DATA2002

Measures of performance

Garth Tarr

THE UNIVERSITY OF SYDNEY

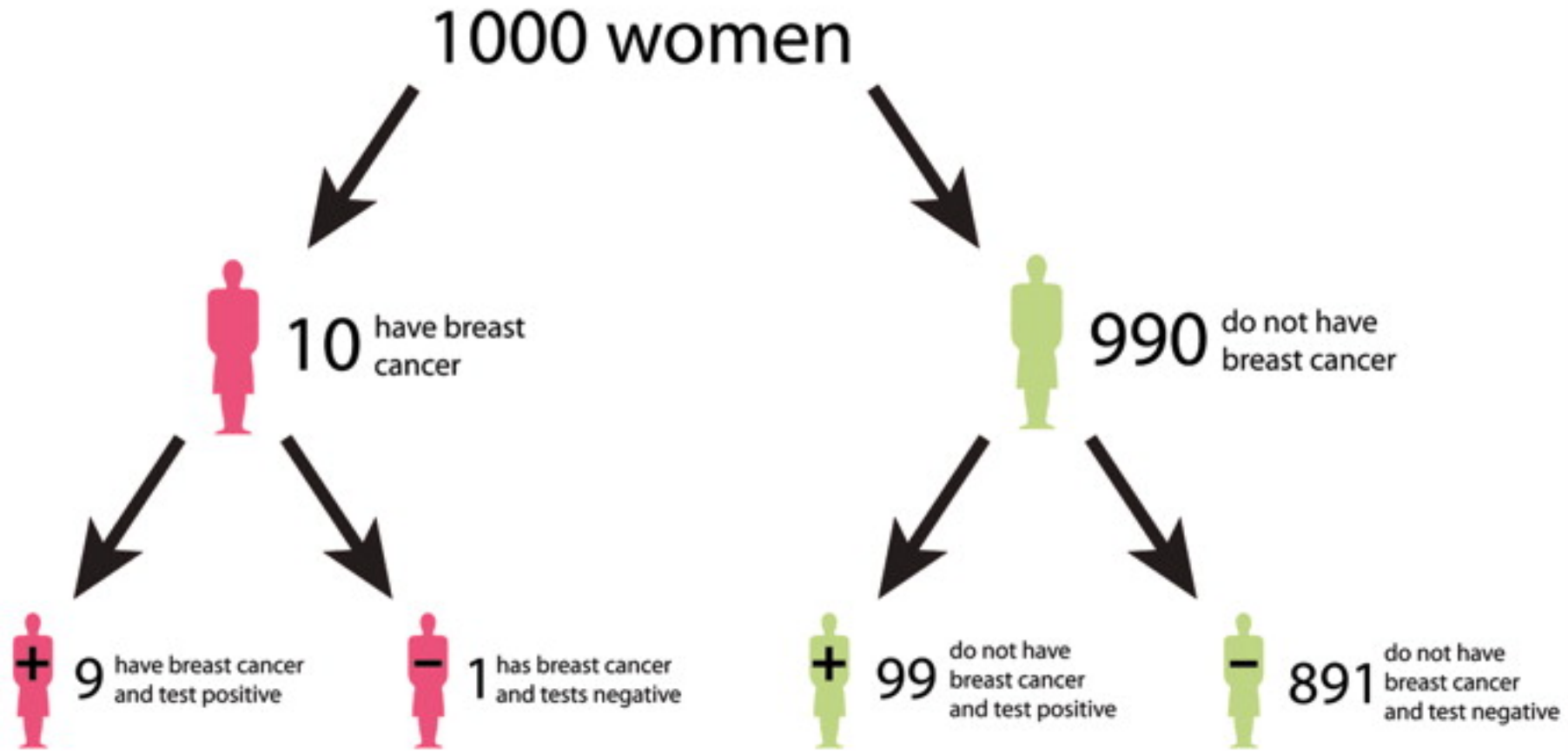
Types of errors

Conditional probability

Evaluating classification models

Types of errors

Breast cancer



? How **good** is the test for breast cancer? How you define **good**? If you test positive, what are the chances that you actually have breast cancer?

Test result vs actual status

In general, **positive** = identified and **negative** = rejected.

Therefore:

True positive = correctly identified ✓

False positive = incorrectly identified ✗

True negative = correctly rejected ✓

False negative = incorrectly rejected ✗

We can summarise this in a table:

	Actual positive	**Actual negative**
Test positive	True positive (TP) ✓	False positive (FP) ✗
Test negative	False negative (FN) ✗	True negative (TN) ✓

W False positives and false negatives

Breast cancer

	Have breast cancer	Do not have breast cancer
Mammogram returns a positive test result	True positive (TP) ✓	False positive (FP) ✗
Mammogram returns a negative test result	False negative (FN) ✗	True negative (TN) ✓

True positive: Mammogram returns a positive test result and you actually have breast cancer

False negative: Mammogram returns a negative test result but you do actually have breast cancer

False positive: Mammogram returns a positive test result but you don't actually have breast cancer

True negative: Mammogram returns a negative test result and you don't actually have breast cancer

What's worse? FN or FP?

Extra notation

Let's formalise this. Let,

- D^+ be the event that an individual **has** a particular disease. The **prevalence** is the marginal probability of disease, $P(D^+)$.
- D^- be the event that an individual **does not have** a particular disease.
- S^+ represent a positive screening test result or a symptom being present.
- S^- represent a negative screening test result or no symptom.

With this notation, we can start to quantify the influence of a risk factor or screening test (S) on the disease outcome (D).

Our table becomes:

	Actual positive D^+	**Actual negative D^-**
Test positive S^+	True positive (TP) ✓	False positive (FP) ✗
Test negative S^-	False negative (FN) ✗	True negative (TN) ✓

Table with counts of outcomes

	Actual positive D^+	Actual negative D^-	
Test positive S^+	a	b	$a+b$
Test negative S^-	c	d	$c+d$
	$a+c$	$b+d$	$a+b+c+d$

- **False negative rate**: $P(S^-|D^+) = \frac{c}{a+c}$
- **False positive rate**: $P(S^+|D^-) = \frac{b}{b+d}$
- **Sensitivity/Recall**: $P(S^+|D^+) = \frac{a}{a+c}$, the probability that the test is positive given that the subject actually has the disease
- **Specificity**: $P(S^-|D^-) = \frac{d}{b+d}$, the probability that the test is negative given that the subject does not have the disease
- **Positive predictive value/Precision**: $P(D^+|S^+) = \frac{a}{a+b}$, the probability that the subject has the disease given that the test is positive
- **Negative predictive value**: $P(D^-|S^-) = \frac{d}{c+d}$, the probability that the subject does not have the disease given that the test is negative
- **Accuracy**: $\frac{a+d}{a+b+c+d}$

𝕎 Sensitivity and specificity and 𝕎 Positive and negative predictive values

Breast cancer

	Breast cancer D^+	No breast cancer D^-	
Test positive S^+	9	99	108
Test negative S^-	1	891	892
	10	990	1000

Prevalence: 1% (10/1000)

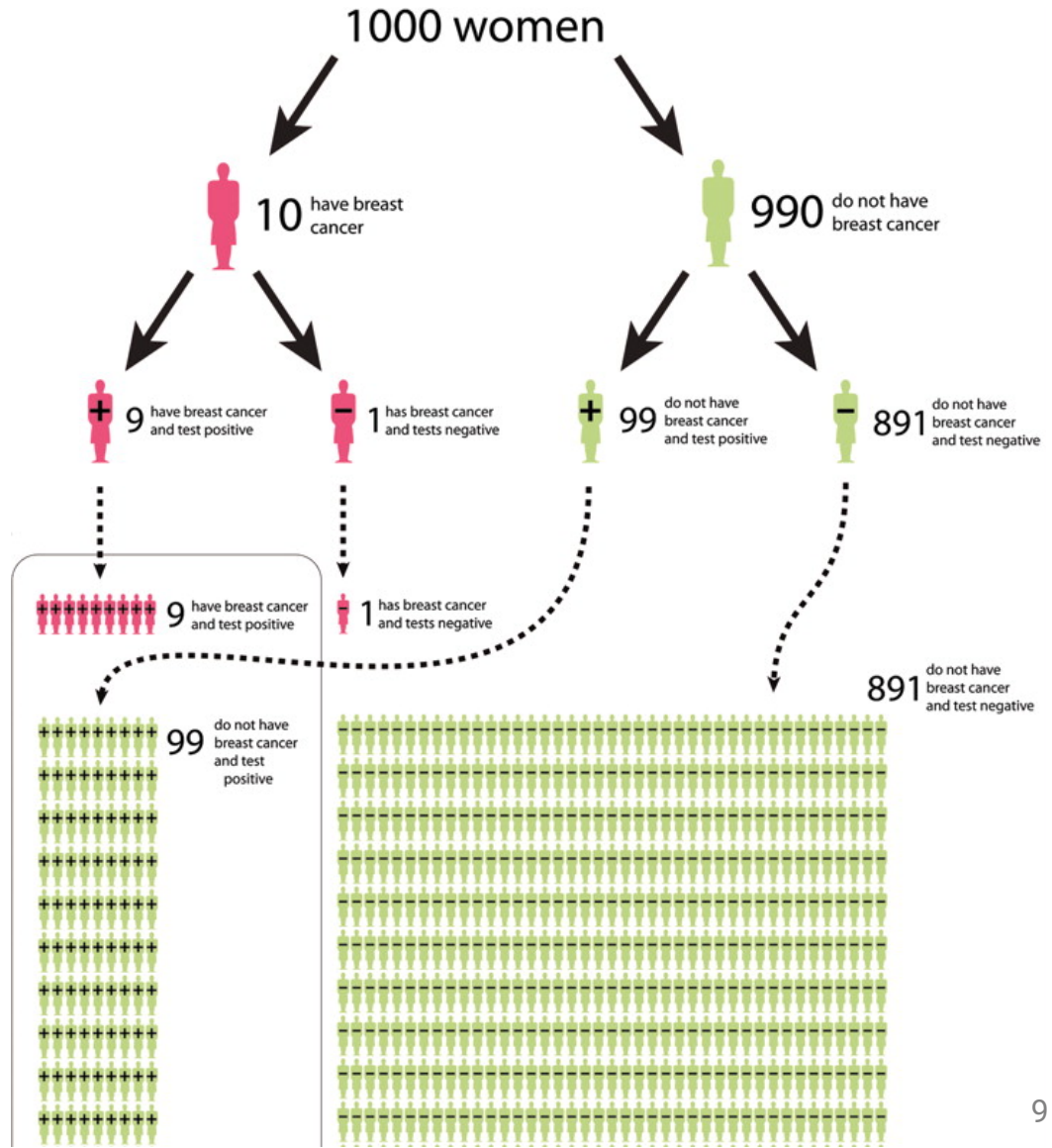
Sensitivity: 90% (9/10)

Specificity: 90% (891/990)

Positive predictive value (or precision)

$$P(D^+|S^+) = 9/(9+99) = 0.08$$

Image source: Spiegelhalter, Pearson, and Short (2011)



Conditional probability

Conditional probability: motivation

It's been suggested that lawn bowls is one of the most dangerous sports in the world (highest death rates).

Given that you're a young adult, should you be worried about playing?

Conditional on being a young adult, the risk of playing lawn bowls is very small.

Even though the probability of dying while playing lawn bowls is higher than many other sports, the overwhelming majority of lawn bowl related deaths are linked to the extreme age of the player.

Conditional probability: definition

- Let B be an event so that $P(B) > 0$
- The conditional probability of an event A given that B has occurred is

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

- Rearranging, we also have
  - $P(A \cap B) = P(A \mid B)P(B)$
  - $P(A \cap B) = P(B \mid A)P(A)$
- For the *classical definition of probability* where we have a finite number of equally likely outcomes, then
  - $P(A)$ is the *proportion* of outcomes in the subset A
  - $P(B)$ is the *proportion* of outcomes in the subset B
  - and $P(A|B)$ can be interpreted as *the proportion of outcomes in B that are also in A*.

Bayes' rule

Bayes' rule allows us to reverse the conditioning set provided that we know some marginal probabilities,

$$\begin{aligned} P(B \mid A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}. \end{aligned}$$

where B^c is the event that B doesn't occur.

Note that we've used the law of total probability

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$$

Example

A study comparing the efficacy of HIV tests, reports on an experiment which concluded that HIV antibody tests have a **sensitivity** of 99.7% and a **specificity** of 98.5%. Suppose that a subject, from a population with a .1% **prevalence** of HIV, receives a positive test result. What is the positive predictive value?

- Mathematically, we want $P(D^+ \mid S^+)$ given the sensitivity, $P(S^+ \mid D^+) = .997$, the specificity, $P(S^- \mid D^-) = .985$, and the prevalence $P(D^+) = .001$

$$P(D^+ \mid S^+) = \frac{P(S^+ \mid D^+)P(D^+)}{P(S^+ \mid D^+)P(D^+) + P(S^+ \mid D^-)P(D^-)}$$

Need to find:

- $P(D^-) = 1 - P(D^+) = 1 - 0.001 = 0.999$
- $P(S^+ \mid D^-) = 1 - P(S^- \mid D^-) = 1 - 0.985 = 0.015$

Using Bayes' formula

$$P(D^+ \mid S^+) = \frac{P(S^+ \mid D^+)P(D^+)}{P(S^+ \mid D^+)P(D^+) + P(S^+ \mid D^-)P(D^-)}$$
$$= \frac{.997 \times .001}{.997 \times .001 + .015 \times .999} = 0.062$$

In this population a positive test result only suggests a 6.2% probability that the subject has the disease I.e. the **positive predictive value** is 6.2% for this test.

- The low positive predictive value is due to low prevalence of disease and the somewhat modest specificity
- Suppose it was known that the subject was an intravenous drug user and routinely had intercourse with an HIV infected partner
- The evidence implied by a positive test result does not change because of the prevalence of disease in the subject's population, only our interpretation of that evidence changes

NIPT

Non-invasive prenatal tests (NIPT) are an increasingly popular way of screening for chromosome conditions and advertise as having very high **accuracy**.

Taylor-Phillips, Freeman, Geppert, et al. (2016) perform a meta-analysis on the accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down syndrome.

They broke their findings down by "general obstetric population" and "high risk propulation".

NIPT

They presented their findings as if they had taken a sample of 100,000 pregnancies from the **general obstetric population**:

- True positives: 417
- False positives: 94
- True negatives: 99471
- False negatives: 18

And a sample of 10,000 pregnancies from the **high risk population**:

- True positives: 324
- False positives: 31
- True negatives: 9636
- False negatives: 9

Calculate the sensitivity, specificity, precision and accuracy for the test in both the general population and the high risk population. Do your findings have any public health implications?

Evaluating classification models

Evaluating a classification model

Imagine a very simple classification model[1] looking to classify a particular type of deformation, kyphosis, in children who have had corrective spinal surgery (Chambers and Hastie, 1991).

The **outcome** is absence or presence of kyphosis after the operation.

The **predictor** is the number of the topmost vertebra operated on.

- If the highest vertebrae operated on was 9th or higher, we predict that kyphosis will be absent after surgery.
- If the highest vertebrae operated on was 8th or lower, we predict that kyphosis will be present after surgery.

The data can be found in the `kyphosis` data frame that comes with the **rpart** package.

1. We'll explore fitting and evaluating classification models in more detail later in semester.

Evaluating a classification model

```
## install.packages("rpart")
data(kyphosis, package = "rpart")
dplyr::glimpse(kyphosis)
```

```
## Rows: 81
## Columns: 4
## $ Kyphosis <fct> absent, absent, present, absent, absent, …
## $ Age      <int> 71, 158, 128, 2, 1, 1, 61, 37, 113, 59, 8…
## $ Number   <int> 3, 3, 4, 5, 4, 2, 2, 3, 2, 6, 5, 3, 5, 4,…
## $ Start    <int> 5, 14, 5, 1, 15, 16, 17, 16, 16, 12, 14, …
```

```
truth = kyphosis$Kyphosis
prediction = ifelse(kyphosis$Start >= 9,
                    "Predict absent",
                    "Predict present")
table(prediction, truth)
```

```
##                   truth
## prediction         absent present
##   Predict absent       56       6
##   Predict present       8      11
```

Evaluating a classification model

```
table(prediction, truth)
```

```
##                  truth
## prediction        absent present
##   Predict absent      56       6
##   Predict present      8      11
```

For our prediction model, define D^+, D^-, S^+, S^-, and find the

- sensitivity
- specificity
- false negative rate
- false positive rate
- positive predictive value
- negative predictive value

Reordering the rows and columns in the table might help!

References

Chambers, J. M. and T. J. Hastie (1991). *Statistical Models in S*. Chapman & Hall/CRC The R Series. Boca Raton, FL: Chapman and Hall/CRC. ISBN: 978-0412830402.

Spiegelhalter, D., M. Pearson, and I. Short (2011). "Visualizing Uncertainty About the Future". In: *Science* 333.6048, pp. 1393-1400. ISSN: 0036-8075. DOI: 10.1126/science.1191181.

Taylor-Phillips, S., K. Freeman, J. Geppert, A. Agbebiyi, O. A. Uthman, J. Madan, A. Clarke, S. Quenby, and A. Clarke (2016). "Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis". In: *BMJ Open* 6.1. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2015-010002.