STUDY DESIGNS

IN BIOMEDICAL RESEARCH

PREVALENCE & PREVALENCE SURVEY

This lecture covers a few fundamental issues in Diagnostic Medicine which are related to prevalence & prevalence survey:

(1) Random Testing
(2) Test-Retest
(3) Screening Efficiency
(4) Prevalence Survey

RANDOM TESTING

Should we conduct “random testing” for a rare disease, such as AIDS?

COMMON ARGUMENTS

- Those against the practice often cite concerns about errors, privacy and confidentiality, and “unwanted consequences” (such as job’s loss).
- Those promoting the practice, eg. policy makers, often want to know “the magnitude of the problem” in order to justify spending - on research as well as interventions.
- But what about scientific merit?

PREDICTIVE VALUES

Recall that both predictive values are functions of disease prevalence, π = Pr(D = +):

$$P^+ = \frac{S^+\pi}{S^+\pi + (1 - S^-)(1 - \pi)}$$

$$P^- = \frac{S^-(1 - \pi)}{S^-(1 - \pi) + (1 - S^+)\pi}$$

EXAMPLES: AIDS SCREENING

Example A: $S^+=.977$, $S^-=.926$, and $\pi=.003$:

$$P^+ = \frac{(.977)(.003)}{(.977)(.003)+(.074)(.997)} = .038 \text{ or } 3.8\%$$

Example B: $S^+=.977$, $S^-=.926$, and $\pi=.20$:

$$P^+ = \frac{(.977)(.20)}{(.977)(.20)+(.074)(.80)} = .767 \text{ or } 76.7\%$$

Note: Current Estimate for USA's AIDS: .3% as above and S^+ and S^- are for ELISA in Weiss, 1985.

The first example is random testing; in the second, it's testing a "high risk" subpopulation, e.g. drug abusers

IMPLICATION

- Predictive values of a screening test depend not only on sensitivity and specificity but on disease prevalence too.
- The higher the prevalence, the higher the positive predictive value; “random screening” or “random testing” might not do much good – many false positives.
- The higher the prevalence, the lower the negative predictive value (but the effect is much weaker for P^-, almost negligible) – in fact, we do not pay much attention to Negative Predictive Value

This agrees with math result:

$$P^+ = \frac{S^+\pi}{S^+\pi + (1 - S^-)(1 - \pi)}$$

$$\frac{dP^+}{d\pi} = \frac{S^+}{[S^+\pi + (1 - S^-)(1 - \pi)]^2}$$

Result: The higher the Prevalence, the higher the Positive Predictive Value.

TO SCREEN OR NOT TO SCREEN?
TEST & RE-TEST

ISSUE FOR BREAST CANCER SCREENING

- The need is not the issue; it decreases BC mortality by 32% (Tabar, 2000; from “the Swedish two-county trial”).
- The test “characteristics” may not be a major issues; sensitivity is low (Kuhl, 2000) but the specificity ranges from 93%-99.7% in high-risk women (Warner, 2001).
- But is forty or fifty “old enough”? (to be at “higher risk” for efficient screening)

SCREENING GUIDLINE

- There are guidelines, by federal panels and/or ACS, but are there any justification? Why 40? Why 50? Or, why not starting at 35?
- Here are some post-hoc overall data by ACS: about 10% or less* are "recalled" for more tests (because the first mammogram is "positive"); 8%-10% of those need biopsy – because mammogram is positive again, and 20% of those with biopsy have cancer. That puts the positive predictive value (of first test) at most 1.6%-2%.

HOW GOOD IS GOOD?

Some investigators imply that a "good test" must yield $P^{+} \geq 50\%$; by either improving its characteristics (S^{+} and S^{-}) or by selecting the population in which the test is used so that the background prevalence is higher.

But if you cannot improve (S^{+} and S^{-}); When Does It Make Sense to Screen?

When Does It Make Sense to Screen?

From:

$$P^+ = \frac{S^+\pi}{S^+\pi + (1-S^-)(1-\pi)}$$

We can "solve" for π:

$$\pi = \frac{(1-S^-)P^+}{S^+ + (1-S^-)P^+ - S^+P^+}$$

Then set a "desirable level" for P+ to obtain "screenable prevalence"

SCREENABLE PREVALENCE

For example, setting P^{+}=.80 or 80%

Table 2.1		Sensitivity, S+				
		0.5	0.9	0.95	0.98	0.99
Specificity, S-	0.5					
	0.9		0.308		0.29	
	0.95			0.174		
	0.98		0.082		0.075	
	0.99					

That is, if S^{+}=S^{-}=.98, we "attain" a positive predictive value of 80% if prevalence $\pi \geq .075$; much higher if the test is not that good.

Specificity has more influence on Screenable Prevalence: $\pi \geq .082$ when (s^{-}=.98,s^{+}=.90) but $\pi \geq .29$ when (s^{-}=.90,s^{+}=.98)

SOME RESULTS OF MEMMOGRAPHY

(Currently) $S^- = .966$, $S^+ = .647$ &

$$\pi = \frac{(1 - S^-)P^+}{(1 - S^-)P^+ + S^+(1 - P^+)}$$

Predictive Value, P	Screenable Prevalence
1%	53 per 100,000
2%	**107 per 100,000**
5%	**276 per 100,000**
10%	581 per 100,000

Prevalences from SEER:

Age Group	Rate
35-39	59 per 100,000
40-44	**119 per 100,000**
45-49	194 per 100,000
50-54	**254 per 100,000**
55-59	313 per 100,000

COMPETING STRATEGIES FOR BREAST CANCER SCREENING

Starting at age 40: Incidence Rate is about 119 per 100,000

Positive Predictive Value is 2%

Negative Predictive Value is 99.96%

Starting at age 50: Incidence Rate is about 254 per 100,000

Positive Predictive Value is 5%

Negative Predictive Value is 99.91%

Would it be justified to reduce from 50 to 40?

CAN IT BE IMPROVED?

- Even starting at 50, positive predictive value is only 5%
- By either improving its characteristics of the test or by selecting the population with higher prevalence, is that possible to reach that threshold of predictive value: $P^{+} \geq 50\%$?
- Improving characteristics of a test such as mammogram is possible but it would take years; in addition, its specificity is already rather high, it is possible to improve but not by much.
- What's about retest?

WHAT ABOUT A RE-TEST?

- If starting at age 40, and if “recalled”, the chance to have cancer would be about 2%. Another recall for biopsy would raise the predictive value (of two positives) to 28% (which is similar to ACS’ data of about 20% - perhaps including younger users).
- If starting at age 50, and if “recalled”, the chance to have cancer would be about 5%. Another recall for biopsy would raise the predictive value to 50%-51%; that qualifies it as a “good” procedure as stipulated by some investigators.

COMPETING STRATEGIES FOR BREAST CANCER SCREENING

Starting at age 40: Incidence Rate is about 119 per 100,000

Positive Predictive Value is 2%

Two recalls raise predictive value to 28%

Starting at age 50: Incidence Rate is about 254 per 100,000

Positive Predictive Value is 5%

Two recalls raise predictive value to 51%

Should Women Start Mammograms at Age 40 or 50?

- For those with "reason" to test, i.e. women with family history (mother or sisters with BC), decision is easier – and should be recommended (by age 50 it might be too late, more than 50% of the BRCA1 or BRCA2 mutation carriers have already developed the disease).
- For others, it may boil down to this not-very-simple question: are you prepared for unwanted consequences? At age 40, 98% of positive mammograms are false positives and, after another recall, 72% of biopsies are negative

SCREENING EFFICIENCY

Screening may save lives; but how do we measure the efficiency of a screening program?

Sensitivity, $S^+ = .647$;

For a woman with Breast Cancer:

(1) The chance to have a positive result is (.647); second mammogram is needed

(2) The chance tested positive twice is $(.647)^2$; biopsy is needed

Specificity, $S^- = .966$;

For a woman without Breast Cancer:

(1) The chance to have a positive result is (.034); second mammogram is needed

(2) The chance tested positive twice is $(.034)^2$; biopsy is needed.

If a woman is chosen randomly, as in mass screenings, the probability that she has breast cancer is p and does not have breast cancer is (1-p). The probability p depends on her age or age group which is available/estimated from some large population database – such as SEER.

"SEER" is Surveillance, Epidemiology, and End Results program of the National Cancer Institute (NCI); SEER collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 28% of the population of the United States. The SEER Program registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status. The SEER Program is the only comprehensive source of population-based information in the United States that includes stage of cancer at the time of diagnosis and patient survival data

STARTING AGE: 40

100,000 women age 40-44

119 with breast cancer (source: SEER)

77 identified by mammograms (sensitivity = .647)

50 identified as positive again (sensitivity = .647)

50 confirmed by biopsies (assume 100% rate): →treatment

12 died if all were not screened (assume 25% death rate)

4 would be saved by mammograms (32% rate by Tabar)

100,000 go through the process, 4 lives saved

100,000 go through the process, 4 lives saved

25,000 go through the process to save 1 life

Age 40:

NNS = 25,000

NNS: Number Needed Screening

STARTING AGE: 50

100,000 women age 50-54

254 with breast cancer (source: SEER)

164 identified by mammograms (sensitivity = .647)

106 identified as positive again (sensitivity = .647)

106 confirmed by biopsies (assume 100% rate): →treatment

27 died if all were not screened (assume 25% death rate)

8.5 would be saved by mammograms (32% rate by Tabar)

100,000 go through the process, 8.5 lives saved

100,000 go through the process, 8.5 lives saved

11,700 go through the process to save 1 life

Age 50:

NNS = 11,700

NNS: Number Needed Screening

NNS, Number Needed Screening (to save a life), is often used as a parameter characterizing a screening procedure. In the evaluation of treatment program, it's **NNT**, Number needed treatment (to save a life or a bad event – such as, for example, a fall with a broken bone).

PREVALENCE SURVEY

If you want to know how many percent of Minnesotans having no health insurance, you would survey n people, at random. If x of the n people in the sample have no health insurance, our estimate is x/n. This estimate, a proportion, is good – i.e. "unbiased"!

What if you want to estimate a disease prevalence? Say, what is the prevalence of HIV infection? or of breast cancer?

Well, you need a disease screening procedure.

But, use of a screening procedure involves errors, false positives and false negatives; is the result, the estimated disease prevalence, good? or if the estimate "unbiased"? (There are false positives and there are false negatives; would they cancel each other?)

AN SIMPLE ILLUSTRATION

Let assume we have great screening procedure but the target disease is rare (in real life, most diseases are rare):

- 98% sensitive
- 97% specific
- .1% prevalence

Example

	Infection=Yes	Infection=No	Total
Test=Positive			
Test=Negative			
Total	100	99900	100000

	Infection=Yes	Infection=No	Total
Test=Positive	98		
Test=Negative	2		
Total	100	99900	100000

	Infection=Yes	Infection=No	Total
Test=Positive		2997	
Test=Negative		96903	
Total	100	99900	100000

	Infection=Yes	Infection=No	Total
Test=Positive	98	2997	3095
Test=Negative	2	96903	90905
Total	100	99900	100000

RESULT

- True prevalence: 100/100,000 = .1%
- Estimated prevalence: 3,095/100,000 = 3.1%
- Not good, we over-estimate it!

	Infection=Yes	Infection=No	Total
Test=Positive	98	2997	3095
Test=Negative	2	96903	90905
Total	100	99900	100000

2% false negative rate applies to 100 diseased persons versus 3% false positive rate applies to 99,900 healthy persons! (still many more false positives even if the false negative rate is 40%)

Is there any way we can improve?

SETTING:

- It's a very simple design
- We have a screening test T; its sensitivity S^+ and specificity S^- have been independently established.
- A "prevalence survey" is conducted in one target population in order to estimate the disease prevalence, $\pi = Pr(D=+)$.
- Data: x of n subjects found "positive".

Simple Solution?

Does this work: to estimate the disease prevalence by the frequency of positive tests: $p_t = x/n$ – ignoring its errors?

This is a good estimate but it is an estimate of $\pi_t = Pr(T=+)$, the "response rate" whereas we want to estimate the disease prevalence, $\pi = Pr(D=+)$. It is good estimator but for a wrong parameter!

How good is p_t as an estimate of prevalence π?

It can be shown that estimator p_t depends not only on disease the prevalence (that it is supposed to estimate) but also on the characteristics of the test, S^+ and S^-. It is badly biased.

It is biased upward, overestimating π

First, we want to establish a relationship between Response Rate, Disease Prevalence, Sensitivity, and Specificity.

$$
\begin{aligned}
\pi_t &= \Pr(T = +) \\
&= \Pr(T = +, D = +) + \Pr(T = +, D = -) \\
&= \Pr(T = + \mid D = +)\Pr(D = +) + \Pr(T = + \mid D = -)\Pr(D = -) \\
&= S^{+}\pi + (1 - S^{-})(1 - \pi) \\
&= \pi + (1 - S^{-})(1 - \pi) - (1 - S^{+})\pi
\end{aligned}
$$

THE BIAS

$$\pi_t = \pi + (1 - S^-)(1 - \pi) - (1 - S^+)\pi$$

$$= E(p_t)$$

$$\mathbf{Bias} = \mathbf{E(p_t) - \pi}$$

$$= \boldsymbol{\pi_t - \pi}$$

$$= \mathbf{(1 - S^-)(1 - \pi) - (1 - S^+)\pi}$$

EXAMPLE

$$E(p_t) = \pi + (1 - S^-)(1 - \pi) - (1 - S^+)\pi$$

$$\text{Bias} = (1 - S^-)(1 - \pi) - (1 - S^+)\pi$$

If $(1-\pi) > \pi$, p_t is most likely biased upward;

Example, for $S^+ = S^- = .9$ and $\pi = .1$;
bias = (.1)(.9)-(.1)(.1) = .08 which is more affected by specificity. Here, the point estimate could be twice the true value.

A NEW POINT ESTIMATE

$$\pi_t = \Pr(T = +) = \Pr(T = +, D = +) + \Pr(T = +, D = -)$$

$$\pi_t = \Pr(T = + \mid D = +)\Pr(D = +) + \Pr(T = + \mid D = -)\Pr(D = -)$$

$$\boldsymbol{\pi_t = S^+\pi + (1 - S^-)(1 - \pi)}$$

$$\boldsymbol{\pi = \frac{\pi_t + S^- - 1}{J}}; J = S^+ + S^- - 1, \text{leading to}$$

$$\boldsymbol{p = \frac{p_t + S^- - 1}{J}}$$

J is called the Youden's Index

EXAMPLE

For example, for $S^{+}=S^{-}=.9$ and $\pi = .1$;

$$\pi_t = S^{+}\pi + (1 - S^{-})(1 - \pi)$$

$$= .18$$

Let suppose $x = 20$ out of

$n = 100$ subjects are positive

$$\mathbf{p_t = .20}$$

$$\mathbf{p = \frac{p_t + S^{-} - 1}{J} = .125}$$

A correction, using p instead of p_t , is a substantial improvement; in addition, if S^+ and S^- are known apriori (without errors), then p is unbiased for π.

$$\pi = \frac{\pi_t + S^- - 1}{J}; J = S^+ + S^- - 1$$

$$p = \frac{p_t + S^- - 1}{J}$$

$$\mathbf{E(p) = \frac{\pi_t + S^- - 1}{J}}$$

$$\mathbf{= \pi}$$

EXAMPLE

- Stamler et al. (1976) surveyed one million people and found 24.7% had DBP>90 mm Hg and 11.6% had DBP>95 mm Hg – using p_t, of course.
- Carey et al. (1976) using elevation of BP in 3 separate readings as the criterion for having hypertension (the "truth") and found $S^+ = .930$ and $S^- = .911$; good characteristics.
- Yet, correcting p_t to get p shows dramatic results: Stamler 24.7% becomes 18.8% and 11.6% becomes 3.2% - estimates p_t and p can differ by a factor of 4!

Measuring “Blood Pressure” is an interesting case where we do not really have “gold standard”; sensitivity and specificity are determined using repeated measures (to establish gold standard)

$$p = \frac{p_t + s^- - 1}{j}$$

$$S^+ = .930;\ S^- = .911$$

$$J = S^+ + S^- - 1 = .841$$

$$p_t = .116$$

$$p = \frac{.116 + .911 - 1}{.841} = .032$$

STANDARD ERROR, SE(p)

$$p = \frac{p_t + S^- - 1}{J}$$

$$Var(p) = \frac{Var(p_t)}{J^2}$$

$$\mathbf{SE(p)} = \frac{\mathbf{1}}{\mathbf{J}} \sqrt{\frac{\mathbf{p_t(1-p_t)}}{\mathbf{n}}}$$

Suggested Exercises:

#1 Calculate the screenable prevalence when:
(a) $S^+=S^-=.99$, and (b) $S^+=S^-=.90$

#2 For the case of Mammograms ($S^-=.966$, $S^+=.647$) for women in the 50's, the positive predictive value of the first positive recall is 5%, verify that two recalls would raise the predictive value to 51%