How likely is a positive finding to be wrong?

Sholom Wacholder, DCEG, NCI

January, 2005, Bethesda

Introduction

- **Dr. Zerhouni's Analogy:**
  - M & M conference and this meeting
  - How can we change practice to do better in the future?
- **Dr. Ransohoff: Should rules of evidence be changed?**
- **I will not talk**
  - **Asking the right question**
    - Determination and assessment
      - Exposure
      - Endpoint
  - **Getting the timing right**
  - **Bias reduction**
    - Design
    - Fieldwork
    - Analysis
- **Instead, I want to question a venerable convention**

Standard practice

- P-values, confidence intervals
  - Do not convey the level of uncertainty about the hypothesis when the statistical test is "significant"
  - Do not provide a sense of the chance that the significant finding is "wrong"
- Is 5% the best criterion for significant two-sided p-value?
- Do all CI's need to be 95% CIs?

Decision making

- Clinic
  - Do I screen?
  - Screening modality
  - Decision rule on the basis of results of the screen?
- Statistics
  - Should I launch a study?
  - Study design
    - Sample size
  - How do I act on the basis of results of a study?
- Parallelism
  - Browner & Newman, JAMA, 1987

Statistical decisions

- Accept or reject null hypothesis?
- Stop a randomized trial to protect participants from excess of serious adverse events?
- Recommend a change in behavior to reduce risk
- Act as if a hypothesis is no longer viable
  - Based on accumulated negative evidence

Basis of Statistical Decisions

- Loss from wrong decisions
  - Two kinds of loss
    - False positive decision
    - False negative decision
    - Just like PPV and NPV
      - Positive and negative predictive value
  - Expected loss depends on
    - Likelihood of each type of wrong decisions
      - Relative magnitude is enough
      - Depends on context
    - Probability the hypothesis is true
      - Unknowable ... but

Standard statistical decision making

- α=0.05 is universal
  - Standard sample size determination
    - Sample size vs power for α=0.05
  - Analysis
  - Prior probability not considered formally
  - Loss from bad decisions not considered
- →, probability that positive report is a false positive is not considered

Example of algebra of false positives for speculative H_A

- Chance alternative hypothesis H_A is true = 0.1%=1/1,000=0.001
- If H_A false ➔ 5% chance of rejection
- If H_A true ➔ 100% chance of rejection
- Pr(reject & H_A false)=0.999*0.05 ≈ 0.050
- Pr(reject & H_A true) =0.001*1.00 = 0.001
- FPRP = Pr(H_A false | rejection)

  ≈ 0.050/(0.001+0.050) ≈ 98%

How likely is a positive finding to be wrong?

- Calculate FPRP (JNCI, 2004)
  - False positive report probability
  - Analogous to 1-PPV
  - Uses p-value, power and prior probability that there is an association
- Base decision on FPRP-based test of "noteworthiness"
  - "reject" if FPRP < 0.2 (or 0.5, perhaps)
  - Or choose α so that FPRP < 0.2 if test rejected
- Interpretation
  - Bayesian
  - Frequentist, but with study-specific α-level
    - Based on prior probability of hypothesis and power
    - Accounting for "loss" from wrong decisions

Essential Formula

FPRP: FALSE POSITIVE-REPORT PROBABILITY

- Prior: π = Pr(association)
- Power: 1-β = Pr(Rejection | association)
- Size: α = Pr(Rejection | no association)
- FPRP = Pr(No association | Rejection)

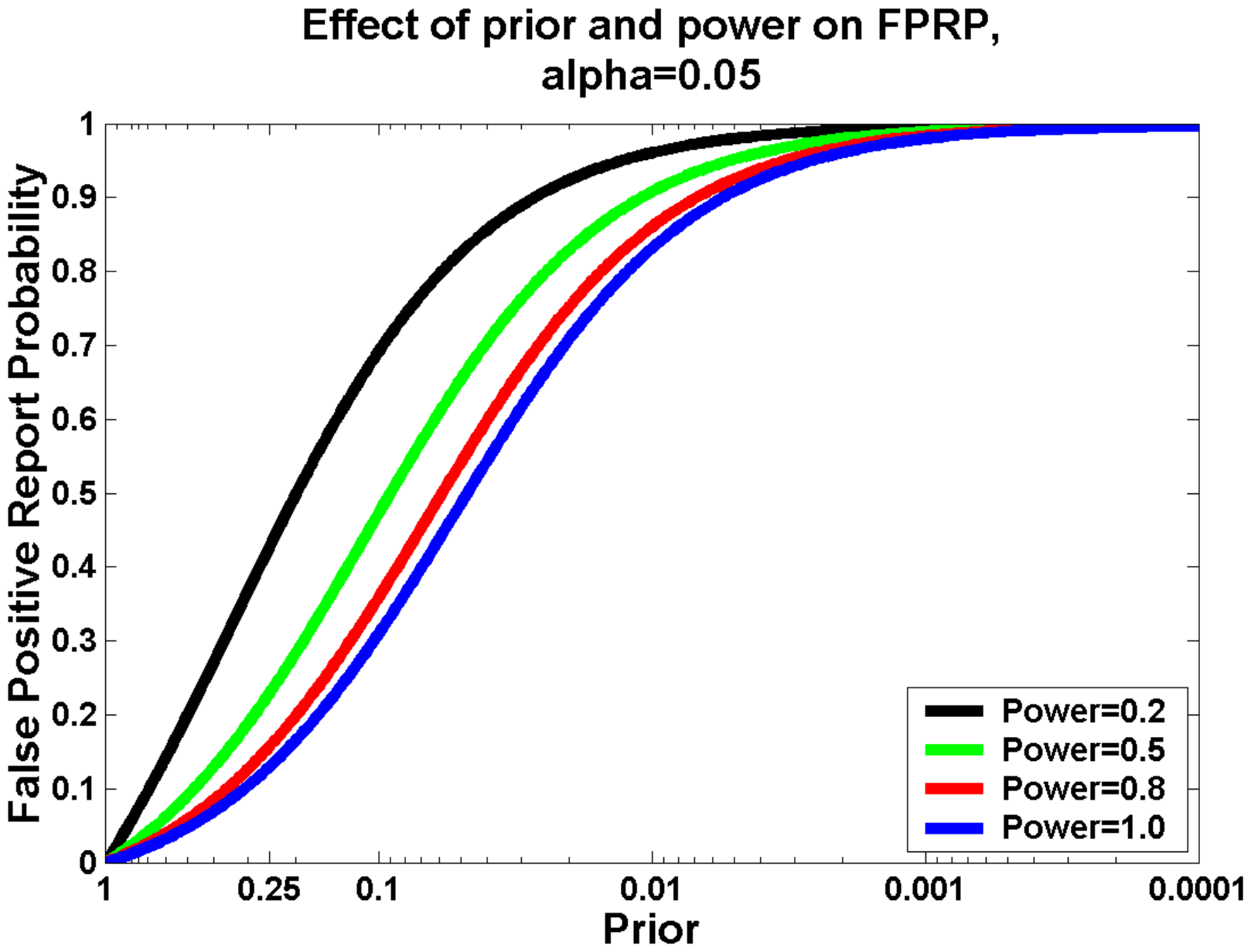
$$FPRP = \frac{\alpha(1-\pi)}{\alpha(1-\pi)+(1-\beta)\pi} = \frac{1}{1+\frac{(1-\beta)}{\alpha}\frac{\pi}{(1-\pi)}}$$

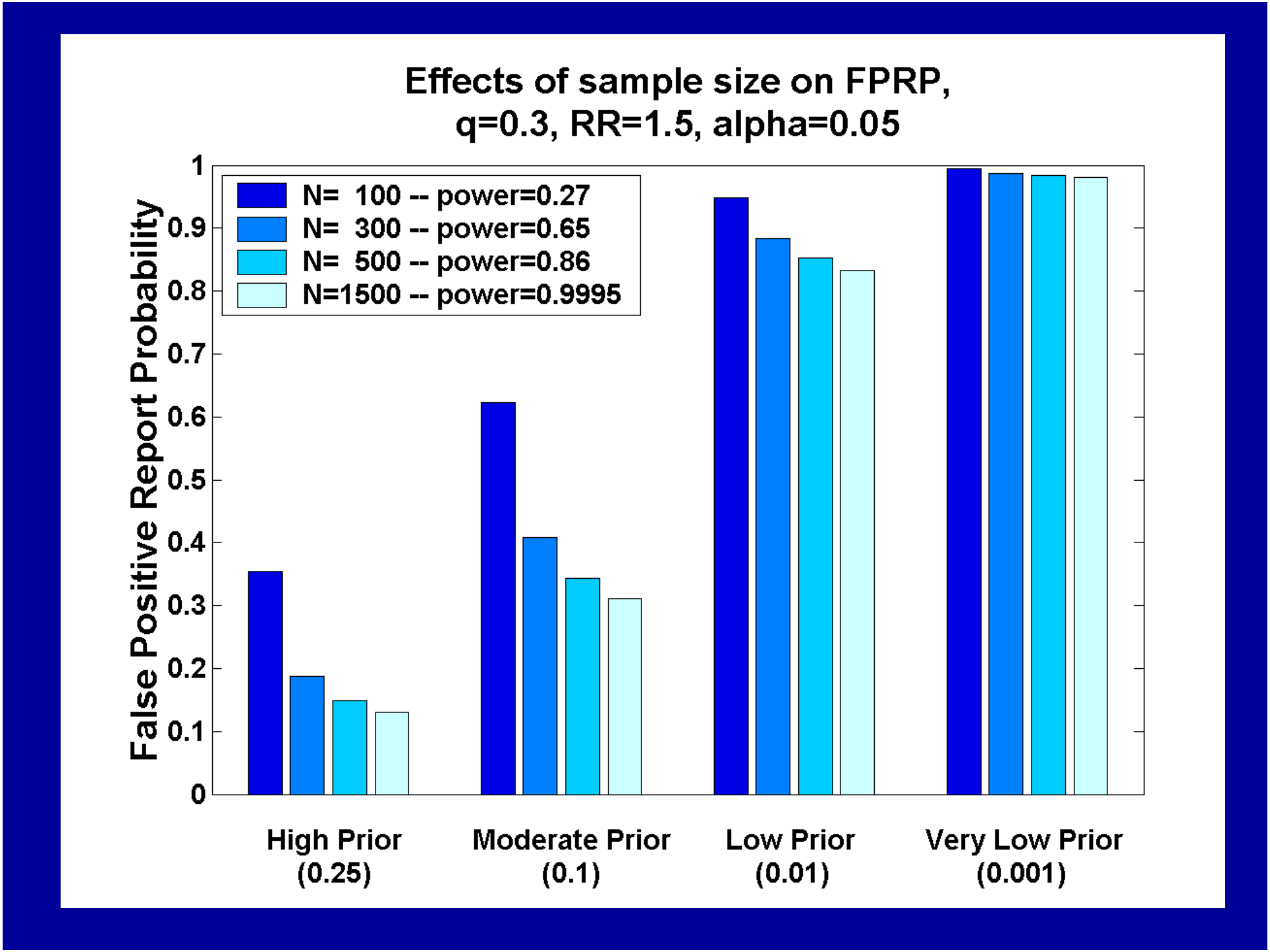
Essential Formula II

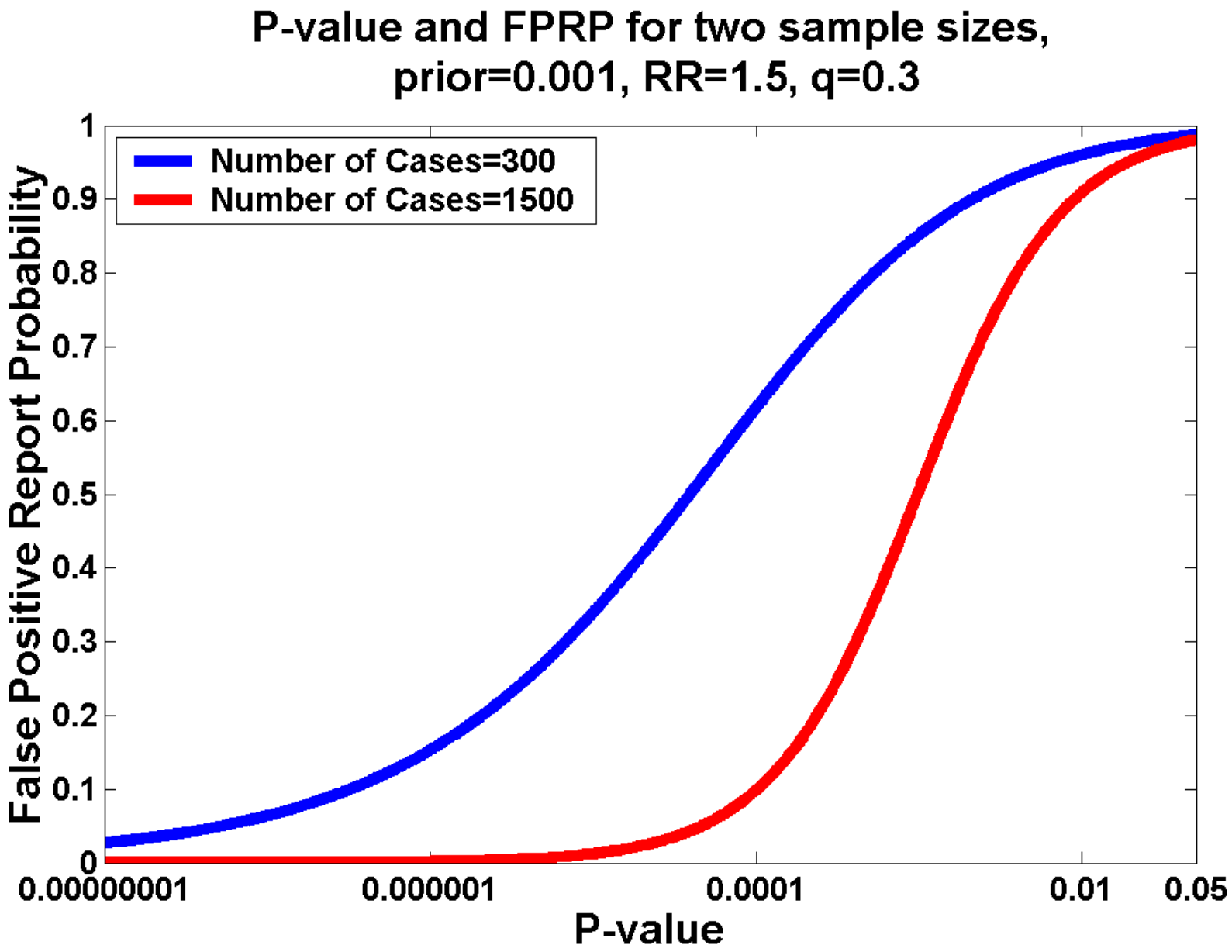
- Prior: π = Pr(association)
- Power: 1-β = Pr(Reject | association)
- Size: α = Pr(Reject | no association)
- FNRP: False Negative Report Probability

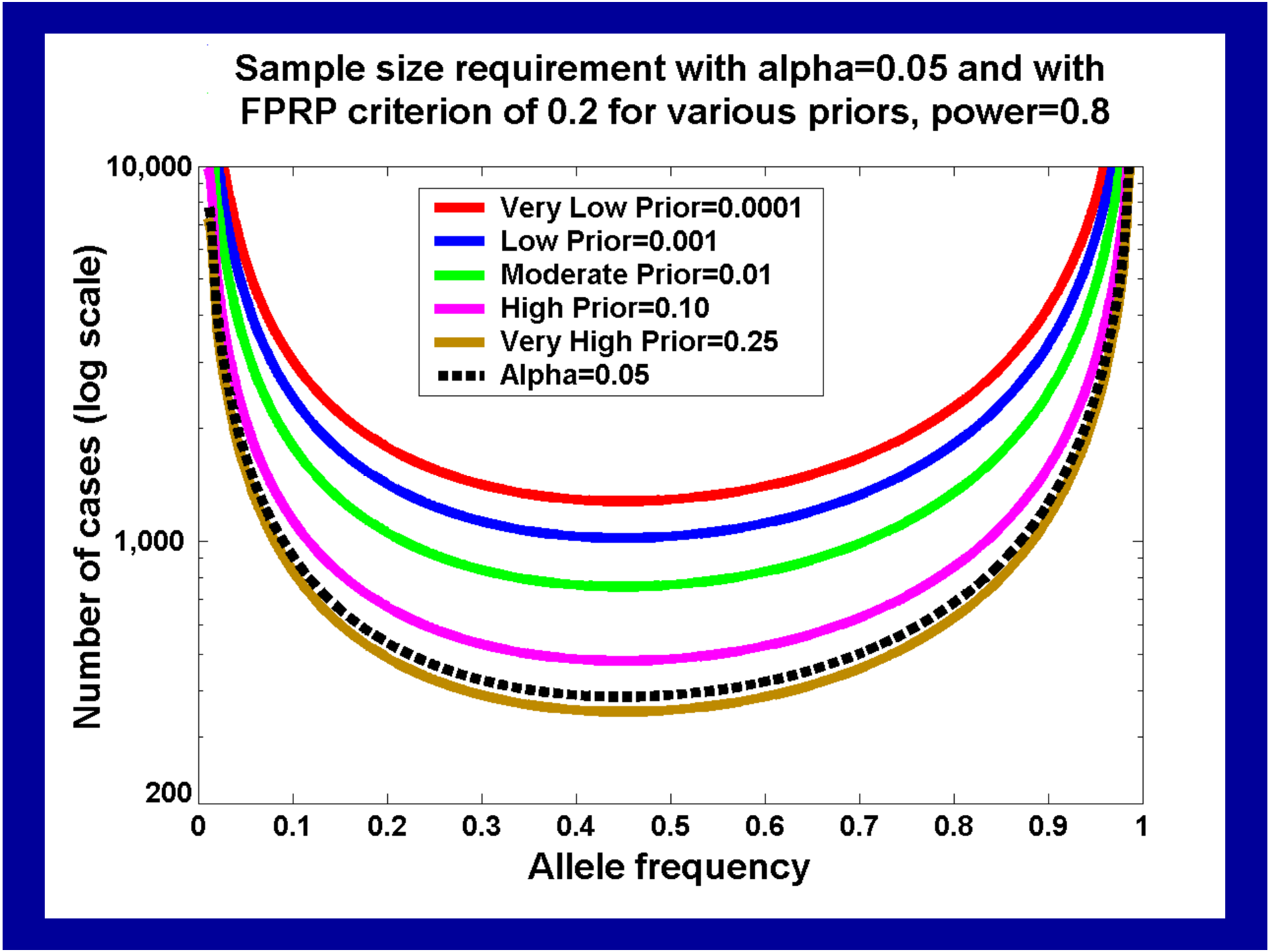
  FNRP= Pr(No association | No Rejection)

$$FNRP = \frac{\beta(\pi)}{\beta(\pi) + (1-\alpha)(1-\pi)} = \frac{1}{1 + \frac{(1-\alpha)}{\beta}\frac{(1-\pi)}{\pi}}$$


Effect of prior and power on FPRP,
alpha=0.05
False Positive Report Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
0.25
0.1
0.01
0.001
0.0001
Prior
Power=0.2
Power=0.5
Power=0.8
Power=1.0


Effects of sample size on FPRP,
q=0.3, RR=1.5, alpha=0.05
False Positive Report Probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
N= 100 -- power=0.27
N= 300 -- power=0.65
N= 500 -- power=0.86
N=1500 -- power=0.9995
High Prior
(0.25)
Moderate Prior
(0.1)
Low Prior
(0.01)
Very Low Prior
(0.001)


P-value and FPRP for two sample sizes,
prior=0.001, RR=1.5, q=0.3
Number of Cases=300
Number of Cases=1500
False Positive Report Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.00000001
0.000001
0.0001
0.01
0.05
P-value


Sample size requirement with alpha=0.05 and with FPRP criterion of 0.2 for various priors, power=0.8
Very Low Prior=0.0001
Low Prior=0.001
Moderate Prior=0.01
High Prior=0.10
Very High Prior=0.25
Alpha=0.05
Number of cases (log scale)
10,000
1,000
200
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Allele frequency

Key question

- What is optimal tradeoff between power and protection from false positives?
  - Universal 95% CI, p<0.05 equally inappropriate for low prior probabilities
  - Bonferroni provides insidious incentive
    - Don't explore additional hypotheses or subgroups
- Or tradeoff between FPRP vs FNRP
  - False negative report probability
- Tradeoff may be different for different audiences
  - Researchers
  - Public

Implication

- Vary the alpha level depending on how likely X is to cause D
  - Bayes approach
  - FPRP: 4-step program
    - Wacholder et al., JNCI, 2004
    - Simple calculation from p-value
    - Spreadsheet for reader, editor

Other sources of false positives

- Confounding
  - For detecting subtle effects
    - Especially confounding by indication in clinical epidemiology
      - HRT
  - P-values are misleading
  - Power reduced
- Poor field work
  - Low response, follow-up compliance rates
  - Poor exposure assessment
    - Lowers power
      - Raises FPRP
  - Differential

Final thoughts (1)

- Observational studies crucial in prevention and clinical epidemiology
  - To motivate trials
  - Ethics
  - Feasibility
    - Post-marketing surveillance

Final thoughts (2)

- To reduce false positives in epidemiologic studies
  - Improve study design
  - Improve study practice
  - Improve statistical approaches
    - Including
      - Explicit consideration of probability that a positive report is a false positive due to random variation
- We scientists cannot figure out how to communicate with public until we figure out what we need to communicate with each other