What about R Commander menus?

Note: Fall 2023

I have not been able to run the EBM plugin successfully! It simply returns an error message — on data sets which have in the past performed perfectly. Thus, until further notice, do not use the EBM plugin. Instead, use commands in the `epiR` package. I'm leaving the text here on the chance the error with the plugin is fixed.

Rcmdr has a plugin that will calculate ARR, RRR and NNT. The plugin is called `RcmdrPlugin.EBM` (Leucuta et al 2014) and it would be downloaded as for any other package via R.

Download the package from your selected R mirror site, then start R Commander.

```
install.packages("RcmdrPlugin.EBM")
```

From within R Commander (Fig. 7.3.6), select

Tools → Load Rcmdr plug-in(s)...

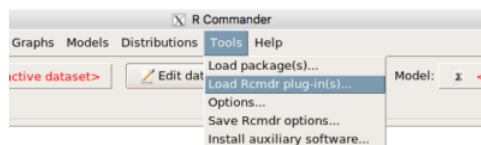


Figure 7.3.6: To install an Rcmdr plugin, first go to Rcmdr → Tools → Load Rcmdr plug-in(s)...

Next, select from the list the plug-in you want to load into memory, in this case, `RcmdrPlugin.EBM` (Fig. 7.3.7).

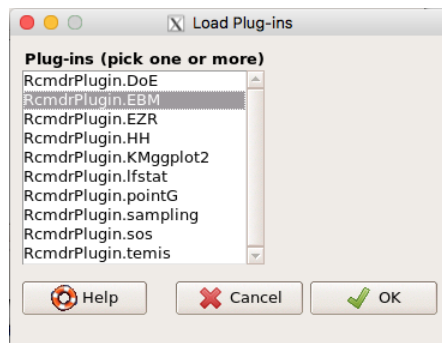


Figure 7.3.7: Select the Rcmdr plugin, then click the “OK” button to proceed.

Restart Rcmdr again (Fig. 7.3.8),

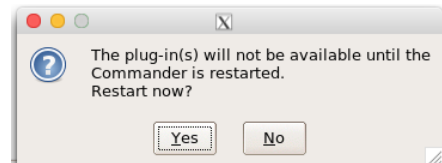


Figure 7.3.8: Select “Yes” to restart R Commander and finish installation of the plug-in.

and the menu “EBM” should be visible in the menu bar (Fig. 7.3.9).

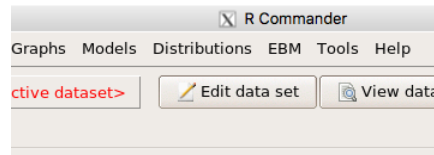


Figure 7.3.9: Copy and Paste Caption here. (Copyright; author via source)

Note that you will need to repeat these steps each time you wish to work with a plug-in, unless you modify your `.RProfile` file. See

Rcmdr → *Tools* → *Save Rcmdr options...*

Clicking on the EBM menu item brings up the template for the Evidence Based Medicine module. We’ll mostly work with 2×2 **tables** (e.g., see Table 7.3.1) , so select the “Enter two-way table...” option to proceed (Fig. 7.3.10).

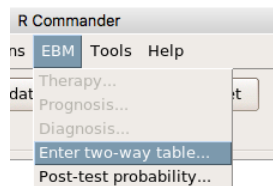


Figure 7.3.10: Select “Enter two-way table...”.

And finally, Figure 7.3.11 shows the two-way table entry cells along with options. We’ll try a problem by hand, then use the EBM plugin to confirm and gain additional insight.

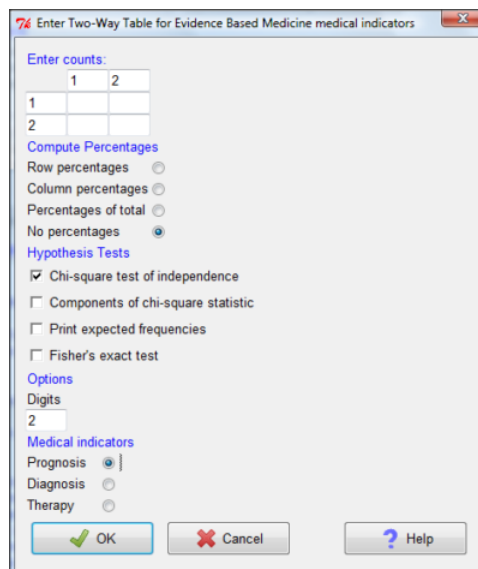


Figure 7.3.11: Two-way table Rcmdr EBM plug-in.

For assessing how good a test or assay is, use the Diagnosis option in the EBM plugin. For situations with treated and control groups, use Therapy option. For situations in which you are comparing exposure groups (e.g., smokers vs non-smokers), use the Prognosis option.

Example

Here's a simple one (problem from Gigerenzer 2002).

About 0.01% of men in Germany with no known risk factors are currently infected with HIV. If a man from this population actually has the disease, there is a 99.9% chance the tests will be positive. If a man from this population is not infected, there is a 99.9% chance that the test will be negative. *What is the chance that a man who tests positive actually has the disease?*

Start with the reference, or base population (Figure 7.3.12). It's easy to determine the rate of HIV infection in the population if you use numbers. For 10,000 men in this group, exactly one man is likely to have HIV ($0.0001 \times 10,000$), whereas 9,999 would not be infected.

For the man who has the disease it's virtually certain that his results will be positive for the virus (because the sensitivity rate = 99.9%). For the other 9,999 men, one will test positive (the false positive rate = $1 - \text{specificity rate} = 0.01\%$).

Thus, for this population of men, for every two who test positive, one has the disease and one does not, so the probability even given a positive test is only $100 \times 1/2 = 50\%$. This would also be the test's Positive Predictive Value.

Note that if the base rate changes, then the final answer changes!

It also helps to draw a tree to help you determine the numbers (Fig. 7.3.12)

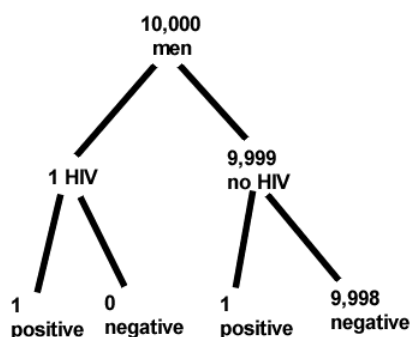


Figure 7.3.12: Draw a probability tree to help with the frequencies.

From our probability tree in Figure 7.3.12 it is straightforward to collect the information we need.

- Given this population, how many are expected to have HIV? **Two**.
- Given the specificity and sensitivity of the assay for HIV, how many persons from this population will test positive? **Two**.
- For every positive test result, how many men from this population will actually have HIV? **One**.

Thus, given this population with the known risk associated, the probability that a man testing positive actually has HIV is 50% ($= \frac{1}{(1+1)}$).

Use the EBM plugin. Select two-way table, then enter the values as shown in Fig. 7.3.13

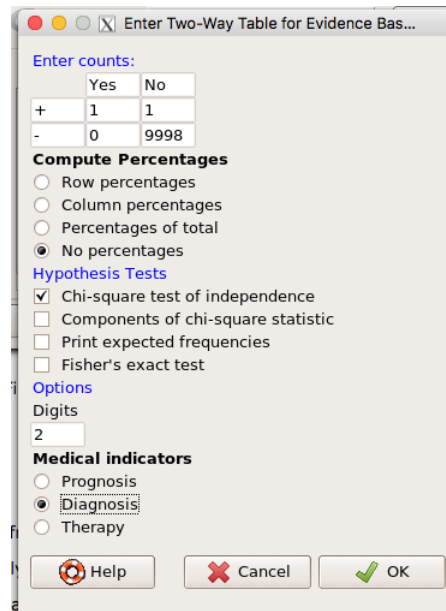


Figure 7.3.13: EBM plugin with data entry.

Select the “Diagnosis” option — we are answering the question: How probable is a positive result given information about sensitivity and specificity of a diagnosis test? The results from the EBM functions are given below.

```
Rcmdr> .Table
      Yes No
+    1    1
-    0 9998
```

```
Rcmdr> fncEBMCrossTab(.table=.Table, .x='', .y='', .ylab='', .xlab='',
Rcmdr+ .percents='none', .chisq='1', .expected='0', .chisqComp='0', .fisher='0',
Rcmdr+ .indicators='dg', .decimals=2)
```

```
# Notations for calculations
      Disease + Disease -
Test +      "a"      "b"
Test -      "c"      "d"
```

```
# Sensitivity (Se) = 100 (95% CI 2.5 - 100) %. Computed using formula: a / (a + c)
# Specificity (Sp) = 99.99 (95% CI 99.94 - 100) %. Computed using formula: d / (b + d)
# Diagnostic accuracy (% of all correct results) = 99.99 (95% CI 99.94 - 100) %. Computed using formula: (a + d) / (a + b + c + d)
# Youden's index = 1 (95% CI 0.02 - 1). Computed using formula: Se + Sp - 1
# Likelihood ratio of a positive test = 9999 (95% CI 1408.63 - 70976.66). Computed using formula: (a + b) / (c + d)
```

```
# Likelihood ratio of a negative test = 0 (95% CI 0 - NaN). Computed using formula: (
# Positive predictive value = 50 (95% CI 1.26 - 98.74) %. Computed using formula: a /
# Negative predictive value = 100 (95% CI 99.96 - 100) %. Computed using formula: d /
# Number needed to diagnose = 1 (95% CI 1 - 40.91). Computed using formula: 1 / [Se -
```

Note that the formulas used to calculate Sensitivity, Specificity, etc., follow our Table 7.3.1 (compare to “Notations for calculations”). The use of EBM provides calculations of our confidence intervals.

Questions

1. The sensitivity of the fecal occult blood test (FOBT) is reported to be 0.68. What is the False Negative Rate?
2. The specificity of the fecal occult blood test (FOBT) is reported to be 0.98. What is the False Positive Rate?
3. For men between 50 and 54 years of age, the rate of colon cancer is 61 per 100,000. If the false negative rate of the fecal occult blood test (FOBT) is 10%, how many persons who have colon cancer will test negative?
4. For men between 50 and 54 years of age, the rate of colon cancer is 61 per 100,000. If the false positive rate of the fecal occult blood test (FOBT) is 10%, how many persons who do not have colon cancer will test positive?
5. A study was conducted to see if mammograms reduced mortality (data from Table 5-1 p. 60 Gigerenzer (2002)). What is the RRR?

Mammogram	Deaths/1000 women
No	4
Yes	3

6. A study was conducted to see if mammograms reduced mortality (data from Table 5-1 p. 60 Gigerenzer (2002)). What is the NNT?

Mammogram	Deaths/1000 women
No	4
Yes	3

7. Does supplemental Vitamin C decrease risk of stroke in Type II diabetic women? In a study conducted on 1923 women, a total of 57 women had a stroke. Of the 57, 14 were in the normal Vitamin C level and 32 were in the high Vitamin C level. What is the NNT between normal and high supplemental Vitamin C groups?
8. Sensitivity of a test is defined as
 - A. False Positive Rate
 - B. True Positive Rate
 - C. False Negative Rate
 - D. True Negative Rate
9. Specificity of a test is defined as
 - A. False Positive Rate
 - B. True Positive Rate
 - C. False Negative Rate
 - D. True Negative Rate
10. In thinking about the results of a test of a null hypothesis, Type I error rate is equivalent to
 - A. False Positive Rate
 - B. True Positive Rate
 - C. False Negative Rate
 - D. True Negative Rate
11. During the Covid-19 pandemic, number of reported cases each day were published. For example, 155 cases were reported for 9 October 2020 by Department of Health. What is the raw incident rate?

This page titled [7.3: Conditional probability and evidence-based medicine](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.