Introduction to Clinical Trials - Day 1

Session 2 - Screening Studies

Presented July 23, 2018

Susanne J. May
Department of Biostatistics
University of Washington

Daniel L. Gillen
Department of Statistics
University of California, Irvine

Screening Trials

Need for exploratory science

- Before we can do a large scale, confirmatory Phase III trial, we must have
  - A hypothesized treatment indication to confirm
    - Disease
    - Patient population
    - Treatment strategy
    - Outcome
  - Comfort with the safety / ethics of human experimentation
- In “drug discovery", in particular, we will not have much experience with the intervention

Phases of investigation

- Preclinical
  - Epidemiology including risk factors
  - Basic science: Physiologic mechanisms
  - Animal experiments: Toxicology
- Clinical
  - Phase I: Initial safety / dose finding
  - Phase II: Preliminary efficacy / further safety
  - Phase III: Confirmatory efficacy / effectiveness
- Approval of indication based on total evidence to date
  - Evidence based medicine
  - (Phase IV: Post-marketing surveillance, REMS)

Case Study: Selenium for cancer prevention

Epidemiologic findings (Clark LC, *Nutr Cancer*. 1984;6(1):13-21)

- Case-control study: Plasma selenium and skin neoplasms:
  - 142 cases (basal cell epithelioma or squamous cell carcinoma); 103 noncancer controls.
    - Odds ratio = 4.39:
      lowest vs highest selenium decile (cases vs controls)

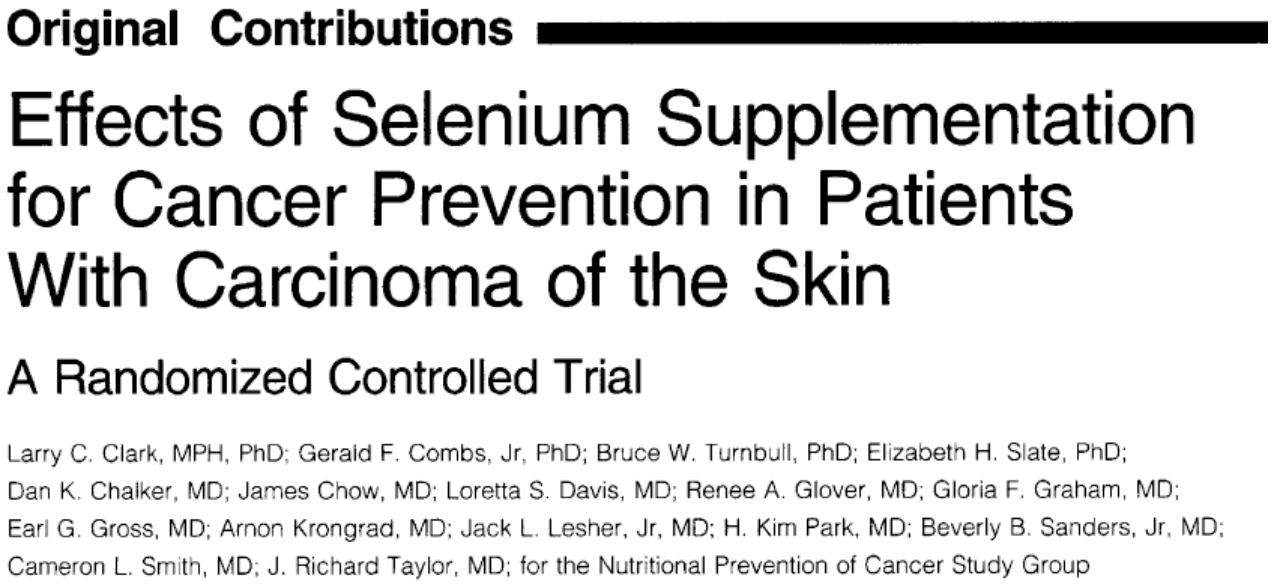
Abstract

Although experimental studies in animals show that selenium may prevent cancer, case-control studies of internal human cancers have been difficult to interpret because neoplastic tissue sequesters selenium. We therefore conducted a case-control study to examine the association between plasma selenium level and skin cancer, a neoplasm with minimal tumor mass at the time of diagnosis. The mean selenium level among patients with either basal cell epithelioma (N = 142), squamous cell carcinoma (N = 48), or both (N = 50), was 0.141 micrograms/g. This was significantly lower than the mean plasma selenium level of the 103 control subjects, which was 0.155 micrograms/g. The noncancer control groups were drawn from current clinic patients and past clinic patients. The logistic estimate of the odds ratio for the lowest versus the highest decile of selenium for all cases combined versus the group of current patient controls was 4.39; for all cases combined versus the past patient controls, the logistic estimate of the odds ratio was 5.81.

Case Study: Selenium for cancer prevention

Follow-up clinical trial (Clark, JAMA 1996; 276:1957-1963)

- Design: RCT (double-blind placebo-controlled; 1983-1991)
  - Dietary supplement: oral selenium (200μg) vs placebo
  - Patients with history of basal or squamous cell skin cancer
  - 1312 patents in seven dermatology clinics in eastern US



Original Contributions

Effects of Selenium Supplementation for Cancer Prevention in Patients With Carcinoma of the Skin

A Randomized Controlled Trial

Larry C. Clark, MPH, PhD; Gerald F. Combs, Jr, PhD; Bruce W. Turnbull, PhD; Elizabeth H. Slate, PhD; Dan K. Chalker, MD; James Chow, MD; Loretta S. Davis, MD; Renee A. Glover, MD; Gloria F. Graham, MD; Earl G. Gross, MD; Arnon Krongrad, MD; Jack L. Lesher, Jr, MD; H. Kim Park, MD; Beverly B. Sanders, Jr, MD; Cameron L. Smith, MD; J. Richard Taylor, MD; for the Nutritional Prevention of Cancer Study Group

Case Study: Selenium for cancer prevention

Clark trial results

- Lung cancer results (incident cases):
  - Selenium: 17 cases; Placebo: 31 cases
  - RR: 0.54 (95%CI: 0.30-0.98; p = 0.04)
- Prostate cancer results (incident cases):
  - Selenium: 13 cases; Placebo: 35 cases
  - RR: 0.37 (95%CI: 0.18-0.71; p = 0.002)

Table 3.—Cancer Incidence by Treatment Group

Cancer Sites, No.	Selenium	Placebo	RR (95% CI)*	*P* Value	HR (95% CI)†	*P* Value
Lung‡	17	31	0.54 (0.30-0.98)	.04	0.56 (0.31-1.01)	.05
Prostate‡	13	35	0.37 (0.18-0.71)	.002	0.35 (0.18-0.65)	.001
Colorectal‡	8	19	0.42 (0.18-0.95)	.03	0.39 (0.17-0.90)	.03
Head and neck	6	8	0.74 (0.21-2.43)	.58	0.77 (0.27-2.24)	.64
Bladder	8	6	1.32 (0.40-4.61)	.62	1.27 (0.44-3.67)	.66
Esophageal	2	6	0.33 (0.03-1.84)	.15	0.30 (0.06-1.49)	.14
Breast	9	3	2.88 (0.72-16.5)	.09	2.95 (0.80-10.9)	.11
Other specific carcinomas	5	9	0.55 (0.14-1.82)	.27	0.54 (0.18-1.62)	.27
Total carcinomas‡§	59	104	0.55 (0.40-0.77)	<.001	0.54 (0.39-0.75)	<.001
Melanomas	8	8	0.97 (0.32-2.96)	.91	0.92 (0.34-2.45)	.87
Leukemia/lymphomas	8	5	1.58 (0.46-6.14)	.41	1.50 (0.49-4.60)	.48
Other specific noncarcinomas	3	3	0.99 (0.13-7.37)	.98	0.99 (0.20-4.94)	.99
Total noncarcinomas	19	16	1.17 (0.57-2.44)	.65	1.16 (0.60-2.27)	.65
Total cancer‡§	77	119	0.63 (0.47-0.85)	.001	0.61 (0.46-0.82)	<.001

*RR indicates relative risk; and CI, confidence interval. P values derived from log rank tests.

Case Study: Selenium for lung cancer prevention

ECOG 5597

Phase III Chemoprevention Trial of Selenium Supplementation in Persons with Resected Stage I Non-Small Cell Lung Cancer

- Design: RCT (double-blind placebo-controlled)
  - Dietary supplement: oral selenium ($200\mu g$) vs placebo
  - Patients with resected stage I NSC lung cancer
  - 1522 patients from ECOG-participating clinics from 2000-2009.
- Results (interim analysis in 2009):
  - 5-year risk of recurrence or death: Selenium: 72%; Placebo 78%
  - Trial stopped early: “not an effective chemoprevention agent."

Case Study: Selenium for prostate cancer prevention

SELECT trial; JAMA. 2009 301(1): 39-51

- Randomized 35,533 men to 4 treatment groups (2×2 factorial:

  Selenium + Vit E placebo
  Selenium placebo + Vit E
  Selenium + Vit E
  Selenium placebo + Vit E placebo

- Follow-up for 4.17-7.33 years over 12 years

Effect of Selenium and Vitamin E on Risk of Prostate Cancer and Other Cancers
The Selenium and Vitamin E Cancer Prevention Trial (SELECT)

Case Study: Selenium for prostate cancer prevention

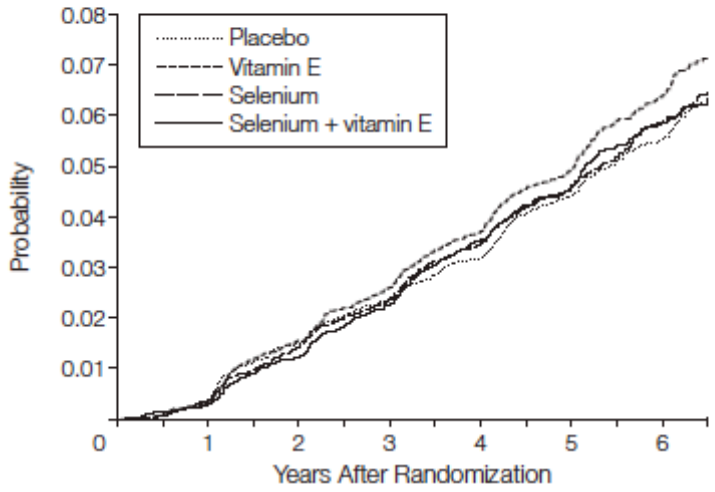
Follow-up clinical trial (Clark, JAMA 1996; 276:1957-1963)

SELECT trial results

Hazard ratios and 99% CI for prostate cancer:

Vit E:	1.13 (0.95 to 1.35)
Selenium:	1.04 (0.87 to 1.24)
Selenium + Vit E:	1.05 (0.88 to 1.25)

Figure 2. Cumulative Incidence of Prostate Cancer Detected Each Year by Intervention Group



No. at risk							
Placebo	8689	8553	8328	8039	7389	4892	2516
Vitamin E	8732	8610	8373	8098	7401	4867	2537
Selenium	8750	8597	8341	8083	7393	4848	2558
Selenium + vitamin E	8700	8585	8371	8097	7428	4894	2580

Compared with placebo, there was a statistically nonsignificant increase in prostate cancer in the vitamin E group (P=.06) and not in the selenium + vitamin E group (P=.52) or the selenium group (P=.62).

Case Study: Selenium for cancer prevention

Summary remarks

The selenium story represents:

- Excellent demonstration of careful evaluation of a hypothesis illustrating:
  - Interplay between careful epidemiology and clinical trials in a range of diseases:
    1. Epidemiology as foundation for major intervention trials.
    2. Demonstrates the importance of confirmatory trials for subgroup effects in large trials.
    3. Large RCT's of the same hypothesis in multiple diseases
    4. The question has been answered??

Screening Trials

Phases of investigation

- Preclinical
  - Epidemiology including risk factors
  - Basic science: Physiologic mechanisms
  - Animal experiments: Toxicology
- Clinical
  - Phase I: Initial safety / dose finding
  - Phase II: Preliminary efficacy / further safety
  - Phase III: Confirmatory efficacy / effectiveness
- Approval of indication based on total evidence to date
  - Evidence based medicine
  - (Phase IV: Post-marketing surveillance, REMS)

Phase II clinical trials : Screening

- Phase II clinical trials seek to establish preliminary evidence of efficacy
- Goals:
  - Screening for any evidence of treatment efficacy
  - Incidence of major adverse effects
  - Decide if worth studying in larger samples
    - Gain information about best chance to establish efficacy
    - Choose population, treatment, outcomes
- This initial screening is essential for achieving the following public health objectives...

Public Health Objective for Clinical Research

Formulating the public health objective

- Ultimate objectives:
  - Discover things that are true
  - Develop the science in order to provide public health benefit (therapies, prevention, etc...)
  - Want high prevalence of truly beneficial therapies/practices among all things (therapies or public health recommendations) that are adopted in practice.
- These objectives are quantified as the *positive predictive value* (PPV) of clinical research
  - Medical studies as diagnostic tests
  - Review of PPV: cervical cancer screening
  - PPV in clinical trials
    - Illustration of practices to increase (or decrease) PPV.

Public Health Objective for Clinical Research

Medical studies as diagnostic tests

- Clinical testing of a new treatment or preventive agent is analogous to using laboratory or clinical tests to diagnose a disease
  - Goal is to find a procedure that identifies truly beneficial interventions
- Not surprisingly, the issues that arise when screening for disease apply to clinical trials
  - Predictive value of a positive test is best when prevalence is high
  - Use screening trials to increase prevalence of beneficial treatments

Public Health Objective for Clinical Research

Diagnostic testing

- We most often characterize the *sensitivity* and *specificity* of a diagnostic/screening test
  - Sensitivity of test: Probability of positive in diseased
    - Sample a cohort of subjects with the disease
    - Estimate the proportion who have a positive test result:

$$\text{Sensitivity} = \Pr[+|D]$$

    - 1 - False Negative Rate
  - Specificity of test: Probability of negative in healthy
    - Sample a cohort of healthy (non-diseased) subjects
    - Estimate the proportion who have a negative test result:

$$\text{Specificity} = \Pr[-|\bar{D}]$$

    - 1 - False Positive Rate

Public Health Objective for Clinical Research

Diagnostic testing

- We are actually interested in the diagnostic utility of the test:
  - Predictive value of a positive test: Probability of disease when test is positive

$$\text{PPV} = \Pr[D|+]$$

  - Predictive value of a negative test: Probability of not diseased when test is negative

$$\text{NPV} = \Pr[\bar{D}|-]$$

Public Health Objective for Clinical Research

Diagnostic testing

- We can compute the predictive value of positive and negative tests using *Bayes rule*:

$$\Pr[D|+] = \frac{\Pr[+|D]\Pr[D]}{\Pr[+|D]\Pr[D] + \Pr[+|\bar{D}]\Pr[\bar{D}]}$$

$$\Pr[\bar{D}|-] = \frac{\Pr[-|D]\Pr[D]}{\Pr[-|D]\Pr[D] + \Pr[-|\bar{D}]\Pr[\bar{D}]}$$

Public Health Objective for Clinical Research

Diagnostic testing

- Key property: Positive and Negative predictive value depends upon sensitivity, specificity, AND prevalence of disease

$$\Pr[D|+] = \frac{\Pr[+|D]\Pr[D]}{\Pr[+|D]\Pr[D] + \Pr[+|\bar{D}]\Pr[\bar{D}]}$$

$$\text{PPV} = \frac{\text{Sens} \times \text{Prev}}{\text{Sens} \times \text{Prev} + (\text{1-Spec}) \times (\text{1-Prev})}$$

$$\Pr[\bar{D}|-] = \frac{\Pr[-|\bar{D}]\Pr[\bar{D}]}{\Pr[-|D]\Pr[D] + \Pr[-|\bar{D}]\Pr[\bar{D}]}$$

$$\text{NPV} = \frac{\text{Spec} \times (\text{1-Prev})}{(\text{1-Sens}) \times \text{Prev} + \text{Spec} \times (\text{1-Prev})}$$

PPV Example: Cervical cancer screening in New Zealand

New Zealand National Cervical Screening Program (NCSP)

- Established in 1991: credited with reducing cervical cancer incidence and mortality.
- Over 70% participation
- Two screening tests (circa 2000)
  - Pap smear ($\sim$\$5): funded by NCSP
  - ThinPrep ($\sim$\$20): offered by some physicians for \$15 fee

ThinPrep versus Pap (Stein 2003)

- Pap smear (Papanicolaou test)
  - Cervical swab on slide for pathologist evaluation
  - Sensitivity $\sim$ 50% (up to 68?%)
  - Specificity $\sim$ 98% (up to 79%)
- “ThinPrep": liquid-based cytology screening test
  - Cervical swab rinsed in tube with liquid preservative
  - Sensitivity $\sim$ 80%
  - Specificity $\sim$ 90%

PPV Example: Cervical cancer screening in New Zealand

NCSP and equitably (*circa 2000*)

- Lower SES communities unable to pay for ThinPrep
- *IF* superior should NCSP adopt ThinPrep?
- Key questions:
  1. Is ThinPrep really more accurate than pap?
  2. What are the potential cost impacts?
     - ThinPrep costs $15 more
     - A positive screening test is referred for colposcopy ($200).
     - Lower specificity might overwhelm budget with unnecessary colposcopies.

PPV Example: Cervical cancer screening in New Zealand

Impact of sensitivity and specificity on the NCSP

- Suppose :
  - 1,000,000 women are screened
  - Prevalence of high grade lesions is 1%:
    - 10,000 with high grade lesion
    - 990,000 without high grade lesion
  - Each positive test is sent for colposcopy

PPV Example: Cervical cancer screening in New Zealand

Impact of sensitivity and specificity on the NCSP

- Suppose NCSP uses pap:
  - Sensitivity = 50%
  - Specificity = 98%

- Results of screening:
  - Number of positive tests:

$$\begin{aligned} \text{True positive tests:} \quad & 10,000 \times 0.50 = 5000 \\ \text{False positive tests:} \quad & 990,000 \times 0.02 = 19,800 \\ \text{PPV}: \quad & \frac{5000}{24,800} = 0.20 \end{aligned}$$

  - Cost:

$$\begin{aligned} \text{Cost of tests:} \quad & \$5.00M \\ \text{Cost of colposcopy:} \quad & \$4.96M \\ \text{Total:} \quad & \sim \$10M \end{aligned}$$

PPV Example: Cervical cancer screening in New Zealand

Impact of sensitivity and specificity on the NCSP

- Suppose NCSP uses ThinPrep:
  - Sensitivity = 80%
  - Specificity = 95%

- Results of screening:

  Number of positive tests:

$$\begin{aligned} \text{True positive tests:} \quad & 10,000 \times 0.80 = 8000 \\ \text{False positive tests:} \quad & 990,000 \times 0.05 = 49,500 \\ \text{PPV}: \quad & \frac{8000}{57,500} = 0.14 \end{aligned}$$

  Cost:

$$\begin{aligned} \text{Cost of tests:} \quad & \$20M \\ \text{Cost of colposcopy:} \quad & \$11.5M \\ \text{Total:} \quad & \sim \$31.6M \end{aligned}$$

PPV Example: Cervical cancer screening in New Zealand

Impact of sensitivity and specificity on the NCSP

- Suppose NCSP uses ThinPrep:
  - Sensitivity = 80%
  - Specificity = 90%

- Results of screening:

  Number of positive tests:

$$\begin{aligned} \text{True positive tests:} \quad & 10,000 \times 0.80 = 8000 \\ \text{False positive tests:} \quad & 990,000 \times 0.1 = 99,000 \\ \text{PPV}: \quad & \frac{8000}{99,900} = 0.075 \end{aligned}$$

  Cost:

$$\begin{aligned} \text{Cost of tests:} \quad & \$20.0M \\ \text{Cost of colposcopy:} \quad & \$21.4M \\ \text{Total:} \quad & \sim \$41.4M \end{aligned}$$

PPV Example: Cervical cancer screening in New Zealand

Summary remarks: public health objective

- Rare diseases:
  - High risk for false positive
  - Important to control *specificity*
- Consequences of a false positive
  - Costs to healthcare system
  - Anxiety costs for women
- Clearly:
  - Weigh costs against risk/consequences of false negative
- Public health objective:
  - Highest PPV for lowest total cost

Formulating the public health objective

PPV as the objective in public health research

- So what is the right answer?
  - Diagnostic testing
    - Identify people with disease who can benefit from care
    - Identify people who should not be treated
  - Public health research?
    - Identify hypotheses that are in fact true
    - Identify hypotheses that are not worthy of further exploration
- What are the consequences of a wrong answer?
  - Diagnostic testing?
    - People do not receive beneficial treatment
    - People receive non-beneficial treatment
  - Public health research?
    - Populations do not receive beneficial practice/care
    - Populations receive non-beneficial practice/care
- Objective:
  - Maximize the proportion of right answers (PPV)

Positive predictive value of research

PPV in research

- A Statistical hypothesis test can be viewed as a test for beneficial treatments.
  - α-level: probability of observing a positive (statistically significant) test in absence of a true treatment effect:
    - Level of significance is 1 - specificity.
    - Choosing $\alpha = 0.05$ gives 95% specificity.
  - Statistical power (β): Probability of observing a positive (statistically significant) test when there is a true treatment effect:
    - Power is sensitivity.
    - Common choice of 80% sensitivity (not usually recommended by me).
  - Prevalence (π_0): the percentage of effective treatments among all tested treatments.

Positive predictive value of research

PPV in research

- Positive predictive value: probability that a statistically significant trial indicates a truly effective treatment.

$$PPV = \frac{\beta\pi_0}{\beta\pi_0 + \alpha(1-\pi_0)}$$

- The probability that our public health recommendation is in fact beneficial.

PPV in biomedical research

Example: The Amgen experience

- CG Begley and LM Ellis:
  " Raise the standards for preclinical cancer research"
  Nature 483:531-533; 2012
  - Over the past decade Amgen scientists tried to confirm the results of 53 'landmark' studies
  - Only 6/53 (11%) of these results were confirmed
  - *"The scientific process demands the highest standards of quality, ethics, and rigour."*
- All true:
  - High standards are an absolute requirement.
  - Also need to note that lack of reproducibility is not surprising if initial false-positive risk is high

The Public Health Objective

Clinical trials as diagnostic tests

- We routinely consider power (β = sensitivity) and type I error (α = 1 - specificity).
- What is the prevalence (π_0)?
  - 9.6% of treatments entering phase I trials are positive in subsequent phase III trials (Biomedtracker, 2016)
    - Lowest in oncology (7.0%)
    - Highest in infectious diseases (17%) (Nature Biotech, 2014)
  - Results of NCI-sponsored trials 1955-2006 (Djulbegovic, 2008)
    - 743 randomized comparisons, 176 (24%) are significant
    - 116 (15%) discover 'breakthrough interventions'.
  - Results of phase II cancer trials (Hay, Nature Biotech, 2016)
    - 28% led to phase III.
  - Prevalence of truly beneficial treatments entering phase II trials is probably less than 10%.

The Public Health Objective
How do clinical trials determine PPV?

Example: Phase II studies as screening tests

- Consider the following approaches to evaluating new treatments:
  1. Study every treatment in a large definitive experiment.
  2. Perform small screening tests, and perform large definitive experiments only in those treatments that pass the screening tests.

- Suppose that we want to evaluate the efficiency of these strategies. Assume:
  - 10% of all treatments actually work.
  - Level of significance = 0.05 (specificity = 0.95).
  - 1,000,000 subjects are available for clinical trials.
  - Power for a clinically important difference:

$$\begin{aligned} 1000 \text{ subjects} &\rightarrow 97.5\% \text{ power} \\ 500 \text{ subjects} &\rightarrow 80\% \text{ power} \\ 50 \text{ subjects} &\rightarrow 15\% \text{ power} \end{aligned}$$

The Public Health Objective
How do clinical trials determine PPV?

Example: Phase II studies as screening tests

- Scenario 1 (only large trials):
  - Suppose we evaluate 1000 new treatments (100 effective and 900 ineffective) with 1000 subjects per trial.
  - On average we have positive tests for:
    - 98 of the 100 effective treatments ($0.975 \times 100 \approx 98$).
    - 45 of the 900 ineffective treatments ($0.05 \times 900 = 45$).
  - PPV: $98/(45 + 98) = 0.69$; that is, only 69% of the 143 treatments identified actually work.

The Public Health Objective
How do clinical trials determine PPV?

Example: Phase II studies as screening tests

- Scenario 2 (preliminary screening trials):
  - (a) Suppose we first screen 12,500 new treatments (1,250 effective and 11,250 ineffective).
    - Using 50 subjects in the screening trials (625,000 total) with 15% power.
    - On average the screening trials give positive tests for:
      - 187 of the 1,250 effective treatments ($0.15 \times 1250 \approx 187$).
      - 562 of the 11,250 ineffective treatments ($0.05 \times 11250 \approx 562$).
    - PV+ for the screening phase: $187/(187 + 562) = 0.25$.
  - (b) Now evaluate the 749 treatments (187 effective and 562 ineffective) from the screening trials.
    - Using 500 subjects per trial (374,500 total) with 80% power.
    - On average these confirmatory trials give positive tests for:
      - 150 of the 187 effective treatments ($0.8 \times 187 \approx 150$).
      - 28 of the 562 ineffective treatments ($0.05 \times 562 \approx 28$).
    - PV+ for confirmatory trials: $150/178 = 0.84$.

The Public Health Objective
How do clinical trials determine PPV?

Example: Phase II studies as screening tests

- Comparison of scenarios:
  - Scenario 1 (large trials only):
    - Use 1,000,000 subjects
    - Screen 1,000 new treatments
    - Adopt 98 effective treatments
    - Adopt 45 ineffective treatments
    - PPV = 98/143 = 0.69
  - Scenario 2 (screening studies followed by large trials):
    - Use 999,500 subjects
    - Screen 12,500 new treatments
    - Adopt 150 effective treatments
    - Adopt 28 ineffective treatments
    - PPV = 150/178 = 0.84

The Public Health Objective
How do clinical trials determine PPV?

Example: Phase II studies as screening tests

- Bottom line:
  - Using the same number of subjects, phase II studies increase the predictive value of a positive study. A greater number of effective treatments are identified due in part to the greater number of treatments screened.
  - (Different choices of statistical power in screening and confirmatory trials can be used to optimize the strategy for a particular setting.)

The Public Health Objective
How do clinical trials determine PPV?

PPV is increased through good experimental practice

- How do we increase PPV in the clinical trials paradigm?

$$PPV = \frac{\beta\pi_0}{\beta\pi_0 + \alpha(1-\pi_0)}$$

1. Increase π_0:
   - Careful planning of preliminary studies (choice of endpoints; patient population)
   - Avoid "novel" and "innovative" ideas
   - Careful specification of hypothesis-driven research (avoid “science by hunch")

The Public Health Objective
How do clinical trials determine PPV?

Sensitivity to π_0 (how likely is it that the new treatment works?)

1a. Trial of an 'incremental' advance for a known compound:

- $\pi_0 = 0.20$; $\alpha_2 = 0.05$; $\beta_2 = 0.15$; $\alpha_3 = 0.05$; $\beta_3 = 0.80$
- Results:

	Trials	True Pos	False Pos	PPV
Phase 2	11765	353	471	0.43
Phase 3	824	282	24	0.92

1b. Trial of a novel and innovative therapy:

- $\pi_0 = 0.01$; $\alpha_2 = 0.05$; $\beta_2 = 0.15$; $\alpha_3 = 0.05$; $\beta_3 = 0.80$
- Results:

	Trials	True Pos	False Pos	PPV
Phase 2	13245	20	656	0.029
Phase 3	675	16	33	0.33

The Public Health Objective
How do clinical trials determine PPV?

PPV is increased through good experimental practice

- How do we increase PPV in the clinical trials paradigm?

$$PPV = \frac{\beta\pi_0}{\beta\pi_0 + \alpha(1-\pi_0)}$$

2. Increase β:

    - Good practice (no missing data, low variation in outcome assessment, good adherence, etc.)
    - Increase sample size.

The Public Health Objective
How do clinical trials determine PPV?

Sensitivity to β_3 (ultimate sensitivity for effective therapies)

2a. Sufficiently powered phase III ($\beta_3 = 0.975$)

- $\pi_0 = 0.10$; $\alpha_2 = 0.05$; $\beta_2 = 0.15$; $\alpha_3 = 0.05$; $\beta_3 = 0.975$
- Results:

	Trials	True Pos	False Pos	PPV
Phase 2	9091	136	409	0.25
Phase 3	545	133	20	0.87

2b. Underpowered phase III:

- $\pi_0 = 0.10$; $\alpha_2 = 0.05$; $\beta_2 = 0.15$; $\alpha_3 = 0.05$; $\beta_3 = 0.50$
- Results:

	Trials	True Pos	False Pos	PPV
Phase 2	15385	231	692	0.25
Phase 3	923	115	35	0.77

The Public Health Objective
How do clinical trials determine PPV?

PPV is increased through good experimental practice

- How do we increase PPV in the clinical trials paradigm?

$$PPV = \frac{\beta\pi_0}{\beta\pi_0 + \alpha(1-\pi_0)}$$

3. Reduce α:

   - Pre-specify outcomes
   - Pre-specify all analyses
   - Avoid multiple comparisons
   - Avoid surrogate outcomes
   - Avoid subgroups

The Public Health Objective
How do clinical trials determine PPV?

Sensitivity to α (false positive risk; specificity)

3a. Relax phase II alpha ($\alpha_2 = 0.20$)

- $\pi_0 = 0.10$; $\alpha_2 = 0.20$; $\beta_2 = 0.15$; $\alpha_3 = 0.05$; $\beta_3 = 0.80$
- Results:

	Trials	True Pos	False Pos	PPV
Phase 2	6780	102	1220	0.077
Phase 3	1322	81	61	0.571

3b. Relax both phase II and III alpha ($\alpha_2 = 0.2$, $\alpha_3 = 0.10$):

- $\pi_0 = 0.10$; $\alpha_2 = 0.20$; $\beta_2 = 0.15$; $\alpha_3 = 0.10$; $\beta_3 = 0.80$
- Results:

	Trials	True Pos	False Pos	PPV
Phase 2	6780	102	1220	0.077
Phase 3	1322	81	122	0.40

The Public Health Objective
How do clinical trials determine PPV?

Summary: PPV as a function of π_0, α, and β

Scenario	π_0	α_2	β_2	α_3	β_3	Drugs Evaluated	True Pos	False Pos	PPV
1	0.10	*	*	0.05	0.800	1000	98	45	0.685
2	0.10	0.05	0.15	0.05	0.800	12500	150	28	0.842
3	0.20	0.05	0.15	0.05	0.800	11765	282	24	0.923
4	0.01	0.05	0.15	0.05	0.800	13265	16	33	0.327
5	0.10	0.05	0.15	0.05	0.975	9091	133	20	0.867
6	0.10	0.05	0.15	0.05	0.500	15385	115	35	0.769
7	0.10	0.20	0.15	0.05	0.800	6780	81	61	0.571
8	0.10	0.20	0.15	0.10	0.800	6780	81	122	0.400

The Public Health Objective
PPV and good science?

Summary remarks: How to get high PPV with fewer trials

- Design for scientifically informative negative trials
  - All trials (positive or negative) must reduce the number of viable hypotheses.
- Accept that *no means no*
  - *(Never give up)*: Avoid inflating α with
    - multiple endpoints
    - subgroup analyses
    - surrogate endpoints
  - *(Try try again)*: Avoid recycling ideas

The Public Health Objective
PPV and good science?

Summary remarks: How to get high PPV with fewer trials

- Assure power (β)
  - Good practice reduces variability
  - Good recruitment/retention
  - Adequate sample size
- Avoid development programs with low pre-test probability (π_0)
  - “Novel" and “innovative" approaches have low π_0.

The Public Health Objective
PPV and good science?

Summary remarks: How to get high PPV with fewer trials

- Financial analysts are even quick to pick up on this! (Grainger, Forbes, 2015)
  - "More insidiously, the pivotal trials often adopt a different end-point, agreed with the regulators, to the Phase 2 trials (where a 'surrogate' end-point was used to predict whether the regulatory end-point is likely to be met). Unless this surrogate is perfect (and few are), some agents that are positive against the surrogate will be ineffective against the regulatory end-point."
  - "Similarly, the pivotal trials need to be performed in less selected patient populations. If the drug is more effective in the defined subset studied in Phase 2 than in the broader population, unexpected failure will again result."

For a discussion across all biomedical studies, see "Why Most Published Research Findings Are False", J. Ioannidis, *PLOS Medicine*, August 2005