# STUDY DESIGNS

## IN BIOMEDICAL RESEARCH

## DESIGNS IN DIAGNOSTIC MEDICINE

# PREVALENCE SURVEY
# Revisited

# SIMPLE 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".

# 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**

**BIAS**: If $S^+$ and $S^-$ are known apriori (without errors), then p is unbiased for $\pi$.

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

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

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

$$= \boldsymbol{\pi}$$

# STANDARD ERROR

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

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

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

In some problems, sensitivity and specificity are not known; they need to be estimated but data from the “main survey” are not enough for this purpose.

# REALISTIC SETTING:

- We have a screening test T; but its sensitivity $S^+$ and specificity $S^-$ are not known.
- The main "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".

The main survey is still a simple design with as much data. However, Sensitivity and specificity are not known and need to be estimated but there are not enough data from the main “prevalence survey” to do so.

Sensitivity and specificity are estimated using two other independent surveys/samples; $S^+$ is estimated by the proportion $s^+$ from a sample of size $n_1$ , and $S^-$ is estimated by the proportion $s^-$ from a sample of size $n_0$. These two precede and supplement the main prevalence survey.

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

We use the same estimator “p” of the simple design with $S^+$ and $S^-$ being estimated from two independent samples by proportions $s^+$ and $s^-$

# Using Taylor series expansion, we can approximate Expected Value and Variance of estimator p

$$E(p) = \pi + \frac{\pi}{J^2}\frac{S^+(1-S^+)}{n_1} - \frac{(1-\pi)}{J^2}\frac{S^-(1-S^-)}{n_0}$$

$$Var(p) = \frac{1}{J^2}\frac{p_t(1-p_t)}{n} + \frac{\pi^2}{J^2}\frac{S^+(1-S^+)}{n_1} + \frac{(1-\pi)^2}{J^2}\frac{S^-(1-S^-)}{n_0}$$

# SOURCES OF BIAS

**When sensitivity and specificity are unknown and are estimated using two other independent samples, "p" is no longer unbiased; as seen from the last two terms of the following formula, the bias come from the estimation of the sensitivity and specificity.**

$$E(p) = \pi + \frac{\pi}{J^2}\frac{S^+(1-S^+)}{n_1} - \frac{(1-\pi)}{J^2}\frac{S^-(1-S^-)}{n_0}$$

$$E(p) = \pi + \frac{\pi}{J^2}\frac{S^+(1-S^+)}{n_1} - \frac{(1-\pi)}{J^2}\frac{S^-(1-S^-)}{n_0}$$

**The bias is negligible if the other two samples $n_1$ and $n_0$ are both large; denominators are much larger than numerators. In other words, the estimator p is asymptotically unbiased.**

# SOURCES OF VARIABILITY

- The first term of the variance due to the prevalence survey itself.
- The last two terms due to our need of estimating sensitivity and specificity.

$$\mathrm{Var}(p) = \frac{1}{J^2}\frac{p_t(1-p_t)}{n} + \frac{\pi^2}{J^2}\frac{S^+(1-S^+)}{n_1} + \frac{(1-\pi)^2}{J^2}\frac{S^-(1-S^-)}{n_0}$$

# PRIORITIES

- Comparing the last two terms in Var(p); in common cases where both $S^+$ and $S^-$ are high but $\pi$ is low, the last term is much larger than the second term.
- That means in the contribution of the two supplement surveys to the variability of the prevalence estimate, the survey to estimate the specificity is dominating.

$$Var(p) = \frac{1}{J^2}\frac{p_t(1-p_t)}{n} + \frac{\pi^2}{J^2}\frac{S^+(1-S^+)}{n_1} + \frac{(1-\pi)^2}{J^2}\frac{S^-(1-S^-)}{n_0}$$

That is, if we have choices, we should focus on a precise estimation of the specificity – i.e. having larger $n_0$.

The question is how much larger?

# OPTIMAL ALLOCATION

- The sensitivity and the specificity have different effects on the estimated disease prevalence and its precision.
- If we regard the sum $m=n_1+n_0$ as fixed and find the choices $n_1$ and $n_0$ which minimize the sum of the last two terms in Var(p); result is:

$$\frac{n_1}{n_0} \cong \frac{\pi}{1-\pi}\sqrt{\frac{S^+(1-S^+)}{S^-(1-S^-)}}$$

## EXAMPLE:

$$\frac{n_1}{n_0} \cong \frac{\pi}{1-\pi}\sqrt{\frac{S^+(1-S^+)}{S^-(1-S^-)}}$$

For example, let $S^+=S^-=.93$, and $\pi = .05$, then $n_0$ should be 19 times as large as $n_1$

$$\frac{n_1}{n_0} \cong \frac{\pi}{1-\pi}\sqrt{\frac{S^+(1-S^+)}{S^-(1-S^-)}} = \frac{(.05)}{(.95)}\sqrt{\frac{(.93)(.07)}{(.93)(.07)}} = \frac{1}{19}$$

In practice, it is easier to find controls than cases. However, realistically, this optimal allocation is very hard to achieve; one survey could be 4 or 5 times larger but not 19 times.

# EXAMPLE

- The ELISA test for AIDS is used to screen donated blood for blood banks.
- An evaluation of ELISA yielded estimates (Weiss, 1985): $s^{+}$=.977 (using $n_1$=88) and $s^{-}$=.926 (using $n_0$=297) – allocation was in the right direction but not yet optimal.

# COMPARISON OF SCREENING TESTS

In a “comparison” we want to see if two tests, usually a new versus a more established one, have the same performance using “statistical test or tests of significance”. The comparison is relatively easy statistically; the more difficult problem is how “express” the “level of difference” if the two screening tests do not have the same level of performance (i.e. statistical test is significant).

# STUDY DESIGNS

- Decisions about which test or tests to recommend for widespread use and which to abandon, assuming that more than one are acceptable, are made on the basis of research studies that compare the accuracies of the tests.
- If each study subject is tested by all tests, we refer to as “paired design”, even more than two tests are under consideration.
- If each study subject is tested by one test, we will refer to the design as “unpaired”.

# USE OF UNPAIRED DESIGNS

- If tests are invasive, cause the patient discomfort, time consuming, or have significant risk associated with them,
- If ethical considerations require unpaired designs be made to minimize the burden on the subject.
- If the performance of one test might interfere with the implementation and/or the result of another; for example, two surgical procedures.
- If we only have observational data

# UNPAIRED DESIGNS

- Follow the same design principles of multi-arm randomized clinical trials.
- Those include well-defined inclusion-exclusion criteria, clear apriori definition of disease and test result - including measurement scale, preparation of study protocol, and randomization to ensure that study arms are balanced with regards to factors affecting test performance and/or result; analysis plan must be in place.
- Blinding – if feasible- may be needed to ensure integrity of disease and test assessments.

# DATA LAYOUT

| CASES | Test Result | | |
|---|---|---|---|
| Test | negative (T=-) | positive (T=+) | Total |
| A | n10(A) | n11(A) | n1(A) |
| B | n10(B) | n11(B) | n1(B) |
| | | | |
| CONTROLS | Test Result | | |
| Test | negative (T=-) | positive (T=+) | Total |
| A | n00(A) | n01(A) | n0(A) |
| B | n00(B) | n01(B) | n0(B) |

**There are four groups of subjects**

# COMPARISON OF TESTS WITH BINARY ENDPOINT

We can perform two separate Chi-square tests, one for cases and one for controls; for an overall level of $\alpha$, each test is performed at $\alpha$ /2 (Also, degree of freedom depends on the number of diagnostic tests involved).

# MEASURING DIFFERENCES

- If the difference between two diagnostic tests are found to be significant; the level of difference should be summarized and presented.
- The two commonly used parameters are the ratio of two sensitivities ($RS^+$) and the ratio of two specificities ($RS^-$); these are ratio of independent proportions, variances are calculated as follows.

# RATIO OF PROPORTIONS

$$r = \frac{p_2}{p_1}$$

$$\ln r = \ln p_2 - \ln p_1$$

$$Var(\ln r) = Var(\ln p_2) + Var(\ln p_1)$$

$$Var(\ln r) \cong \frac{1}{p_2^2}\frac{p_2(1-p_2)}{n_2} + \frac{1}{p_1^2}\frac{p_1(1-p_1)}{n_1}$$

$$\cong \frac{1-p_2}{n_2 p_2} + \frac{1-p_1}{n_1 p_1}$$

# RATIOS OF SENSITIVITIES & SPECIFICITIES

$$RS^{+}(A,B) = \frac{S_A^{+}}{S_B^{+}}$$

$$Var\{\ln\ RS^{+}(A,B)\} = \frac{1-S_A^{+}}{n_1(A)} + \frac{1-S_B^{+}}{n_1(B)}$$

$$RS^{-}(A,B) = \frac{S_A^{-}}{S_B^{-}}$$

$$Var\{\ln\ RS^{-}(A,B)\} = \frac{1-S_A^{-}}{n_0(A)} + \frac{1-S_B^{-}}{n_0(B)}$$

# EXAMPLE

**Data from Pepe's book: a randomized study of chronic villus sampling (CVS: Test B) versus early amniocentesis (EA: Test A) for fetus abnormality ( Disease D)**

| CASES | Test Result | | |
|---|---|---|---|
| Test | negative (T=-) | positive (T=+) | Total |
| EA | 6 | 116 | 122 |
| CVS | 13 | 11 | 124 |
| | | | |
| CONTROLS | Test Result | | |
| Test | negative (T=-) | positive (T=+) | Total |
| EA | 4844 | 34 | 4878 |
| CV | 4765 | 111 | 4876 |

# TESTS OF SIGNIFICANCE

| CASES | Test Result | | |
|---|---|---|---|
| Test | negative (T=-) | positive (T=+) | Total |
| EA | 6 | 116 | 122 |
| CVS | 13 | 111 | 124 |
| | | | |
| CONTROLS | Test Result | | |
| Test | negative (T=-) | positive (T=+) | Total |
| EA | 4844 | 34 | 4878 |
| CV | 4765 | 111 | 4876 |

$$\text{Cases}: \chi^2 = \frac{246\{(6)(111)-(13)(116)\}^2}{(19)(127)(122)(124)} = 4.78$$

$$\text{Controls}: \chi^2 = \frac{9754\{(4844)(11)-(4765)(34)\}^2}{(9609)(145)(4878)(4876)} = 41.54$$

# CONFIDENCE INTERVALS

| CASES | Test Result | | |
|---|---|---|---|
| Test | negative (T=-) | positive (T=+) | Total |
| EA | 6 | 116 | 122 |
| CVS | 13 | 111 | 124 |
| | | | |
| CONTROLS | Test Result | | |
| Test | negative (T=-) | positive (T=+) | Total |
| EA | 4844 | 34 | 4878 |
| CV | 4765 | 111 | 4876 |

$$RS^{+}(A,B) = \frac{S_A^{+}}{S_B^{+}}$$

$$Var\{\ln RS^{+}(A,B)\} = \frac{1-S_A^{+}}{n_1(A)} + \frac{1-S_B^{+}}{n_1(B)}$$

$$RS^{-}(A,B) = \frac{S_A^{-}}{S_B^{-}}$$

$$Var\{\ln RS^{-}(A,B)\} = \frac{1-S_A^{-}}{n_0(A)} + \frac{1-S_B^{-}}{n_0(B)}$$

$$RS^{+} = \exp\{\ln\frac{.951}{.895} \pm (2.28)\sqrt{\frac{1-.951}{122} + \frac{1-.895}{124}}\} = (.981,1.151)$$

$$RS^{-} = \exp\{\ln\frac{.993}{.977} \pm (2.28\sqrt{\frac{1-.993}{4878} + \frac{1-.977}{4876}}\} = (1.007,1.024)$$

**We should use +/-1.96 if each test is at 5%**

# PAIRED DESIGNS

- If feasible, paired designs are more desirable.
- Most important, only valid if tests do not interfere with each other; be cautious because interference can be subtle.
- Also paying attention to cooperation of the subjects; “order” should/may be randomized.

# ADVANTAGES

- More efficient because impact of between-subject variability is minimized.
- Possibilities of confounding are eliminated,
- One can examine characteristics of subjects where tests yield different results; this can lead to insight about test performance and, sometimes, strategies for improving tests.
- One can assess the value of applying combinations of tests compared to single tests.

# DATA LAYOUT

| CASES | | Test A Result | | Total |
|---|---|---|---|---|
| | | negative | positive | |
| Test B Result | negative | a1 | b1 | n10(B) |
| | positive | c1 | d1 | n11(B) |
| Total | | n10(A) | n11(A) | n1 |
| | | | | |
| CONTROLS | | Test A Result | | Total |
| | | negative | positive | |
| Test B Result | negative | a0 | b0 | n00(B) |
| | positive | c0 | d0 | n01(B) |
| Total | | n00(A) | n01(A) | n0 |

**One group of cases and one group of controls**

| CASES | | Test A Result | | Total |
|---|---|---|---|---|
| | | negative | positive | |
| Test B Result | negative | a1 | b1 | n10(B) |
| | positive | c1 | d1 | n11(B) |
| Total | | n10(A) | n11(A) | n1 |
| | | | | |
| CONTROLS | | Test A Result | | Total |
| | | negative | positive | |
| Test B Result | negative | a0 | b0 | n00(B) |
| | positive | c0 | d0 | n01(B) |
| Total | | n00(A) | n01(A) | n0 |

**The marginal frequencies for a paired design correspond to the entries for an unpaired design.**

# COMPARISON OF TESTS WITH BINARY ENDPOINT

We can perform two separate Chi-square tests, one for cases and one for controls; for an overall level of $\alpha$, each test is performed at $\alpha$ /2 (Also, degree of freedom depends on the number of diagnostic tests involved).

# Not regular Chi-Square test; it's McNemar Chi-Square test

# MEASURING DIFFERENCES

- If the difference between two diagnostic tests are found to be significant; the level of difference should be summarized and presented.
- The two commonly used parameters are the ratio of two sensitivities ($RS^+$) and the ratio of two specificities ($RS^-$); these are ratio of independent proportions, variances are calculated as follows.

| CASES | | Test A Result | | Total |
|---|---|---|---|---|
| | | negative | positive | |
| Test B Result | negative | a1 | b1 | n10(B) |
| | positive | c1 | d1 | n11(B) |
| Total | | n10(A) | n11(A) | n1 |
| | | | | |
| CONTROLS | | Test A Result | | Total |
| | | negative | positive | |
| Test B Result | negative | a0 | b0 | n00(B) |
| | positive | c0 | d0 | n01(B) |
| Total | | n00(A) | n01(A) | n0 |

**Note: term $d_1$ are in both numerator and denominator.**

$$S_A^+ = n_{11}(A) / n_1$$

$$S_B^+ = n_{11}(B) / n_1$$

$$RS^+ = \frac{S_A^+}{S_B^+} = \frac{b_1 + d_1}{c_1 + d_1}$$

| | Test A Result | | Total |
|---|---|---|---|
| | negative | positive | |
| negative | a1 | b1 | n10(B) |
| positive | c1 | d1 | n11(B) |
| | n10(A) | n11(A) | n1 |
| | | | |
| | Test A Result | | Total |
| | negative | positive | |
| negative | a0 | b0 | n00(B) |
| positive | c0 | d0 | n01(B) |
| | n00(A) | n01(A) | n0 |

**From:**
**Cheng and Macaluso:**
**Epidemiology 8, pp 104-106, 1997**

$$RS^{+}(A,B)=\frac{b_1+d_1}{c_1+d_1};Var\{\ln\ RS^{+}(A,B)\}=\frac{b_1+c_1}{(b_1+d_1)(c_1+d_1)}$$

$$RS^{-}(A,B)=\frac{b_0+a_0}{c_0+a_0};Var\{\ln\ RS^{-}(A,B)\}=\frac{b_0+c_0}{(b_0+a_0)(c_0+a_0)}$$

# EXAMPLE

**Data from Pepe's book: a paired study of exercise stress test (EST: Test B) versus chest pain history (CPH: Test A) for diagnosing coronary artery (Disease D)**

| CASES | Test A Result | | Total |
|---|---|---|---|
| | negative | positive | |
| negative | 25 | 183 | 208 |
| positive | 29 | 786 | 815 |
| | 54 | 969 | 1023 |
| | | | |
| CONTROLS | Test A Result | | Total |
| | negative | positive | |
| negative | 151 | 176 | 327 |
| positive | 46 | 69 | 115 |
| | 197 | 245 | 442 |

# TESTS OF SIGNIFICANCE

| CASES | Test A Result | | Total |
|---|---|---|---|
| | negative | positive | |
| negative | 25 | 183 | 208 |
| positive | 29 | 786 | 815 |
| | 54 | 969 | 1023 |
| | | | |
| CONTROLS | Test A Result | | Total |
| | negative | positive | |
| negative | 151 | 176 | 327 |
| positive | 46 | 69 | 115 |
| | 197 | 245 | 442 |

$$\text{Cases}: \chi^2 = \frac{(183-29)^2}{183+29} = 111.87$$

$$\text{Controls}: \chi^2 = \frac{(176-46)^2}{176+46} = 76.13$$

# CONFIDENCE INTERVALS

| CASES | Test A Result | | Total |
|---|---|---|---|
| | negative | positive | |
| negative | 25 | 183 | 208 |
| positive | 29 | 786 | 815 |
| | 54 | 969 | 1023 |
| | | | |
| CONTROLS | Test A Result | | Total |
| | negative | positive | |
| negative | 151 | 176 | 327 |
| positive | 46 | 69 | 115 |
| | 197 | 245 | 442 |

$$RS^{+}(A,B)=\frac{183+786}{29+783}=1.189;Var\{\ln RS^{+}(A,B)\}=\frac{183+29}{(183+786)(29+786)}=(.016)^{2}$$

$$RS^{-}(A,B)=\frac{176+69}{46+69}=2.130;Var\{\ln RS^{-}(A,B)\}=\frac{176+46}{(176+69)(46+69)}=(.089)^{2}$$

$$RS^{+}=\exp\{\ln(1.189)\pm(2.28)(.016)\}=(1.146,1.232)$$

$$RS^{-}=\exp\{\ln(2.130)\pm(2.28)(.089)\}=(1.738,2.609)$$

Again, we should use +/- 1.96 if each test is at 5%

## Suggested Readings:

Search, find, and read the article by Cheng and Macaluso: Epidemiology 8, pp 104-106, 1997

# Suggested Exercises:

#1. Re-examine the case of the example where we assume that there is a good screening procedure (98% sensitive and 97% specific) and let consider 2 other examples, a low-risk sub-population (prevalence is .5%) and a higher-risk sub-population (prevalence is 10%). In each case, calculate the positive predictive value and the negative predictive value.

#2. Prove that, in the 2-by-2 cross-classification of D(+,-) versus T(+,-), Odds Ratio is equal to 1 if & only if J=0

**#3.** In a previous study (Anderson et al, 2001) of environmental tobacco smoke, we compared two groups of non-smoking women, $n_1 = 23$ women had male partners who smoke in the home and $n_0 = 22$ women who had male partners who did not smoke. Urine samples were obtained and analyzed and the comparison based on a number of chemicals, among then cotinine (a metabolite of nicotine, in nmol/mL) and NNAL and its glucuronide, NNAL-Gluc (NNAL and NNAL-Gluc are metabolites of the tobacco-specific lung carcinogen called NNK, in pmol/mL). Data (cotinine, NNAL+NNAL-Gluc) are given in the following page (ND is for "not detectable", the limit of detection for cotinine is .003 nmol/mL and for NNAL and NNAL-Gluc is .005 pmol/mL; one case has missing value for NNAL+NNAL-Gluc):

Non-exposed women: (ND,ND), (ND,ND), (ND,ND), (ND,ND), (ND,ND), (ND,.008), (.003,ND), (.003,.015), (.006,ND), (.007,ND), (.007,ND), (.007,.018), (.008,ND), (.008,ND), (.009,ND), (.01,ND), (.012,ND), (.016,ND), (.017,ND), (.019,.047), (.025,ND), and (.03,ND).

Exposed women: (ND,.067), (.003,.009), (.003,.012), (.007,.039), (.008,ND), (.008,.010), (.008,.011), (.009,ND), (.011,.037), (.017,.072), (.018,-), (.021,.083), (.036,.022), (.037,.032), (.042,.063), (.046,ND), (.053,.210), (.076,.041), (.099,.018), (.101,.031), (.111, .018), (.122,.282), and (.200,.027).

3A. Determine the optimal cutpoint for cotinine and the corresponding sensitivity and specificity; or

3B. Determine the optimal cutpoint for NNAL+NNAL- Gluc and the corresponding sensitivity and specificity.