STUDY DESIGNS

IN BIOMEDICAL RESEARCH

INTRODUCTORY DIAGNOSTIC MEDICINE

(Major Examples & Simple Designs)

#1. Should women start mammograms at age 40 or at age 50? Many would say the earlier the better because early detection saves life. But federal panel recommended against that, why? Which side are you on?

#2. Should we conduct random testing for HIV/AIDS? Are you concerned about errors? (with very large sample, standard error is minimal). Are you concerned about confidentiality and unwanted consequences? Do you want to know the truth?

#3. Suppose 1000 people were tested for HIV infection and 10 were found positive; that puts the estimate at 1%; are you confident in this estimate? Could it substantially under estimate or overestimate the true rate? Would it help just increase the sample size?

We use the term “diagnosis” to aim at the act or process of predicting a not yet observable condition - such as a disease - using an observable characteristics (clinical observations and/or laboratory test results); referred to as separators or predictors.

The process leads to a quick, easy, and economical way to classify individuals as “diseased” (condition present) or “healthy” (condition absent). Think of regression!

Some Simple Examples: Skin test for TB (tuberculosis), Urine tests for Early Pregnancy, etc

- Screening is a population-based process (Public Health) whereas Diagnosis is individually-based (Medicine).
- However, the difference is not in the make-up of the processes but in their uses.
- For the purpose of learning Biostatistics, we focus on screening but make no strong distinction between the terms “diagnosis” and “screening”; differences, if any, are minor – two terms are used exchangeably.

OTITIS MEDIA

- The 2nd most prevalent disease on earth, affects 90% of children by age 2; Costs 3.8 billions in direct costs (physician visits, tube placements, antibiotics, etc…) in 1995 dollars; Causes hearing loss, learning disabilities, and other middle-ear sequelae.
- As a children's disease, It is the most common diagnosis at physician visits ahead of well-child, URI, injury, and sore throat; It is responsible for 24.5 million physician visits in 1990.

OTITIS MEDIA DIAGNOSIS

- Diagnosis is usually made on clinical grounds - i.e. using otoscopic characteristics of the tympanic membrane – or ear drum (color, position, appearance, and mobility); or persistent earache or bubbles of air (ear exam).
- Otoscopy - from ear's exam - is often supplemented by tympanometry, a method that measures the compliance of the tympanic membrane; two (continuous) characteristics have emerged as leading potential predictors of middle ear fluid: the "static admittance" and the "tympanometric width" (background in physics).

DIABETES

- Diabetes is a disease in which the body is unable to properly use and store glucose (a form of sugar); glucose backs up in the bloodstream causing one's "blood glucose" to rise too high.
- Two types of diabetes: Type 1 (juvenile-onset or insulin-dependent - about 10% of all cases - in whom the body stops producing insulin) and Type 2 (adult-onset or non insulin-dependent - about 90% of all cases - in whom the body does not produce enough or unable to use insulin properly - called "insulin resistance").

DIABETES DIAGNOSIS

- There may be symptoms (eg. being very thirsty, blurry vision, etc…) or risk factors (eg. family history, being overweight, etc …): Not enough!
- Primary diagnosis, however, is based on "plasma glucose" often measured/tested early in the morning before eating meals (called "fasting plasma glucose"); the other major factor is "A1c", a measurement for three-month accumulation average.

THYROID GLAND

- Thyroid is a small bowtie or butterfly-shaped gland, located in your neck, wrapped around the windpipe, behind and below the Adam's Apple area.
- The hypothalamus - part of the brain - releases Thyroid-releasing hormone (TRH) leading to the release of Thyroid-stimulating hormone (TSH); THS tells the thyroid to make thyroid hormones and release into bloodstream as a feedback process.

THYROID HORMONES

- The thyroid produces several hormones, of which two are very important - one is Triiodothyronine (called T3) and the other is thyroxine (called T4)
- These hormones help oxygen get into cells; the hormones then help cells to convert oxygen and calories into energy making thyroid the "master gland of metabolism"

HYPERTHYROIDISM

- There are hypothyroidism & hyperthyroidism, the latter is more consequential.
- When your thyroid starts producing too much thyroid hormones, your body goes into overdrive, causing an increased heart rate, increased blood pressure, and burning more calories more quickly.
- These lead to weight loss, anxiety, muscle weakness, tremors, loss of concentration, etc...

THYROID DISEASE DIAGNOSIS

- Like the case of diabetes, there are risk factors (family history, menopause - the disease is more prevalent among women, being exposed to radiation); and many more symptoms: Not enough
- Primary diagnosis, however, is based on blood test to measure the levels of three major hormones: T3, T4, and TSH - all are on continuous scale.

PROSTATE

- The prostate is part of a man's reproductive system. It is a gland surrounding the neck of the bladder.
- A healthy prostate is about the size of a walnut and is shaped like a donut. The urethra (the tube through which urine flows) passes through the hole in the middle of that "donut". Because of that, if the prostate grows too large, it squeezes the urethra causing a variety of urinary problems.

PROSTATE CANCER

- Cancer begins in cells, building blocks of tissues
- When normal process goes wrong, new cells form unnecessarily and old cells do not die when they should. Extra mass of cells called a tumor; and malignant tumors are cancer.
- No one knows the exact causes of prostate cancer … yet, but age is a significant factor. Most men with prostate cancer are over 65; if they live long enough a large proportion of men would eventually have prostate cancer.

PROSTATE CANCER SCREENING

- There are risk factors (age, family history) and symptoms (inability to urinate, frequent urination at night, etc…): Not enough
- Common screening is a blood test to measure prostate-specific antigen (PSA).
- However, a high level could be caused by benign prostatic hyperplasia (BPH – growth of benign cells); so the test is not specific.
- Another/newer candidate is “Cathepsin B”.

BREAST CANCER

- Breast Cancer is an uncontrolled proliferation of cells (when normal process goes wrong, new cells form unnecessarily and old cells do not die when they should); extra cells form tumors, some are malignant.
- It's a very diverse disease of many varying histological subtypes; different subtypes make it more difficult to treat and to screen.
- The lifetime risk for American women is 1 in 8 – up from 1 in 20 in 1960; In 2009, there were over 200,000 new cases – majority are invasive.

SCREENING FOR BREAST CANCER

- Genetic predisposition, genes BRCA1 and BRCA2, accounts for only 5% to10% of all breast cancer cases.
- No obvious risk factors other than family history and age (& gender); by the age of 50 years, more than 50% of the BRCA1 or BRCA2 mutation carriers have already developed the disease.
- Existing screening methods are: Self Breast Exam, Ultrasound, Magnetic Resonance Imaging (MRI), and Mammography.

BREAST SELF-EXAMINATION

A large randomized trial (n = 266,064) in Shanghai (1989-1991) lead to the following conclusion: “Women who choose to practice breast self-examination should be informed that its efficacy is unproven and that it may increase their chances of having a benign breast biopsy”; after 10 years of follow-up, breast cancer mortality rates in 2 groups were identical (JNCI 94: 1445-1457, 2002).

In short, it might do more harm than good.

ULTRASONOGRAPHY (US)

- The term "ultrasound" applies to all acoustic energy with a frequency above human hearing.
- Medical Ultrasonography is an ultrasound-based diagnostic imaging technique used to visualize muscles and internal organs, their size, structures and possible pathologies.
- More popular in OBGYN for prenatal care but not so popular for Breast Cancer Screening in general; often only used for BC during pregnancy to avoid radiation (of mammography).
- It is about as sensitive but a little less specific than mammography; specificity ranges 80-93%; it picks up a few more benign tumors.

MAGNETIC RESONANCE IMAGING (MRI)

- Magnetic resonance imaging (MRI) is a non-invasive method used to render images of the inside of an object.
- It uses radio waves and a strong magnetic field to provide remarkably clear and detailed pictures of internal organs and tissues.
- It requires specialized equipment to evaluate body structures that may not be as visible with other imaging methods; e.g. you can see not only the organs but even blood vessels too.

ADVANTAGES OF MR IMAGING

- Use of MRI first reported in 1985.
- MRI not associated with ionizing radiation; no known long-term side effects.
- MRI is not impaired by dense parenchyma; sensitivity improves;
- MRI could measure not only physiological but functional properties of tissues as well.
- However, for now, breast MR imaging is not used routinely in a screening setting. Why?

MAMMOGRAPHY

- Mammography is the process of using low-dose X-rays to examine the human breast; It uses doses of ionizing radiation to create image
- It is used to detect and diagnose breast disease or tumor, both in women with or without breast complaints or symptoms - i.e. more routine.
- Modern mammography has only existed since 1969, when the first x-ray machines used just for breast imaging became available. Technology has advanced, so that today's mammogram is very different even from those of the mid-1980s.

THE ISSUE

- 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 the major issue; 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%.

THE SCREENING PROCESS

- Any screening idea, like mammography, must go through a two-stage process
- Stage I: Developmental Stage
  The question here is: Does the idea work?
- Stage II: Applicational Stage
  The question here is: Does it work for me? (i.e. the user)

THE DEVELOPMENTAL STAGE

- In the Developmental Stage, the basic question is: Does the idea work? It's the investigator's burden to prove to public or regulatory agencies.
- Approach: Trying the test's idea on a "pilot population" where one compares the test results versus truth; we have data.

KEY PARAMETERS

- Two parameters:
  Sensitivity, $S^+ = Pr(T=+|D+)$
  Specificity, $S^- = Pr(T=-|D=-)$
- Sensitivity is the probability to correctly identify a diseased individual and Specificity the probability of correctly identify a healthy individual.

HYPOTHESIS TESTING: AN ANALOGY

- **Sensitivity is (1 - false negative rate);**
  a false negative is a "type II error", $S^+ = 1 - \beta$.
- **Specificity is (1 - false positive rate);**
  a false positive is a "type I error" , $S^- = 1 - \alpha$.
- **Clearly, it is desirable that a test or screening should be highly sensitive and highly specific; both α and β are small.**

Sensitivity is (1 - false negative rate);
a false negative is a “type II error”,
$S^+ = 1 - \beta$

Sensitivity is the analog of “Statistical Power”
(& False positive rate is type I error, alpha)

Most recently:

- **Mammography** (Breast Cancer) is about **96.6% specific and 64.7% sensitive**.
- **Pap Smear** (Cervical Cancer) is about **98.5% specific and 40.6% sensitive** (higher here).
- **Fasting Blood Sugar** (Diabetes) is about **96.9% specific and 64ï% sensitive**.

Most tests are highly specific (and somewhat less sensitive); so why do we have so much problem with false positives?

Very briefly: Most diseases – including all cancers - are "rare" (low prevalence and incidence). There are many more healthy people, there are more false positives – even the false positive rate is low (e.g. 2% of 1,000,00 healthy people vs. 30% of 200 people with disease)

Here Diagnostics and Statistics agree on one thing: focus on having very small Type I error (than Type II error); remember the magic cut-points for p-values: 0.01 or 0.05

And let see some statistical details:

Consider the "response/test rate":

π_t = Pr(T=+).

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

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

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

We have a relationship between Response Rate, Disease Prevalence, Sensitivity, and Specificity.

Why Specificity more dominating?

Reason: larger coefficient, $(1-\pi)$ versus π.

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

EXAMPLE #1

Scenario	Sensitivity	Specificity	Prevalence	Response Rate
1A	0.9	0.9	0.002	0.1016
1B	0.9	0.9	0.005	
1C	0.9	0.9	0.200	
2A	0.9	0.8	0.002	0.2014
2B	0.9	0.8	0.005	
3A	0.8	0.9	0.002	0.1014

Compare: 1A versus 2A, then 1A versus 3A; then Response Rate versus Prevalence (by investigating 1A vs. 1B vs. 1C- yourself)

The design in stage I involves two samples, those with the disease and those without.

Sensitivity and Specificity are estimated by the corresponding sample proportions. We can easily obtain standard errors and form confidence intervals.

THE APPLICATIONAL STAGE

- In the Applicational Stage, the basic question is: Does it work for "me"? It's the user's concern.
- Problem: One can't resolve the concern, like comparing the test result versus the truth, i.e. no data (one person; truth is known).

KEY PARAMETERS

- Two parameters:
  Positive Predictive Value, $P^+ = Pr(D=+|T=+)$ &
  Negative Predictive Value, $P^- = Pr(D=-|T=-)$
- Positive predictive value is the probability having an accurate positive result and negative predictive value is the probability having an accurate negative result; (Perhaps, users are more often concerned about P^+ than P^-).

ESTIMATION OF PARAMETERS

In stage II, there are no data other than one "incomplete observation", that of the user of which we know the test result but not the disease status. It is not possible to "estimate" the Predictive Values; we could only obtain the values of these parameters using the Bayes' theorem indirectly then "approximate" them.

BAYES' RULE

$$\Pr(B \mid A) = \Pr(B \text{ and } A)/\Pr(A)$$

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A \text{ and } B) + \Pr(A \text{ and Not } B)}$$

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A \mid B)\Pr(B) + \Pr(A \mid \text{Not } B)\Pr(\text{Not } B)}$$

We apply the Bayes' Rule twice:

(1) Let B=(D=+) and A=(T=+) to obtain P+

(2) Let B=(D=-) and A=(T=-) to obtain P-

POSITIVE PREDICTIVE VALUE

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A \mid B)\Pr(B) + \Pr(A \mid \text{Not } B)\Pr(\text{Not } B)}$$

Let A = (T=+) and B = (D=+), we have:

$$\Pr(D = + \mid T = +) = \frac{\Pr(T = + \mid D = +)\Pr(D = +)}{\Pr(T = + \mid D = +)\Pr(D = +) + \Pr(T = + \mid D = -)\Pr(D = -)}$$

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

NEGATIVE PREDICTIVE VALUE

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A \mid B)\Pr(B) + \Pr(A \mid \text{Not } B)\Pr(\text{Not } B)}$$

Let A = (T=-) and B = (D=-), we have:

$$\Pr(D = - \mid T = -) = \frac{\Pr(T = - \mid D = -)\Pr(D = -)}{\Pr(T = - \mid D = -)\Pr(D = -) + \Pr(T = - \mid D = +)\Pr(D = +)}$$

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

PREDICTIVE VALUES

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 π=**.003**:

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

Example B: S^+=.977, S^-=.926, and π=**.20**:

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

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

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

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

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

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

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

Result: The higher the Prevalence, the lower the Negative Predictive Value. However, the effect is much weaker here; the derivative is negative but near zero.

Versus :

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

EXAMPLE #2:

1A vs. 1B, 1A vs. 2A, & 1A vs. 3A

Since all cancers are "rare" (low prevalence and incidence), specificity has its dominating effect on the positive predictive value but not much on the negative predictive value.

Scenario	Sensitivity	Specificity	Prevalence	P+	P-
1A	0.9	0.9	0.002	0.0178	0.9998
1B	0.9	0.9	0.005	0.0433	0.9994
2A	0.9	0.8	0.002	0.0089	0.9997
2B	0.9	0.8	0.005		
3A	0.8	0.9	0.002	0.0158	0.9996
3B	0.8	0.9	0.005		

Suggested Exercises:

#1. Calculate the response rates left open in Example 1 (scenarios 1B and 1C)

#2. Calculate the positive predictive values and negative predictive values left open in Example 2 (scenarios 2B and 3B)