BIOSTATISTICS METHODS

FOR TRANSLATIONAL & CLINICAL RESEARCH

DIAGNOSTIC MEDICINE

Part A: Concepts & Designs

DEFINITIONS

- **DIAGNOSIS: The act or process of identifying or determining the nature of a disease through examination.**
- **SCREENING: The act or process of separating, or sifting out by means of an appraisal or a selection.**

From a statistical point of view, screening tests are diagnostic tests; they are procedures geared toward detecting a "condition" (eg. disease). The context in which they are applied or use in practice sets them apart.

"Screening tests" are applied in large scale; therefore, they must be non-invasive and inexpensive. A positive result is usually not followed by treatment, but with further, more definitive procedures; thus, their accuracy can be somewhat less than perfect.

A GENERIC TERMINOLOGY

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

CRITERIA FOR USEFUL SCREENING

(1) Disease should be serious or potentially so; if the disease has no serious consequences in terms of longevity or quality of life, there is no benefit to be gained from diagnosing it.

(2) Disease should be relatively prevalent in the target population; otherwise, benefit (of/used as screening) is rather minimal.

(3) Disease should be treatable.

(4) Treatment should be available to those tested positive; sometimes treatment exists but not accessible because of the cost or practical or social reasons.

CRITERIA FOR USEFUL SCREENING

(5) The test should not harm the individuals. Of course, all tests have negative impact above money and time; they may cause physical or emotional discomfort and sometimes are even dangerous to the well being of the subject. The overriding principle is that these should be reasonable and not outweighed the potential benefits.

(6) The test should accurately classify diseased and non-diseased individuals; false negatives leave diseased subjects untreated and false positives result in unnecessary treatments- both should be minimal.

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.
- 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" .

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…).
- 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 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.**
- **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 and it contributes a secretion to the semen.**
- **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".**
- **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…)
- 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” – among others.

Sometimes screening is more complicated than just making "one measurement". For example, having one high measurement of PSA might not mean anything because – as mentioned - a high level could be caused by benign prostatic hyperplasia . There should be "some pattern" of measurements.

A PROSTATE CANCER MODEL

- **Serum PSA in patients diagnosed with prostate cancer follows an exponential growth curve.**
- **A retrospective study of banked serum samples (Carter et al., *Cancer Research 52*, 1992) showed that the exponential growth begins 7-9 years before the tumor is detected clinically.**

EXPONENTIAL GROWTH MODEL

$$PSA_t = PSA_0 \exp(\beta_1 t)$$

$$Y_t = \ln PSA_t$$

$$= \beta_0 + \beta_1 t$$

This is a "A Simple Linear Regression Model" with PSA used on log scale.

AIDS

- Acquired Immunodeficiency syndrome (AIDS) is a severe manifestation of infection with the Human Immunodeficiency Virus (HIV, identified in 1983).
- The virus destroys the immune system leading to opportunistic infections of the lungs, brain, eyes, and other organs; Consequences include debilitating weight loss, diarrhea, and several forms of cancer.
- Currently, 40 millions living with AIDS; about 5 millions newly infected and 3 millions deaths in 2004 – most affected region is Sub-Sahara Africa. Diagnosed by blood tests (for example, using $CD4^+$ T-cell count, or $CD8^+$ T-cell count).

ULCERS

- An ulcer is a break in the lining of the stomach or in the duodenum (first part of small intestine); Gastric/peptic ulcers cost 3.2 billions in1975 dollars.
- Most ulcers are caused by H. Pylori (identified in 1982), a bacterium living in the pylorus (the passage connecting the stomach and the duodenum).
- The two Australian physicians who discovered this bacteria won Nobel prize in 2005.
- Diagnosed by a blood test, even a breath test.

INFECTIONS

- **There are many bacterial and viral diseases such as AIDS and ulcers.**
- **Bacterial and viral diseases are often easier, and more accurately, to be detected because: (i) the disease is "better defined", and (ii) they are all implicated by a common marker: some form of agent-specific antibodies (provided that we know/have the right assay!)**

THE DIAGNOSIS PROCESS

- It starts with an idea, it could be accidental or the result of a long search. The idea then goes 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 testee/user; or when does it work? Or to whom does it work?)

THE DEVELOPMENTAL STAGE

- **In the Developmental Stage, the basic question is: Does the idea work? It's the investigator's (or producer's) burden to prove.**
- **Approach: Trying the test's idea on a "pilot population" where one compares the test results versus truth; the "true diagnosis" may be based on a more refined/accurate method or evidence emerged after the passage of time (Note: here we have data).**

TEST RESULT

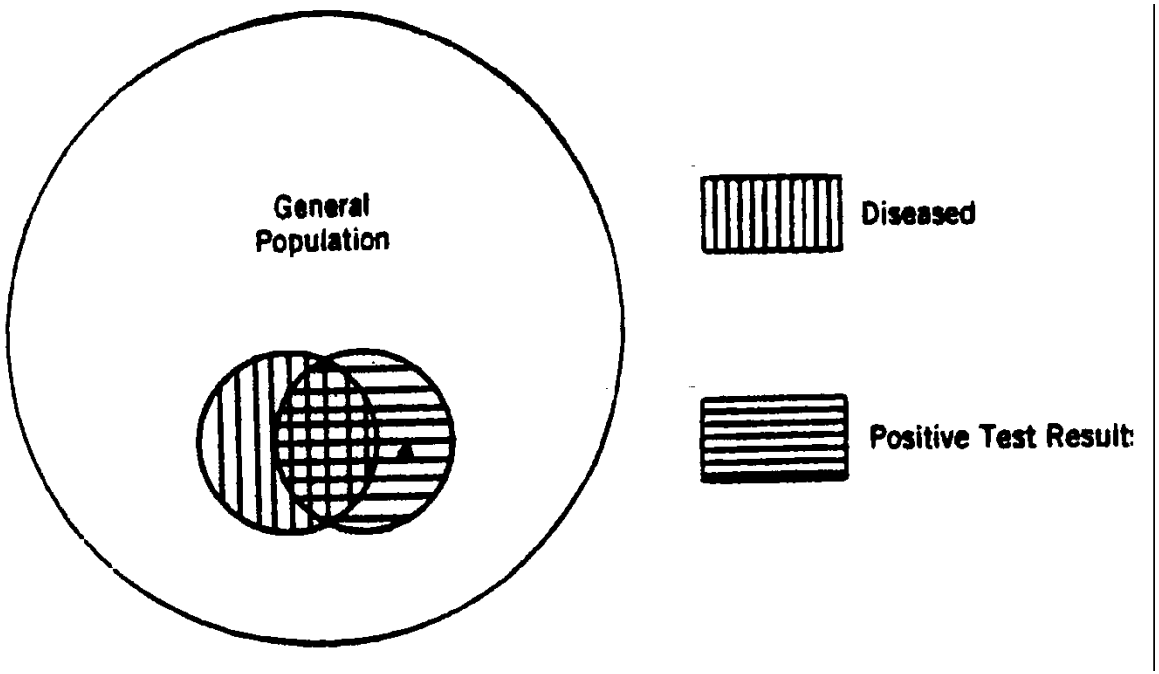
Some tests yield dichotomous results, such as the presence or absence of a bacterium or some specific DNA sequence; but many may involve assessments measured on ordinal scale (eg. 5-point scale for mammograms) or continuous scale (eg. blood glucose). Let start with the simple binary case.

KEY PARAMETERS

- Let “D” and “T” denote the true diagnosis and the test result, respectively
- The key parameters are two conditional probabilities):
  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

The idea, in the developmental stage, was to classified people as “diseased” (condition present) or “healthy” (condition absent) based on certain measurement (from blood or urinary components). The basic question is “How high is high?” or “How low is low?”.

MISCLASSIFICATION



	Test=Positive	Test=Negative
Diseased	True Positive	False Negative
Healthy	False Positive	True Negative

PARAMETERS AND ERRORS

- Sensitivity is (1 - false negativity); false negativity is the “rate” (%) of false negatives.
- Specificity is (1 - false positivity); false positivity is the “rate” (%) of false positives.
- Clearly, it is desirable that a test or screening should be highly sensitive and highly specific; both error rates are small.
- Sensitivity and (1- Specificity) are also referred to as “true positive rate” and “false positive rate”.

PARAMETER ESTIMATION

- Sensitivity and specificity can simply be estimated as "proportions" s^+ and s^- from the two samples;
- Sensitivity is the proportion of diseased individuals detected as positive by the test; specificity is the proportion of healthy individuals detected as negative.
- Standard errors and 95% confidence intervals, for example, are calculated accordingly.

$$\text{sensitivity} = \frac{\text{number of diseased individuals who screen positive}}{\text{total number of diseased individuals}}$$

$$\text{specificity} = \frac{\text{number of healthy individuals who screen negative}}{\text{total number of healthy individuals}}$$

Example

CERVICAL CANCER

This "Pap" Test is highly specific (Specificity=98.5%) but not very sensitive (Sensitivity = 40.6%). If a healthy person is tested, the result is almost sure negative; but if a woman with cancer is tested the chance is 59.4% that the disease is undetected.

	Test		
True	-	+	Totals
-	23,362	362	23,724
+	225	154	379

$$\text{Sensitivity} = \frac{154}{379} = 0.406 \text{ or } 40.6\%$$

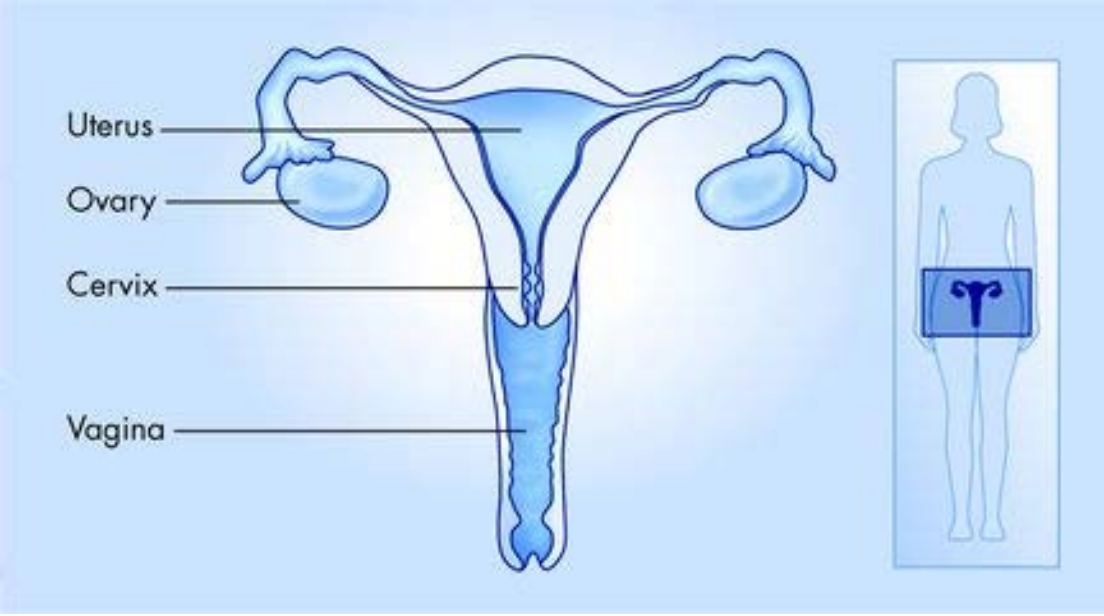
$$\text{Specificity} = \frac{23362}{23724} = 0.985 \text{ or } 98.5\%$$

Cervical Cancer Overview

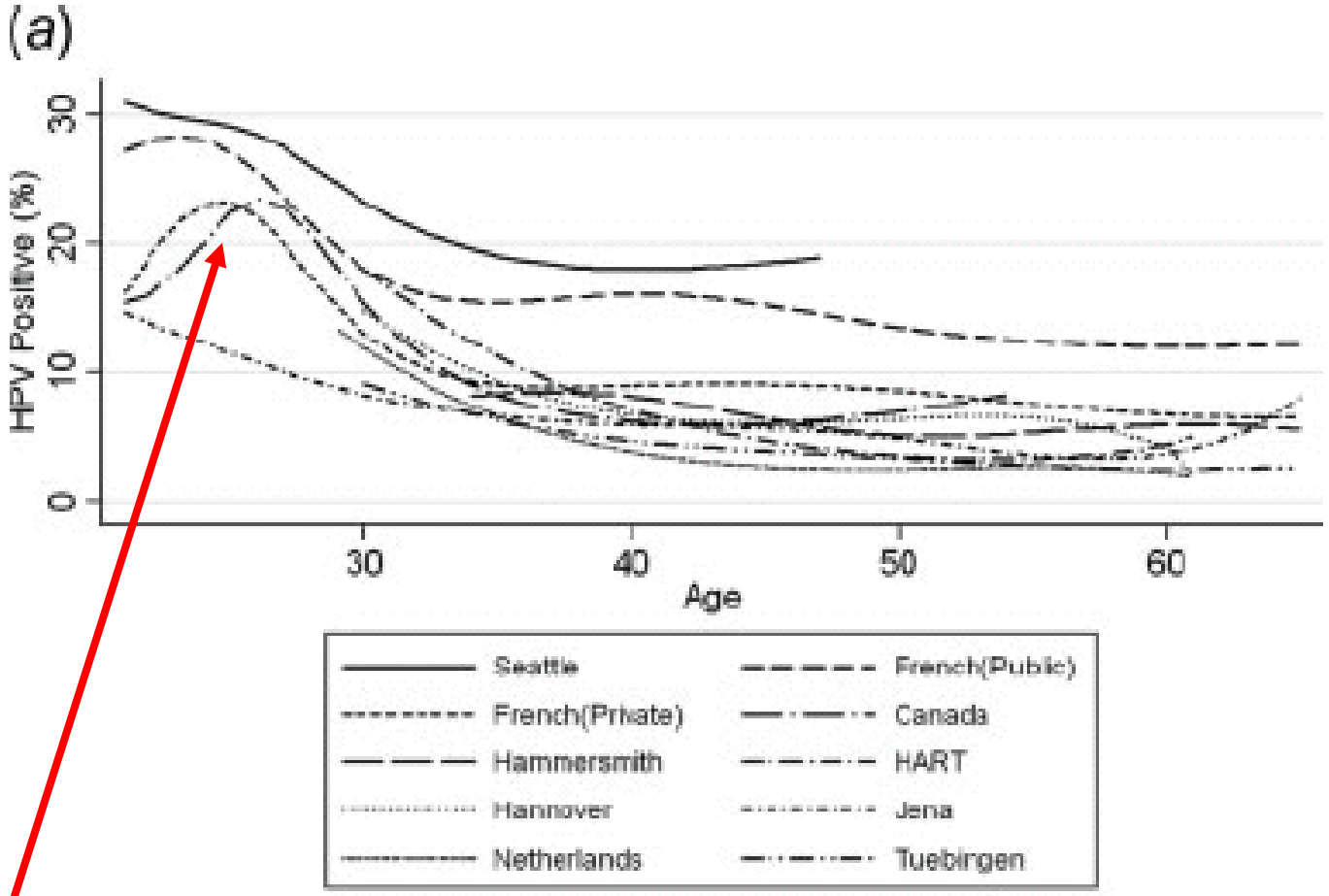
- A cancer that forms in the tissues of the organ connecting the uterus and vagina
- A slow-growing often asymptomatic cancer
- 11,070 new cases in the U.S. in 2008 (NCI)
- 3,870 deaths in the U.S. in 2008 (NCI)
- Worldwide, it's No. 2 "most common cancer" for women.

CERVICAL CANCER: QUESTIONS

- **What else do we currently know about screening for cervical cancer?**
- **Why should screening change in the era of successful vaccination programs?**



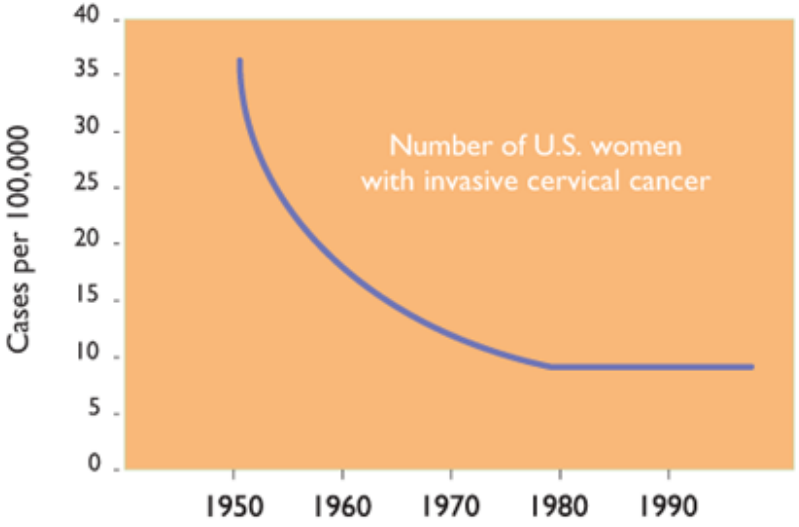
HPV prevalence by age



Peak age (in the 20's)

Benefits of Screening

- **Earlier detection can lead to a decrease in mortality from cancer not diagnosed until later stages of disease**
- **Can lead to longer life and improved quality of life for cancer patients**



***74%* reduction in incidence of cervical cancer in USA during 1955-1992**

THE PAP TESTS

- **The “Pap” test or Pap Smear test is an important part of women’s health care. The smeared cells or cell suspension is placed on a glass slide, stained with a special dye (Pap stain), and viewed under a microscope.**
- ***Pap test* look for abnormal cells in the lining of the cervix before they have the chance to become pre-cancer or cervical cancer. It may be not sensitive due to cases in early stage.**
- **FDA approved in 1990s a liquid-based technique, called “Thin-layer Pap” with improved sensitivity and specificity – plus 2 DNA tests.**

Virus & Vaccine

- **Certain types of the human papillomavirus (HPV) infection is a primary cause; for most women who have HPV, the virus will go away on its own. If not go away, abnormal cell can develop in the lining of the cervix; if not found early & treated, precancer and then cervical cancer can develop.**
- **An HPV vaccine, called Gardasil; approved by the FDA in 2006. However, even with the vaccine, screening is still highly recommended by experts.**

THE APPLICATIONAL STAGE

- **In the Applicational Stage - when the product is on the market, the basic question is: Does it work for "me"? (the user/testee; or when does it work?); It's the user's (or consumer's) concern.**
- **<u>Problem</u>: One can't resolve the concern, like comparing the test result versus the truth, because if one knows the truth one would not need the test; and waiting for evidence to emerge after some passage of time may be "too late".**
- **Need to know if you could "trust" before trying!**

KEY PARAMETERS

- Again, let “D” and “T” denote the true diagnosis and the test result, “D” is unknown in this stage.
- The key parameters are two “other” conditional probabilities:
  Positive Predictive Value, $P^+ = Pr(D=+|T=+)$
  Negative Predictive Value, $P^- = Pr(D=-|T=-)$
- Positive predictive value, or positive predictivity, is the probability have an accurate positive result and negative predictive value is the probability have an accurate negative result; Perhaps, users are more often concerned about P^+ than P^-.

Simple Illustration: Should We Conduct "RANDOM TESTING" For Diseases, Such As AIDS?

- 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.

EXAMPLES

Assumptions: Let assume

- **Complete privacy**
- **Complete confidentiality**
- **There is a good/reliable screening procedure (say, 98% sensitive and 97% specific)**
- **Consider 2 examples, a low-risk sub-population (Example A, prevalence is .1%) and a high-risk sub-population (Example B, prevalence is 20%) – then see what we can learn.**

Example A:

	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

RESULTS

- **True prevalence: 100/100,000 = .1%**
  Estimated prevalence: 3,095/100,000 = 3.1%
  Not good for policy makers: Over estimate more than 30 times;
- **P^+ = 98/3,095 = 3.2%; Not good for users: very low Positive Predictive Value (P^+)**

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

Example B:

	Infection=Yes	Infection=No	Total
Test=Positive			
Test=Negative			
Total	20000	80000	100000

	Infection=Yes	Infection=No	Total
Test=Positive	19600		
Test=Negative	400		
Total	20000	80000	100000

	Infection=Yes	Infection=No	Total
Test=Positive		2400	
Test=Negative		77600	
Total	20000	80000	100000

	Infection=Yes	Infection=No	Total
Test=Positive	19600	2400	22000
Test=Negative	400	77600	78000
Total	20000	80000	100000

RESULTS

- **True prevalence: 20,000/100,000 = 20%**
  Estimated prevalence: 22,000/100,000 = 22%
  Good for policy makers: 22% versus 20%
- **P^+ = 19600/22,000 = 89.1%; Good for users that they can "trust" the results (before using): reasonable Predictive Value (P^+)**

	Infection=Yes	Infection=No	Total
Test=Positive	19600	2400	22000
Test=Negative	400	77600	78000
Total	20000	80000	100000

SIMPLE OBSERVATIONS

- **Predictive values of a screening test depend not only on sensitivity and specificity but on disease prevalence as well. The higher the prevalence, the higher predictive values.**
- **We should only "screen" high-risk sub-populations; "random screening" does not do anyone any good! Not to users/testees, not to policy makers (Note: Current Estimate for USA's AIDS: .3%).**

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.**
- **<u>Approach</u>: 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**
- **At the present time, mammography is about 96.6% specific and 64.7% sensitive.**

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^-).**

Issue #1:

ESTIMATION OF PARAMETERS

- **Unlike sensitivity & specificity, predictive values P^+ and P^- cannot be estimated directly because there are no data.**
- **However, they can be "estimated" indirectly using the Bayes' theorem or Bayes' rule. The formulas are simple.**

PREDICTIVE VALUES

Both **predictive values are functions of disease prevalence**, $\pi = \Pr(D = +)$:

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

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

Issue #2: Again, Should We Conduct "RANDOM TESTING" For Diseases, Such As AIDS?

- 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?

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% as above and S^+ and S^- are for ELISA in Weiss, 1985.

IMPLICATION

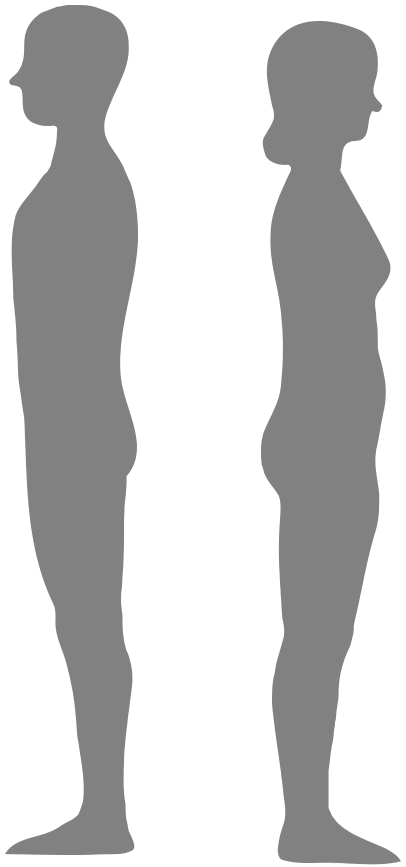
- Predictive values of a screening test depend not only on sensitivity and specificity but on disease prevalence too.
- The higher the prevalence, the higher the positive predictive value; “random screening” or “random testing” might not do much good – many false positives.
- The higher the prevalence,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

MORE ABOUT 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 <u>many</u> 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.**

2009 Estimated US Cancer Cases

Men 766,130		Women 713,220	
Prostate	25%	**Breast**	**27%**
Lung & bronchus	15%	Lung & bronchus	14%
Colon & rectum	10%	Colon & rectum	10%
Urinary bladder	7%	Uterine corpus	6%
Melanoma of skin	5%	Non-Hodgkin lymphoma	4%
Non-Hodgkin lymphoma	5%	Melanoma of skin	4%
Kidney & renal pelvis	5%	Thyroid	4%
Leukemia	3%	Kidney & renal pelvis	3%
Oral cavity	3%	Ovary	3%
Pancreas	3%	Pancreas	3%
All Other Sites	19%	All Other Sites	22%



#1 in Incidence

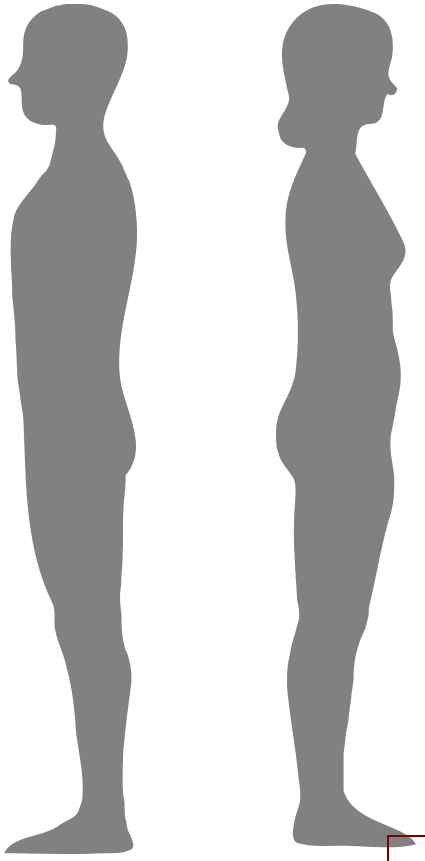
Lifetime Probability of Developing Cancer, U.S. Women, 2003-2005

Site	Risk
All sites†	1 in 3
Breast	**1 in 8**
Lung & bronchus	1 in 16
Colon & rectum	1 in 20
Uterine corpus	1 in 40
Non-Hodgkin lymphoma	1 in 53
Urinary bladder‡	1 in 84
Melanoma§	1 in 58
Ovary	1 in 72
Pancreas	1 in 75
Uterine cervix	1 in 145

#1 in Lifetime Risk

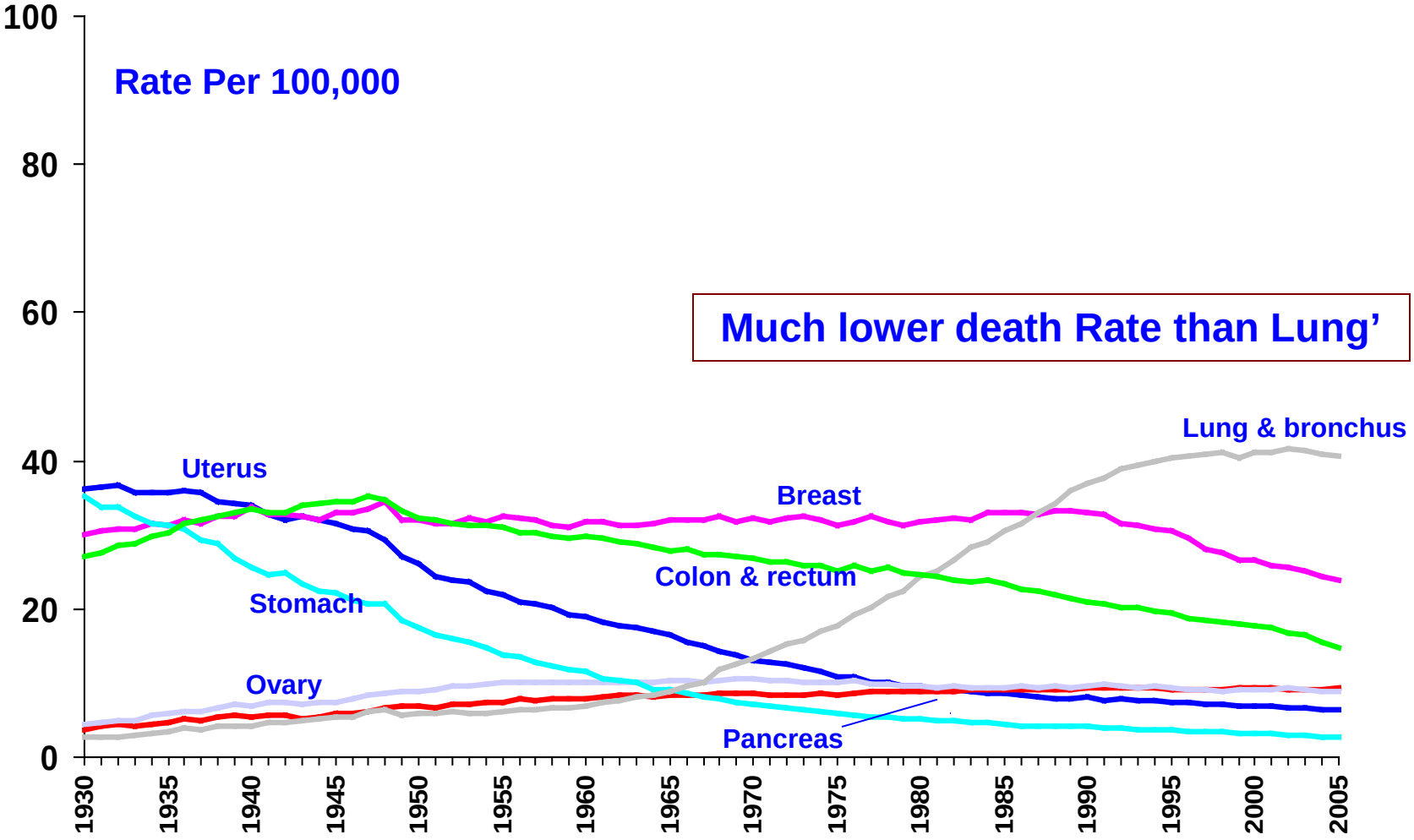
2009 Estimated US Cancer Deaths

	Men 292,540	Women 269,800	
Lung & bronchus	30%	26%	Lung & bronchus
Prostate	9%	**15%**	**Breast**
Colon & rectum	9%	9%	Colon & rectum
Pancreas	6%	6%	Pancreas
Leukemia	4%	5%	Ovary
Liver & intrahepatic bile duct	4%	4%	Non-Hodgkin lymphoma
Esophagus	4%	3%	Leukemia
Urinary bladder	3%	3%	Uterine corpus
Non-Hodgkin lymphoma	3%	2%	Liver & intrahepatic bile duct
Kidney & renal pelvis	3%	2%	Brain/ONS
All other sites	25%	25%	All other sites



#2 “Cancer Killer”

Cancer Death Rates* Among Women, US,1930-2005



*Age-adjusted to the 2000 US standard population.
Source: US Mortality Data 1930-2005, National Center for Health Statistics

Trends in Five-year Relative Survival (%)* Rates, US, 1975-2004

Site	1975-1977	1984-1986	1996-2004
• All sites	50	54	66
• **Breast (female)**	**75**	**79**	**89***
• Colon	52	59	65
• Leukemia	35	42	51
• Lung and bronchus	13	13	16
• Melanoma	82	87	92
• Non-Hodgkin lymphoma	48	53	65
• Ovary	37	40	46
• Pancreas	3	3	5
• Prostate	69	76	99
• Rectum	49	57	67
• Urinary bladder	74	78	81

Not as bad as you think

***It's 91% in 2009**

SCREENING FOR BC

- **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

- **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%.

*It's 3%-4% for women age 40

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?

Issue #3:

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"

THE PAP TEST: NOT THAT BAD!

- The “Pap” test or Pap Smear test, or cytology test, is an important part of women’s health care. The smeared cells or cell suspension is placed on a glass slide, stained with a special dye (Pap stain), and viewed under a microscope. It is used to detect cervical cancer as well as some vaginal or uterine infections.
- As for cervical cancer, **it is still not very sensitive**, especially cases in early stage. However, because it is highly specific (could be about 99%), its positive predictive value is high **making it suitable for “<u>case identification</u>”.**

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?

- Unfortunately, very often, neither maneuvers - by either improving its characteristics or by selecting the population with higher prevalence may be possible to yield $P^{+} \geq 50\%$; That's may be reasonable but too much to ask, even tests useful clinically may not pass!
- For AIDS, maybe one should only screen "high-risk" sub-populations, like drug IV abusers or prisoners; but what's breast cancer, what should we do? We know that early detection is proven to save lives.

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 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.

SCREENING EFFICACIES

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%

Even at age 35-39, if your mammogram is positive, there is still a 1% chance that you have breast cancer. Is 1% a worthy chance?

For some, when it comes to saving life, no chance is a slim chance; there are other costs but no cost is as pricey as life.

But, remember that false positives are not without consequences.

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

Issue #4: To form an INDEX measuring "**Diagnostic Competence**"

- Other things (cost, ease of application, etc…) being equal, **a test with larger values of both sensitivity and specificity is obviously better**.
- If not that clear cut, one has to consider the relative costs associated with 2 forms of error.
- If the 2 types of error are equally important, it may be desirable to have a single index to measure the "diagnostic competence" of the test.
- Could it be "**Overall Agreement**" Pr(D=T)?

OVERALL AGREEMENT

- The simplest measure would be the “overall agreement”, Pr (T=D).
- However, unlike sensitivity and specificity, the **overall agreement is influenced by the disease prevalence**.

YOUDEN'S INDEX

- One measure, **Youden's Index** (Cancer, 1950), has been popular
- If the 2 types of error are equally important, the Youden's Index J is defined as:

$$\mathbf{J = 1 - (\alpha + \beta) = S^{+} + S^{-} - 1}$$

The Youden's index J special with interesting characteristics: (i) it is based on a simple principle: **small sum of errors** (when neither one has priority), (ii) its value is larger when both sensitivity and specificity are high, and (iii) **It does not depend on the disease prevalence**.

YOUDEN'S INDEX MEASURES DIAGNOSTIC POWER

- In using a diagnostic marker/test, the "**gains**" are
  (i) $[P^+ - \pi]$, and
  (ii) $[P^- - (1-\pi)]$
- Youden's Index J is some kind of weighted **average of those gains**:

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

Issue #5:

When does a process qualify as a test?

- To decide if a “process” is a “test”, the minimum criterion it must pass is that it detects disease better than by chance alone
- That a process can only qualify as a test if **it selects diseased persons with higher probability than pure guessing: $P^{+}>\pi$.**

BASIC QUALIFICATION

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

$$P^+ > \pi \Leftrightarrow S^+ > 1-S^- \Leftrightarrow J > 0$$

$$\pi = \Pr(D = +)$$

$$P^+ = \Pr(D = + \mid T = +)$$

Issue #6:

THE ISSUE OF COST

For a Woman Age 40 – 44

$$100 + \frac{119}{100{,}000}[(.647)(100) + (.647)^2(5500)]$$

$$+ [1 - \frac{119}{100{,}000}][(.034)(100) + (.034)^2(5500)]$$

$$= \$112.81/\text{year}$$

Calculations are based on these cost estimates

Mammogram $80 – $120;

Needle Biopsy $5000 – $6000

SCREENING COSTS AT AGE 50

For a Woman Age 50 – 54

$$100 + \frac{254}{100{,}000}[(.647)(100) + (.647)^2(5500)]$$

$$+ [1 - \frac{254}{100{,}000}][(.034)(100) + (.034)^2(5500)]$$

$$= \$115.98/\text{year}$$

Calculations are based on these cost estimates

Mammogram $80 – $120

Needle Biopsy $5000 – $6000

ESTIMATED COSTS TO SAVE A LIFE

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)

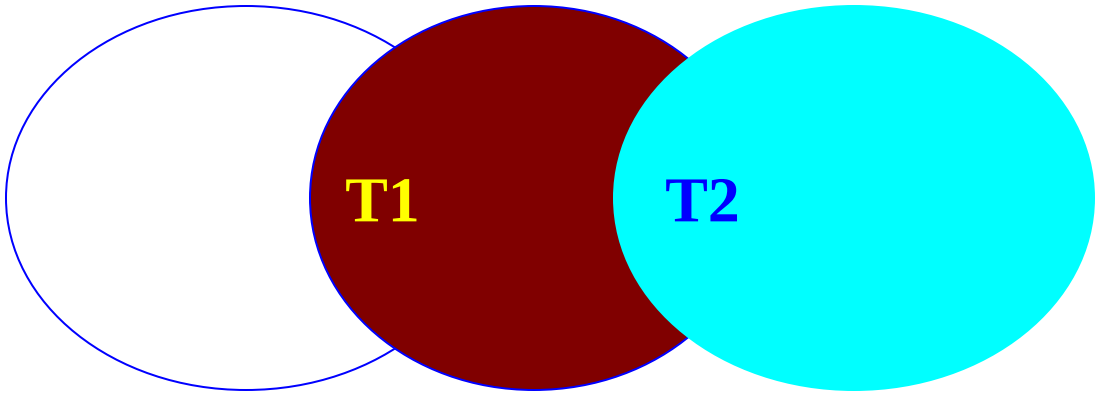
100,000 go through the process, 4 lives saved

25,000 go through the process to save 1 life (called NNS)
Age 40: NNS = 25,000 [Cost = (25,000)(112.81)=$2.82M/yr)
Age 50: NNS = 11,700 [Cost = (11,700)(115.98)=$1.36M/yr)

Prevalence Surveys

Clinical Research

Population Research



Laboratory Research

"Translational Research" consists of efforts to bring discoveries or works in the laboratories (Basic Sciences: biology, biochemistry etc…) to the "bedside" (Medicine) - referred to as "**T1**" or to "community interventions" (Public Health) – referred to as "**T2**".

DESIGNS

- **Estimation of disease prevalence often follows one of 2 different designs; let call them Design 1 and Design 2.**
- **In Design 1, we assume that sensitivity and specificity are known; in its more common version, Design 2, sensitivity and specificity are unknown.**

DESIGN #1

- This is the most simple scenario/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" by the test.

EASY SOLUTION?

Could we estimate of the disease prevalence by the frequency of positive tests: $p_t = x/n$?

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, $\pi = Pr(D=+)$.

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

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

For rare disease, $(1-\pi) >> \pi$, and $(1-S^{-})$ and $(1-S^{+})$ are often roughly equal; the bias is more likely **positive** – therefore, we likely **"over-estimate" the disease prevalence.**

POINT ESTIMATE, p

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

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

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

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

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

J is the Youden's Index

EXAMPLE #1B

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

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

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

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!

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

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

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

$$p_t = .116$$

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

A correction, using p instead of p_t , is a substantial improvement; **if S^+ and S^- are known apriori, then p is unbiased for π.**

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

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

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

$$= \boldsymbol{\pi}$$

STANDARD ERROR, SE(p)

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

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

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

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

Result: The "precision" of estimation of the prevalence **depends only on the size of Youden's index** rather than any function of sensitivity and specificity.

DESIGN #2

- This is still a simple scenario/design
- We have a screening test T; but its sensitivity S^+ and specificity S^- are not known.
- 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" by the test.

Sensitivity and specificity need to be estimated but there are not enough data from this “prevalence survey” to do so.

Sensitivity and specificity are estimated using two other independent 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.

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

We use the same estimator “p” of Design #1 with **S^+ and S^- being estimated from two (2) independent samples by proportions s^+ and s^-**

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.

However, the bias is negligible if the other two samples n_1 and n_0 are both large.

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

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.

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

PRIORITIES

- We already knew that positive predictive value is much more affected by the value of specificity.
- Now comparing the last two terms in Var(p); in common cases where both S^+ and S^- are high but π is low, **the last term is dominating.**
- That means the contribution of variability in the estimate of the specificity is usually the dominant term in the calculating the precision of the estimated disease prevalence.

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

ROC Curve

Diagnostic tests have been presented as always having dichotomous outcomes. In some cases, the result of the test may be binary, but in many cases it is based on the dichotomization of a continuous separator or **biomarker –** some factor correlated the absence or presence of the disease.

To deal with a continuous separator, we need a well-know **graph** called the Receiver Operating Characteristic curve or "**ROC curve**".

If the idea, in the developmental stage, was to classified people as “diseased” (condition present) or “healthy” (condition absent) based on certain continuous measurement (from blood or urinary components); then **we need to “dichotomize” the measurement**: for example, if the measurement is “high” then he’s classified as “diseased” – if it’s “low”, the subject is “healthy”. But the basic question is **“How high is high?”** or **“How low is low?”.**

A SIMPLE **PLAUSIBLE MODEL**

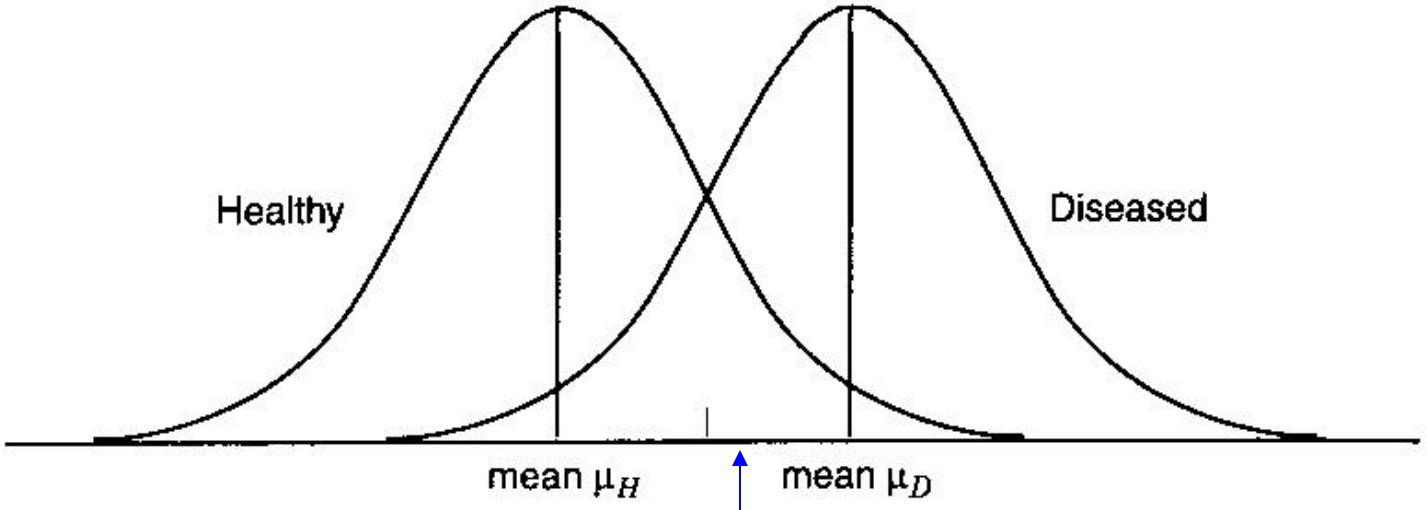


Figure Graphical display of a translational model of diseases.

T=- cutpoint T=+

Separator Y is normally distributed with the same variance, but different means; no matter where you "cut", **both** errors result! More important, specificity & sensitivity are functions of the "cutpoint" y.

SENSITIVITY

- With our assumption that larger values of Y are associated with the diseased population, the sensitivity, Pr(T=+|D=+), associated with a **cutpoint Y=y** is:

$$\mathbf{S^+(y) = Pr(Y>y|D=+)} = \text{“true positive rate”}$$

$$= 1 - \Pr(Y \le y|D=+) = 1 - F^+(y)$$

- where $F^+(y) = \Pr(Y \le y|D=+)$ is the **cumulative distribution function (cdf)** of Y for the diseased population (or population of cases).

SPECIFICITY

- With our assumption that larger values of Y are associated with the diseased population, the specificity, Pr(T=-|D=-), associated with a cutpoint Y=y is:

$S^-(y) = Pr(Y \leq y | D=-)$ = $F^-(y)$, or

$1 - S^-(y) = 1 - F^-(y)$ = "false positive rate"

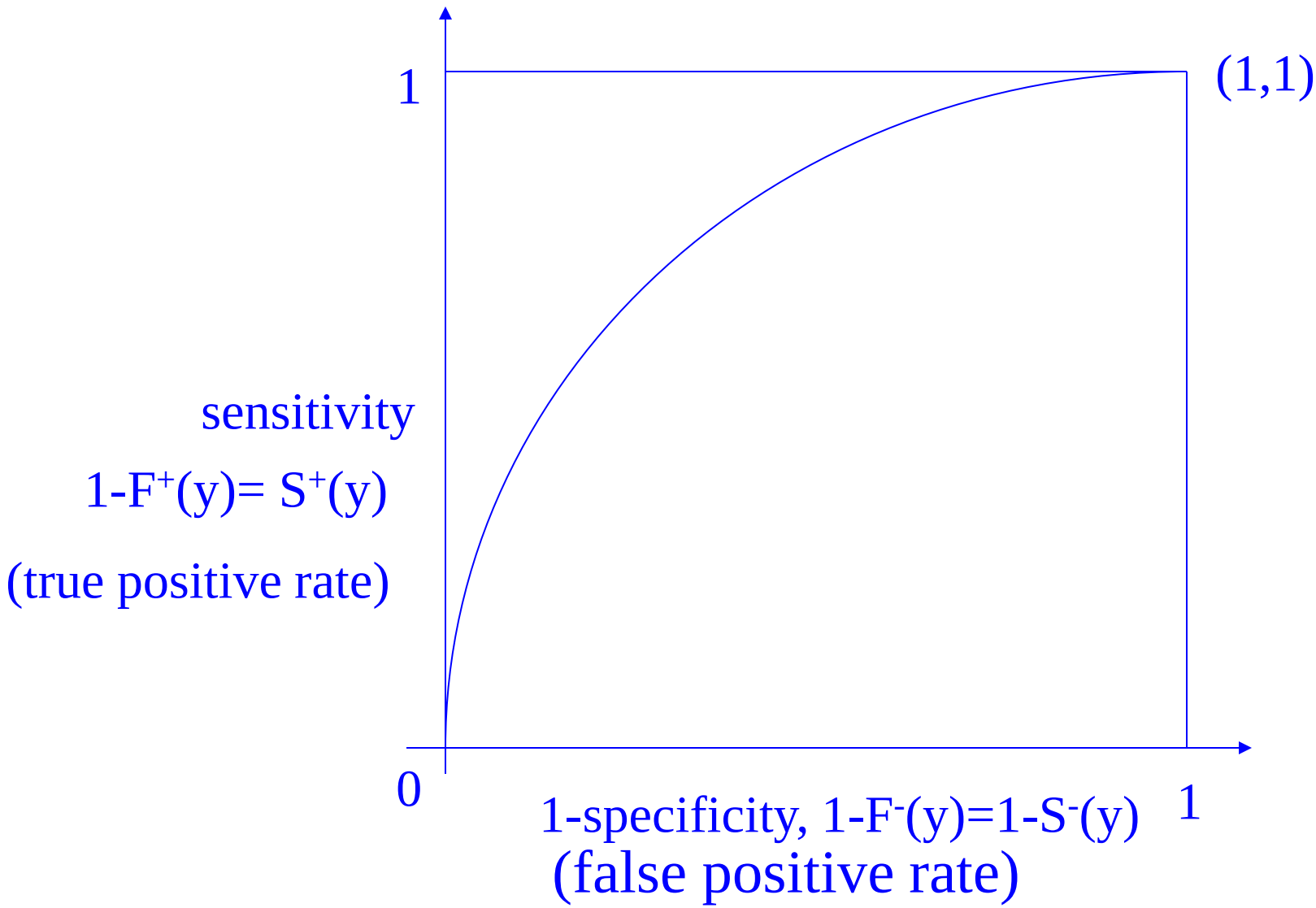
- where $F^-(x)$ is the cumulative distribution function (cdf) of Y for the non-diseased or healthy population.

ROC FUNCTION & ROC CURVE

- **A function "R" from [0,1] to [0,1] that "maps" false positive rate to true positive rate, $(1-F^-(y))$ to $(1-F^+(y))$, is called the "ROC function":**

  $R[1-F^-(y)] = 1-F^+(y)$ or $R[1-S^-(y)] = S^+(y)$
- **The graph of R(.) is called the "ROC curve"**
- **The ROC curve, the graph of sensitivity, $S^+(y)$, versus (1-specificity), $(1-S^-(y))$, is generated as the "cutpoint" y moves through its range of possible values.**

The “ROC function” maps “**sensitivity against (1-specificity)**” or “**true positive rate against false positive rate**”. It maps a survival function against another survival function”. In statistical terms, it maps “**statistical power against type I error**”.


1
(1,1)
sensitivity
1-F+(y)= S+(y)
(true positive rate)
0
1
1-specificity, 1-F-(y)=1-S-(y)
(false positive rate)

“ROC” Curve

As cutpoint "y" moves, $S^+(y)$ and $S^-(y)$ change; the ROC curve is a graphical device to show all possible combinations of sensitivity and specificity.

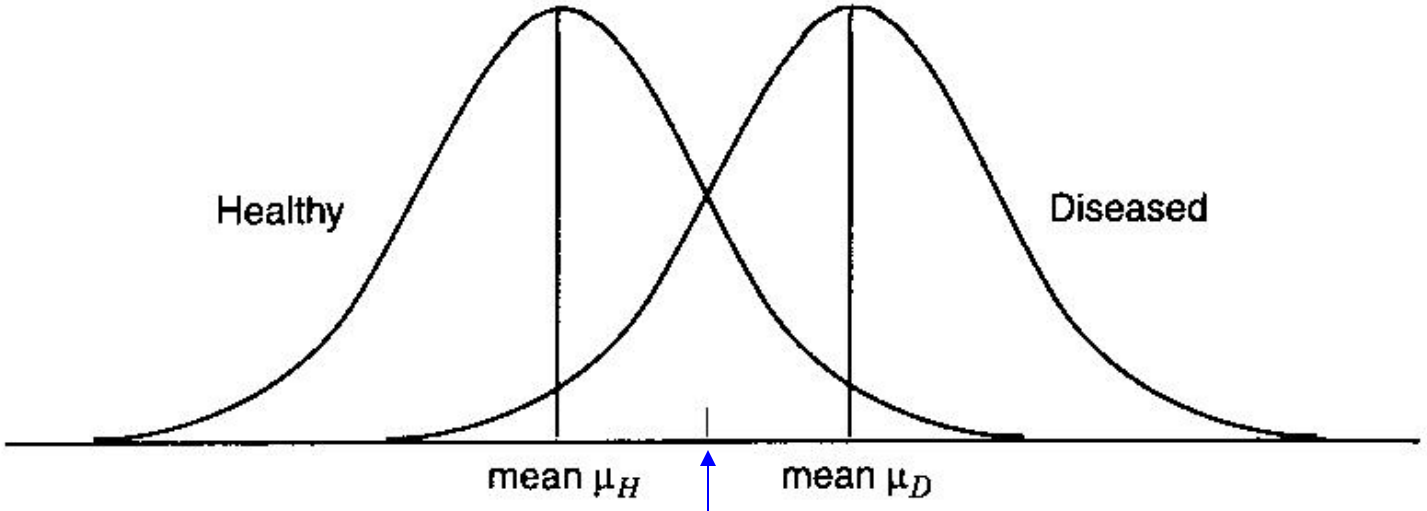


Figure Graphical display of a translational model of diseases.

A **Simple Model**: Separator Y is normally distributed with the same variance, but different means.

Issue: **How to estimate the ROC** curve given two independent samples, $\{y_{0i};\ i=1,\ldots,n_0\}$ and $\{y_{1j},\ j=1,\ \ldots,\ n_1\}$ from n_0 controls and n_1 cases?

EMPIRICAL ESTIMATE

- The simplest way to estimate R(.) is to replace cdfs $F^+(y)$ and $F^-(y)$ by their empirical estimates $p^+(y)$ and $p^-(y)$; **$p^+(y)$ is the proportion** of the n_1 observations y_{1j}'s of the cases which are less than or equal to y, and $p^-(y)$ is defined similarly.
- This is a non-parametric estimate and **$\{1-p^-(y), 1-p^+(y)\}$** is an unbiased estimator of $\{1-F^-(y), 1-F^+(y)\}$ but, as Bamber (1975) put it, "the sample ROC (for continuous Y) can never be anything but **a finite set of points**".
- If there are no ties in the combined sample of y_{0i}'s and y_{1j}'s, there **$(n_0 * n_1)$ points**.

CONNECTING THE DOTS

- Steck (1971) actually made an attempt to connect the dots, turning them into a step function.
- He combined 2 samples & in the usual increasing order.
- He described the empirical estimator as "a **random walk** from the bottom-left corner (0,0) to the top-right corner (1,1) – and **read the combined order sample <u>from largest to smallest</u> -** whose **next step is $1/n_1$ <u>up</u> or $1/n_0$ <u>to the right</u>** according to whether the next observation in the ordered combined sample is a case's measurement (y_1) or a control's measurement (y_0)".

Index for DIAGNOSTIC ACCURACY

- ROC curve is a **graphical device** to show all possible combinations of sensitivity and specificity but, for simplicity, it is desirable to **reduce an entire curve to a single quantitative index of diagnostic accuracy**.
- Possibilities include the difference between means of Y for the two populations, those with disease and those without; and the ratio of variances. However, the most popular one has been the **area under the ROC curve**.
- The area under the curve has a powerful **interpretation** and it is **related to other well-known statistics** making it easier to learn its statistical properties.

Suppose that an observation y_1 is randomly sampled from the diseased population and another random observation Y_0 is independently sampled from the non-diseased population; and let $\Pr(Y_1 > Y_0)$ denote the probability of the event that the Y_1 observation is larger than the Y_0 observation; we have:

$$A = \Pr(Y_1 > Y_0)$$

$$A = \int (1 - F^+) d(1 - F^-)$$

$$A = \int_0^1 R(u) du$$

$$A = \text{Area under ROC curve}$$

INTERPRETATION

- The area under the ROC curve measures the size and the importance of the difference between two populations, those with the disease and those without it.
- It tells how accurately a given diagnostic test differentiates two populations by giving the **probability of correct ranking**; the probability of separating a case from a control: a measure of "**separation power**"
- If Y_0 and Y_1 are identically distributed, A = 1/2; the maximum value of 1.0 is attained if and only if the cases' distribution lies entirely above that of controls.
- **Area under the ROC curve, A, is related to the Wilcoxon and Mann-Whitney statistics.**

AN ALTERNATIVE INDEX

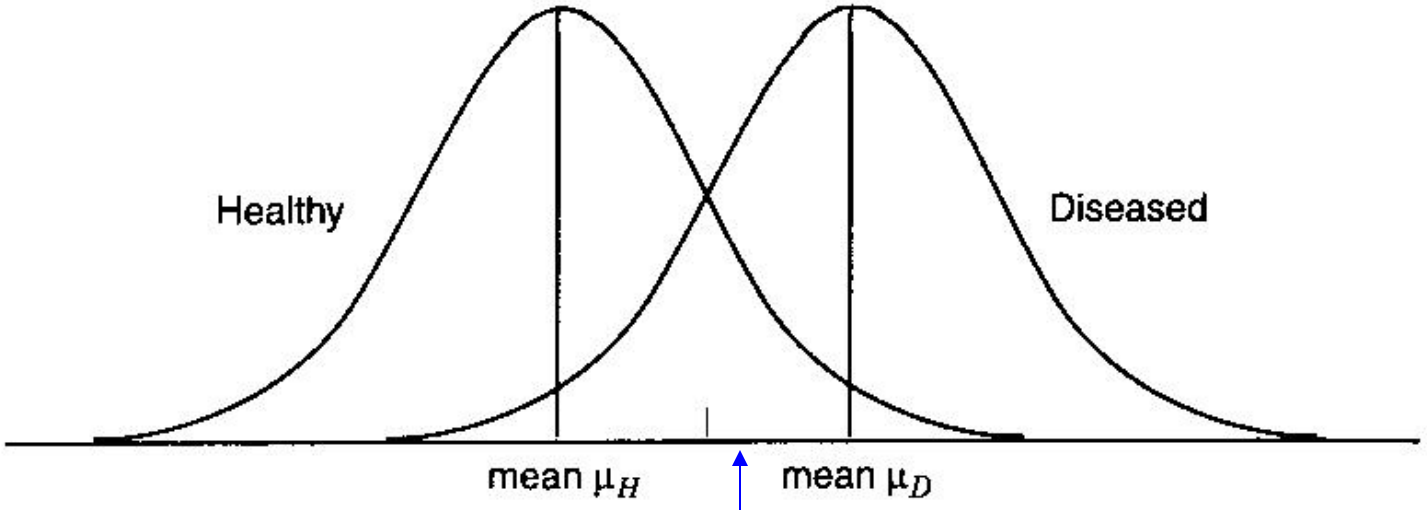


Figure Graphical display of a translational model of diseases.

T=- cutpoint T=+

Separator Y is normally distributed with the same variance, but different means; no matter where you "cut", both errors result!

The sizes of these errors depend on the "standardized distance"

$$\mathbf{d = (\mu_D - \mu_H)/\sigma}$$

Are the two indices, “A” and “d” different?

Yes, different numerical values but “**statistically equivalent**”:

$$d = \sqrt{2}\,\Phi^{-1}(A)$$

So, what is special about index “d”?

It has a very powerful interpretation in terms of disease development!

INTERPRETATION OF "d"

Under logistic model and Suppose Y is normally distributed with the same variance but different means for Pr(Y=y|D=+) and Pr(Y=y|D=-), then:

$$d = \beta_1 \sigma$$

The value of Index "d" is equal to the log(Odds Ratio) due to a change of "one SD" in the value of the marker Y

Solution for Optimization

Back to the case of “Diagnosis”. We all know that, for example, high PSA likely indicates prostate cancer; but **how high** it is **to classify a man as having prostate cancer**?

If we set the cut-point too high, we would miss cases – that is “**low sensitivity**”; if we set the cut-point too low, we would have many false positives – that is “**low specificity**”!

"How high is high?" or "How low is low?". In practice, cutpoints are formed but formed arbitrarily because we fail to form and justify a criterion or criteria.

We need an "**optimal cutpoint**" ; but what do we mean by "optimal"? "**Good**", but what it is good for? May be more than one solution because there are different criteria.

Basic Strategy/Criterion:

To determine an "optimal cutpoint" for a continuous marker by **maximizing the "Youden's Index"** of the dichotomized test.

Using this strategy, when using the resulting dichotomized test in a prevalence survey we would obtain an estimate with minimal error.

SIMPLE EMPIRICAL SOLUTION

- Pool the two samples and arrange in increasing order
- At **each midway between two data points**, calculate the Sensitivity S^+ and Specificity S^-; then the Youden's Index $J = S^+ + S^- - 1$
- Locate the cutpoint corresponding to the maximum value of J.

A NON-PARAMETRIC SOLUTION

- The ROC function R(.) maps ($U = 1-S^{-)}$ on the horizontal axis to ($V = S^{+)}$ on the vertical axis: $V = R(U)$
- The Youden's Index ($J = S^{+} + S^{-} - 1 = R(U) - U$) is maximized when: $0 = R'(U) - 1$, or $R'(U) = 1$.
- **Process:** (i) Smooth empirical estimate by any smoothing technique, (ii) **Locate the point with (slope = 1)** to obtain specificity, then (iii) Go to control sample to get cut-point. Of course, this is more difficult than the empirical solution.

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

SUGGESTED EXERCISE

#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.