3.4 Applications

“Evidence-Based Medicine”: Screening Tests and Disease Diagnosis

Clinical tests are frequently used in medicine and epidemiology to diagnose or screen for the presence \((T^+)\) or absence \((T^-)\) of a particular condition, such as pregnancy or disease. Definitive disease status (either \(D^+\) or \(D^-\)) is often subsequently determined by means of a “gold standard,” such as data resulting from follow-up, invasive radiographic or surgical procedures, or autopsy. Different measures of the test’s merit can then be estimated via various conditional probabilities. For instance, the sensitivity or true positive rate of the test is defined as the probability that a randomly selected individual has a positive test result, \(T^+\), given that he/she actually has the disease. Other terms are defined similarly; the following example, using a random sample of \(n = 200\) patients, shows how they are estimated from the data.

<table>
<thead>
<tr>
<th>Disease Status</th>
<th>(D^+)</th>
<th>(D^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive ((T^+))</td>
<td>16 (= TP)</td>
<td>9 (= FP)</td>
</tr>
<tr>
<td>Negative ((T^-))</td>
<td>4 (= FN)</td>
<td>171 (= TN)</td>
</tr>
<tr>
<td>20</td>
<td>180</td>
<td>200</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{True Positive rate} & = P(T^+ \mid D^+) = \frac{16}{20} = .80 \\
\text{False Positive rate} & = P(T^+ \mid D^-) = \frac{9}{180} = .05 \\
\text{“Sensitivity”} & = \frac{16}{20} = .80 \\
1 - \text{specificity} & = \frac{9}{180} = .05 \\
\text{False Negative rate} & = P(T^- \mid D^+) = \frac{4}{20} = .20 \\
\text{True Negative rate} & = P(T^- \mid D^-) = \frac{171}{180} = .95 \\
1 - \text{sensitivity} & = \frac{4}{20} = .20 \\
\text{“Specificity”} & = \frac{171}{180} = .95
\end{align*}
\]
In order to be able to apply this test to the general population, we need accurate estimates of its **predictive values** of a positive and negative test, \(PV_+ = P(D+ \mid T+)\) and \(PV_- = P(D- \mid T-)\), respectively. We can do this via the basic definition

\[
P(B \mid A) = \frac{P(B \cap A)}{P(A)}
\]

which, when applied to our context, becomes

\[
P(D+ \mid T+) = \frac{P(D+ \cap T+)}{P(T+)} \quad \text{and} \quad P(D- \mid T-) = \frac{P(D- \cap T-)}{P(T-)},
\]

often written

\[
PV_+ = \frac{TP}{TP + FP} \quad \text{and} \quad PV_- = \frac{TN}{FN + TN}.
\]

Here, \(PV_+ = \frac{16}{25} = 0.64\) and \(PV_- = \frac{171}{175} = 0.977\).

However, a more accurate determination is possible, with the use of...

**Bayes’ Formula:**

\[
P(B \mid A) = \frac{P(A \mid B) P(B)}{P(A \mid B) P(B) + P(A \mid B^c) P(B^c)}
\]

which, when applied to our context, becomes

\[
P(D+ \mid T+) = \frac{P(T+ \mid D+) P(D+)}{P(T+ \mid D+) P(D+) + P(T+ \mid D-) P(D-)},
\]

i.e.,

\[
PV_+ = \frac{(\text{Sensitivity})(\text{Prevalence})}{(\text{Sensitivity})(\text{Prevalence}) + (\text{False Positive rate})(1 - \text{Prevalence})}
\]

and

\[
P(D- \mid T-) = \frac{P(T- \mid D-) P(D-)}{P(T- \mid D-) P(D-) + P(T- \mid D+) P(D+)},
\]

i.e.,

\[
PV_- = \frac{(\text{Specificity})(1 - \text{Prevalence})}{(\text{Specificity})(1 - \text{Prevalence}) + (\text{False Negative rate})(\text{Prevalence})}.
\]

All the ingredients are obtainable from the table calculations, except for the baseline **prevalence** of the disease in the population, \(P(D+)\), which is usually grossly overestimated by the corresponding sample-based value, in this case, \(20/200 = .10\). We must look to outside published sources and references for a more accurate estimate of this figure.
Suppose that we are able to determine the **prior probabilities:**

\[ P(D+) = 0.04 \quad \text{and therefore,} \quad P(D-) = 0.96. \]

Then, substituting, we obtain the following **posterior probabilities:**

\[ PV^+ = \frac{(0.80)(0.04)}{(0.80)(0.04) + (0.05)(0.96)} = 0.40 \quad \text{and} \quad PV^- = \frac{(0.95)(0.96)}{(0.95)(0.96) + (0.20)(0.04)} = 0.99. \]

Therefore, a positive test result increases the probability of having this disease from 4% to 40%; a negative test result increases the probability of *not* having the disease from 96% to 99%. Hence, this test is extremely specific for the disease (i.e., low false positive rate), but is not very sensitive to its presence (i.e., high false negative rate). A physician may wish to use a screening test with higher sensitivity (i.e., low false negative rate). However, such tests also sometimes have low specificity (i.e., high false positive rate), e.g., MRI screening for breast cancer. An ideal test generally has both high sensitivity and high specificity (e.g., mammography), but are often expensive. Typically, health insurance companies favor tests with three criteria: cheap, fast, and easy, e.g., Fecal Occult Blood Test (**FOBT**) vs. colonoscopy.
“Evidence-Based Medicine”: Receiver Operating Characteristic (ROC) Curves

Originally developed in the electronic communications field for displaying “Signal-to-Noise Ratio” (SNR), these graphical objects are used when numerical cutoff values are used to determine $T^+$ versus $T^-$.

Example: Using blood serum markers in a screening test ($T$) for detecting fetal Down’s syndrome ($D$) and other abnormalities, as maternal age changes.

**Triple Test:** Uses three maternal serum markers (alpha-fetoprotein, unconjugated oestriol, and human gonadotrophin) to calculate a woman’s individual risk of having a Down syndrome pregnancy.

<table>
<thead>
<tr>
<th>Maternal age (years)</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>40.4</td>
<td>97.4</td>
</tr>
<tr>
<td>25</td>
<td>43.9</td>
<td>97.0</td>
</tr>
<tr>
<td>30</td>
<td>51.9</td>
<td>95.2</td>
</tr>
<tr>
<td>35</td>
<td>70.5</td>
<td>87.1</td>
</tr>
<tr>
<td>40</td>
<td>89.8</td>
<td>63.7</td>
</tr>
</tbody>
</table>

Source: Canick JA, George JK. Multiple marker screening for fetal Down syndrome. Contemporary Ob/Gyn, 1992, April, pp. 3-12.
The True Positive rate (from 0 to 1) of the test is graphed against its False Positive rate (from 0 to 1), for a range of age levels, and approximated by a curve contained in the unit square. The farther this graph lies above the diagonal – i.e., the closer it comes to the ideal level of 1 – the better the test. This is often measured by the Area Under Curve (AUC), which has a maximum value of 1, the total area of the unit square. Often in practice, the “curve” is simply the corresponding polygonal graph (as shown), and AUC can be numerically estimated by the Trapezoidal Rule. (It can also be shown that this value corresponds to the probability that a random pregnancy can be correctly classified as Down, using this screening test.) Illustrated below are the ROC curves corresponding to three different Down syndrome screening tests; although their relative superiorities are visually suggestive, formal comparison is commonly performed by a modified version of the Wilcoxon Rank Sum Test (covered later).
**Further Applications: Relative Risk and Odds Ratios**

Measuring degrees of association between disease \((D)\) and exposure \((E)\) to a potential risk (or protective) factor, using a prospective **cohort study**: 

\[
\text{Given: Exposed (E+) and Unexposed (E-)} \quad \quad \text{Investigate: Association with D+ and D-}
\]

From the resulting data, various probabilities can be estimated. Approximately,

### Disease Status

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Diseased ((D+))</th>
<th>Nondiseased ((D-))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed ((E+))</td>
<td>(p_{11})</td>
<td>(p_{12})</td>
</tr>
<tr>
<td>Unexposed ((E-))</td>
<td>(p_{21})</td>
<td>(p_{22})</td>
</tr>
<tr>
<td></td>
<td>(p_{11} + p_{21})</td>
<td>(p_{12} + p_{22})</td>
</tr>
</tbody>
</table>

\[
P(D+ | E+) = \frac{P(D+ \cap E+)}{P(E+)} = \frac{p_{11}}{p_{11} + p_{12}} \quad \quad P(D- | E+) = \frac{P(D- \cap E+)}{P(E+)} = \frac{p_{12}}{p_{11} + p_{12}}
\]

\[
P(D+ | E-) = \frac{P(D+ \cap E-)}{P(E-)} = \frac{p_{21}}{p_{21} + p_{22}} \quad \quad P(D- | E-) = \frac{P(D- \cap E-)}{P(E-)} = \frac{p_{22}}{p_{21} + p_{22}}
\]

\[\Rightarrow \quad \text{Odds of disease, given exposure} = \frac{P(D+ | E+)}{P(D- | E+)} = \frac{p_{11}}{p_{12}} = \frac{p_{11} + p_{12}}{p_{11}} \quad \frac{p_{12}}{p_{12}} \Rightarrow \frac{p_{11}}{p_{12}}\]

\[\Rightarrow \quad \text{Odds of disease, given no exposure} = \frac{P(D+ | E-)}{P(D- | E-)} = \frac{p_{21}}{p_{22}} \quad \frac{p_{22}}{p_{22}} \Rightarrow \frac{p_{21}}{p_{22}}\]

### Odds Ratio: \(OR\)

\[
OR = \frac{P(D+ | E+)}{P(D- | E+)} \div \frac{P(D+ | E-)}{P(D- | E-)} = \frac{p_{11}}{p_{12}} \div \frac{p_{21}}{p_{22}} = \frac{p_{11}p_{22}}{p_{21}p_{12}} \quad \text{“cross product ratio”}
\]

**Comment:** If \(OR = 1\), then “odds, given exposure” = “odds, given no exposure,” i.e., no association exists between disease \(D\) and exposure \(E\). **What if \(OR > 1\) or \(OR < 1\)?**

### Relative Risk: \(RR\)

\[
RR = \frac{P(D+ | E+)}{P(D+ | E-)} = \frac{p_{11}}{p_{21}} \div \frac{p_{11} + p_{12}}{p_{21} + p_{22}} = \frac{p_{11}(p_{21} + p_{22})}{p_{21}(p_{11} + p_{12})} \quad \text{“cross product ratio”}
\]

**Comment:** \(RR\) directly measures the effect of exposure on disease, but \(OR\) has better statistical properties. However, if the disease is rare in the population, i.e., if \(p_{11} \approx 0\) and \(p_{21} \approx 0\), then \(RR = \frac{p_{11}(p_{21} + p_{22})}{p_{21}(p_{11} + p_{12})} \approx \frac{p_{11}p_{22}}{p_{12}p_{21}} = OR\).
Recall our earlier example of investigating associations between lung cancer and the potential risk factors of smoking and coffee drinking. First consider the former:

### Lung Cancer

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Diseased (D+)</th>
<th>Nondiseased (D−)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed (E+)</td>
<td>.12</td>
<td>.04</td>
</tr>
<tr>
<td>Not Exposed (E−)</td>
<td>.03</td>
<td>.81</td>
</tr>
</tbody>
</table>

\[
P(D+ | E+) = \frac{P(D+ \cap E+)}{P(E+)} = \frac{.12}{.16} = \frac{3}{4} \quad \text{and therefore,} \quad P(D− | E+) = \frac{.04}{.16} = \frac{1}{4}.
\]

A random smoker has a 3 out of 4 (i.e., 75%) probability of having lung cancer;
a random smoker has a 1 out of 4 (i.e., 25%) probability of not having lung cancer.

Therefore, the **odds** of the disease, given exposure, is \(\frac{3}{4} \times \frac{1}{4} = 3\).

The probability that a random smoker has lung cancer is 3 times greater than the probability that he/she does not have it.

\[
P(D+ | E−) = \frac{P(D+ \cap E−)}{P(E−)} = \frac{.03}{.84} = \frac{1}{28} \quad \text{and therefore,} \quad P(D− | E−) = \frac{.81}{.84} = \frac{27}{28}.
\]

A random nonsmoker has a 1 out of 28 (i.e., 3.6%) probability of having lung cancer;
a random nonsmoker has a 27 out of 28 (i.e., 96.4%) probability of not having lung cancer.

Therefore, the **odds** of the disease, given no exposure, is \(\frac{1}{28} \times \frac{27}{28} = \frac{1}{27}\).

The probability that a random nonsmoker has lung cancer is \(\frac{1}{27} = .037\) times the probability that he/she does not have it.

Or equivalently,

The probability that a random nonsmoker does not have lung cancer is 27 times greater than the probability that he/she does have it.

#### Odds Ratio: \(OR\)

\[
OR = \frac{\text{odds}(D+ | E+)}{\text{odds}(D+ | E−)} = \frac{3/4}{1/28} = \frac{81}{1} = 81.
\]

The odds of having lung cancer among smokers are 81 times greater than the odds of having lung cancer among nonsmokers.

#### Relative Risk: \(RR\)

\[
RR = \frac{P(D+ | E+)}{P(D+ | E−)} = \frac{3/4}{1/28} = \frac{81}{1} = 21.
\]

The probability of having lung cancer among smokers is 21 times greater than the probability of having lung cancer among nonsmokers.

The findings that \(OR \gg 1\) and \(RR \gg 1\) suggest a strong association between lung cancer and smoking. (But how do we formally show that this is significant? Later…)}
Now consider measures of association between lung cancer and caffeine consumption.

| Caffeine | Diseased ($D^+$) | Nondiseased ($D^-$) | $P(D^+ | E^+)$ | $P(D^+ | E^-$) |
|----------|------------------|---------------------|----------------|---------------|
| Exposed ($E^+$) | .06 | .34 | .15 | .85 |
| Not Exposed ($E^-$) | .09 | .51 | .15 | .85 |

$P(D^+ | E^+)$, the probability of having lung cancer among caffeine consumers, is equal to the probability of having lung cancer among caffeine non-consumers. 

### Odds Ratio: $OR$

$$OR = \frac{\text{odds}(D^+ | E^+)}{\text{odds}(D^+ | E^-)} = \frac{.15}{.15} = 1.$$  

The odds of having lung cancer among caffeine consumers are equal to the odds of having lung cancer among caffeine non-consumers.

### Relative Risk: $RR$

$$RR = \frac{P(D^+ | E^+)}{P(D^+ | E^-)} = \frac{.15}{.15} = 1.$$  

The probability of having lung cancer among caffeine consumers is equal to the probability of having lung cancer among caffeine non-consumers.

**NOTE:** The findings that $OR = 1$ and $RR = 1$ are to be expected, since $D^+$ and $E^+$ are independent! Thus, no association exists between lung cancer and caffeine consumption. (In truth, there actually is a spurious association, since many coffee drinkers also smoke, which commonly leads to lung cancer. In this context, smoking is a variable that confounds the association between lung cancer and caffeine, and should be adjusted for. For a well-known example of a study where this was not done carefully enough, with substantial consequences, see MacMahon B., Yen S., Trichopoulos D., et. al., Coffee and Cancer of the Pancreas, *New England Journal of Medicine*, March 12, 1981; 304: 630-33.)
Adjusting for Age (and other confounders)

Once again, consider the association between lung cancer and smoking in the earlier example. A legitimate argument can be made that the reason for such a high relative risk ($RR = 21$) is that age is a confounder that was not adequately taken into account in the study. That is, there is a naturally higher risk of many cancers as age increases, regardless of smoking status, so “How do you tease apart the effects of age versus smoking, on the disease?” The answer is to adjust, or standardize, for age. First, recall that relative risk $RR = \frac{P(D+ | E+)}{P(D+ | E−)}$ by definition, i.e., we are confining our attention only to individuals with disease ($D+$), and measuring the effect of exposure ($E+$ vs. $E$–). Therefore, we can restrict our analysis to the two cells in the first column of the previous $2 \times 2$ table. However, suppose now that the probability estimates are stratified on age, as shown:

<table>
<thead>
<tr>
<th>Age</th>
<th>$n_i^+$</th>
<th>$x_i^+$</th>
<th>$p_i^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>250</td>
<td>5</td>
<td>$5/250 = .02$</td>
</tr>
<tr>
<td>60-69</td>
<td>150</td>
<td>15</td>
<td>$15/150 = .10$</td>
</tr>
<tr>
<td>70-79</td>
<td>100</td>
<td>40</td>
<td>$40/100 = .40$</td>
</tr>
<tr>
<td>Total</td>
<td>$500$</td>
<td>$60$</td>
<td>$60/500 = .12$ (as before)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>$n_i^-$</th>
<th>$x_i^-$</th>
<th>$p_i^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>300</td>
<td>3</td>
<td>$3/300 = .01$</td>
</tr>
<tr>
<td>60-69</td>
<td>200</td>
<td>8</td>
<td>$8/200 = .04$</td>
</tr>
<tr>
<td>70-79</td>
<td>100</td>
<td>7</td>
<td>$7/100 = .07$</td>
</tr>
<tr>
<td>Total</td>
<td>$600$</td>
<td>$18$</td>
<td>$18/600 = .03$ (as before)</td>
</tr>
</tbody>
</table>

For each age stratum ($i = 1, 2, 3$),

$n_i^+ = \#$ individuals in the study who were exposed ($E+$), regardless of disease status

$n_i^- = \#$ individuals in the study who were not exposed ($E$–), regardless of disease status

$x_i^+ = \#$ of exposed individuals ($E+$), with disease ($D+$)

$x_i^- = \#$ of unexposed individuals ($E$–), with disease ($D+$)

Therefore,

$p_i^+ = x_i^+ / n_i^+ = \text{proportion of exposed individuals (E+), with disease (D+)}$

$p_i^- = x_i^- / n_i^- = \text{proportion of unexposed individuals (E–), with disease (D+)}$
From this information, we can imagine a combined table of age strata for $D+$:

<table>
<thead>
<tr>
<th>Age</th>
<th>$n_i = n_i^+ + n_i^-$</th>
<th>$p_i^+$</th>
<th>$p_i^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>550</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>60-69</td>
<td>350</td>
<td>.10</td>
<td>.04</td>
</tr>
<tr>
<td>70-79</td>
<td>200</td>
<td>.40</td>
<td>.07</td>
</tr>
<tr>
<td>Total</td>
<td>$n = 1100$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now, to estimate the “age-adjusted” numerator $P(D^+ \mid E^+)$ of $RR$, we calculate the weighted average of the proportions $p_i^+$, using their corresponding combined sample sizes $n_i$ as the weights. That is,

$$P(D^+ \mid E^+) \approx \frac{\sum n_i p_i^+}{\sum n_i} = \frac{(550)(.02) + (350)(.10) + (200)(.40)}{550 + 350 + 200} = \frac{126}{1100} = 0.1145$$

and similarly, the “age-adjusted” denominator $P(D^+ \mid E^-)$ of $RR$ is estimated by the weighted average of the proportions $p_i^-$, again using the same combined sample sizes $n_i$ as the weights:

$$P(D^+ \mid E^-) \approx \frac{\sum n_i p_i^-}{\sum n_i} = \frac{(550)(.01) + (350)(.04) + (200)(.07)}{550 + 350 + 200} = \frac{33.5}{1100} = 0.0305$$

whereby we obtain

$$RR_{adj} = \frac{P(D^+ \mid E^+)}{P(D^+ \mid E^-)} = \frac{126}{33.5} = 3.76.$$ 

Note that in this example, there is a substantial difference between the adjusted and unadjusted risks. The same ideas extend to the “age-adjusted” odds ratio $OR_{adj}$. 