When is a finding in molecular epidemiology ready for prime time?

Sholom Wacholder
DCEG, NCI

ACE, Boston
September 13, 2004

Off-peak

- Speculative
- How to rule it in or out?
  - Replication
  - Sensitivity analysis
- Too soon to think about
  - intervention?
  - mechanistic interpretation?

Prime Time

- Established
- Appropriate to think about
  - intervention?
  - mechanistic interpretation?

Why pick on *molecular* epidemiology?

- Same problems exist in all epidemiology
  - In all science
  - Sam Shapiro re *pharmaco*epidemiology
    - Be skeptical of your own findings
    - “What might have gone wrong?”
  - Personal incentive for scientist:
    - oversell results to get attention
      - Resist skepticism
  - Role of sensitivity analysis
    - Only if one identifies important issues

Why pick on *molecular* epidemiology?

- Because track record is especially poor for identifying genes that cause complex disease
  - Too many false positives in “association” (*sic*) studies
    - Lohmueller
    - Hirschhorn
    - Ioannidis
  - Linkage studies don’t do so well either
    - E.g., diabetes, prostate ca

Why the poor track record?

- Because variation in even the “best-candidate genes” rarely cause meaningful elevation in risk of complex diseases
- We cannot resist the temptation:
  - “We spent $M on data collection”
  - “We have the DNA.”
  - “Why not look broadly?”
- Don’t resist the temptation

Evaluation of studies must change

- Old days
  - High prior probability for pre-specified hypotheses needed to get funding
  - Small studies
- Now
  - Large studies
  - Vague hypotheses: “genes cause disease”
    - No single gene justifies study by itself
  - But high prior that small number of 30K genes may, together, have reasonable PAR

J Natl Cancer Inst 2004;96:434–42

COMMENTARY

Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies

Sholom Wacholder, Stephen Chanock, Montserrat Garcia-Closas, Laure El ghormli, Nathaniel Rothman

EDITORIALS

Betting Odds and Genetic Associations

Duncan C. Thomas, David G. Clayton

Essential formula:

FPRP: FALSE POSITIVE-REPORT PROBABILITY

PRIOR: $\pi = \text{Pr(rejection} | \text{association)}$

POWER: $1 - \beta = \text{Pr(rejection} | \text{association)}$

SIZE: $\alpha = \text{Pr(rejection} | \text{association)}$

FALSE POSITIVE-REPORT PROBABILITY:

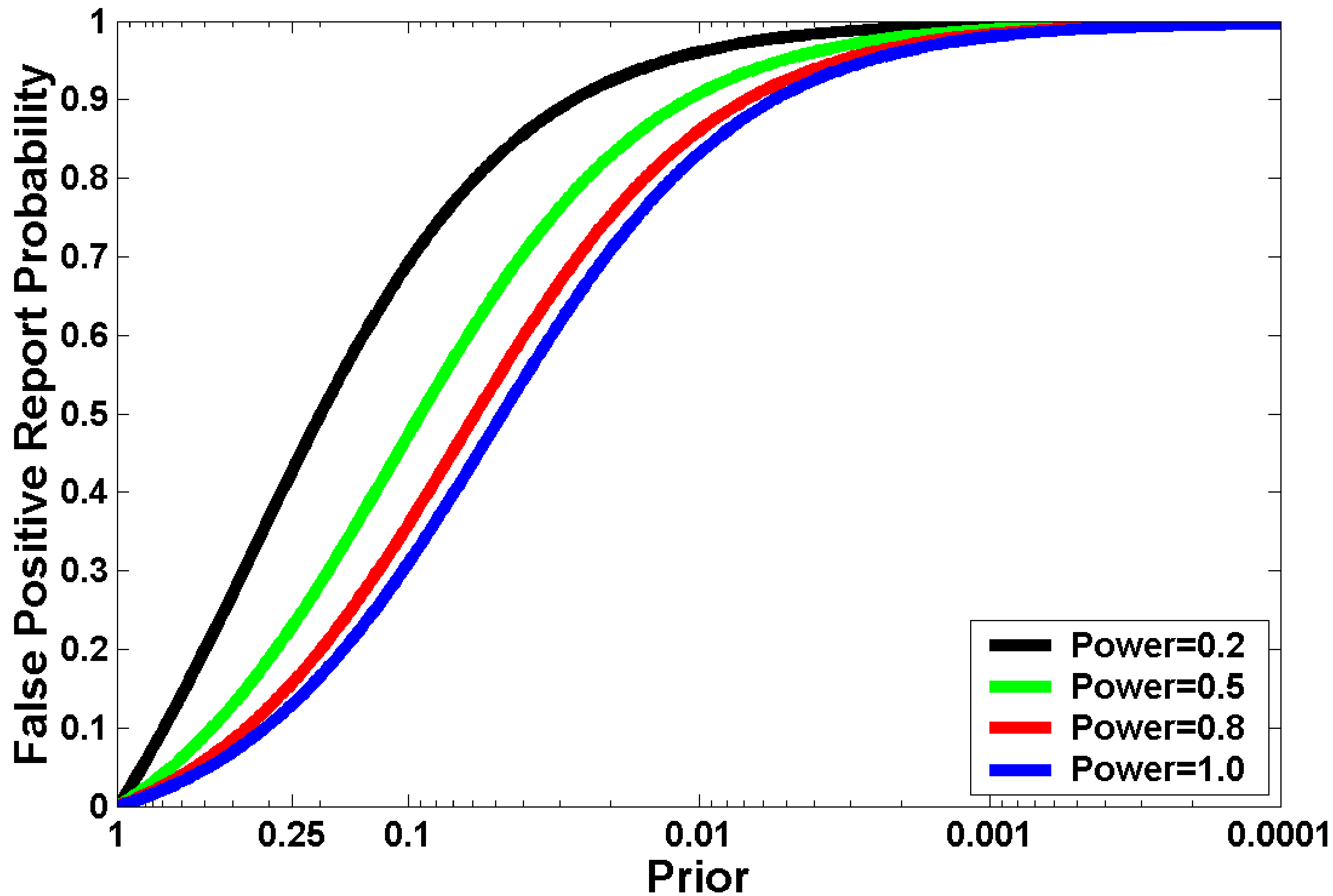
$$FPRP = \text{Pr(No association} | \text{Rejection)}$$

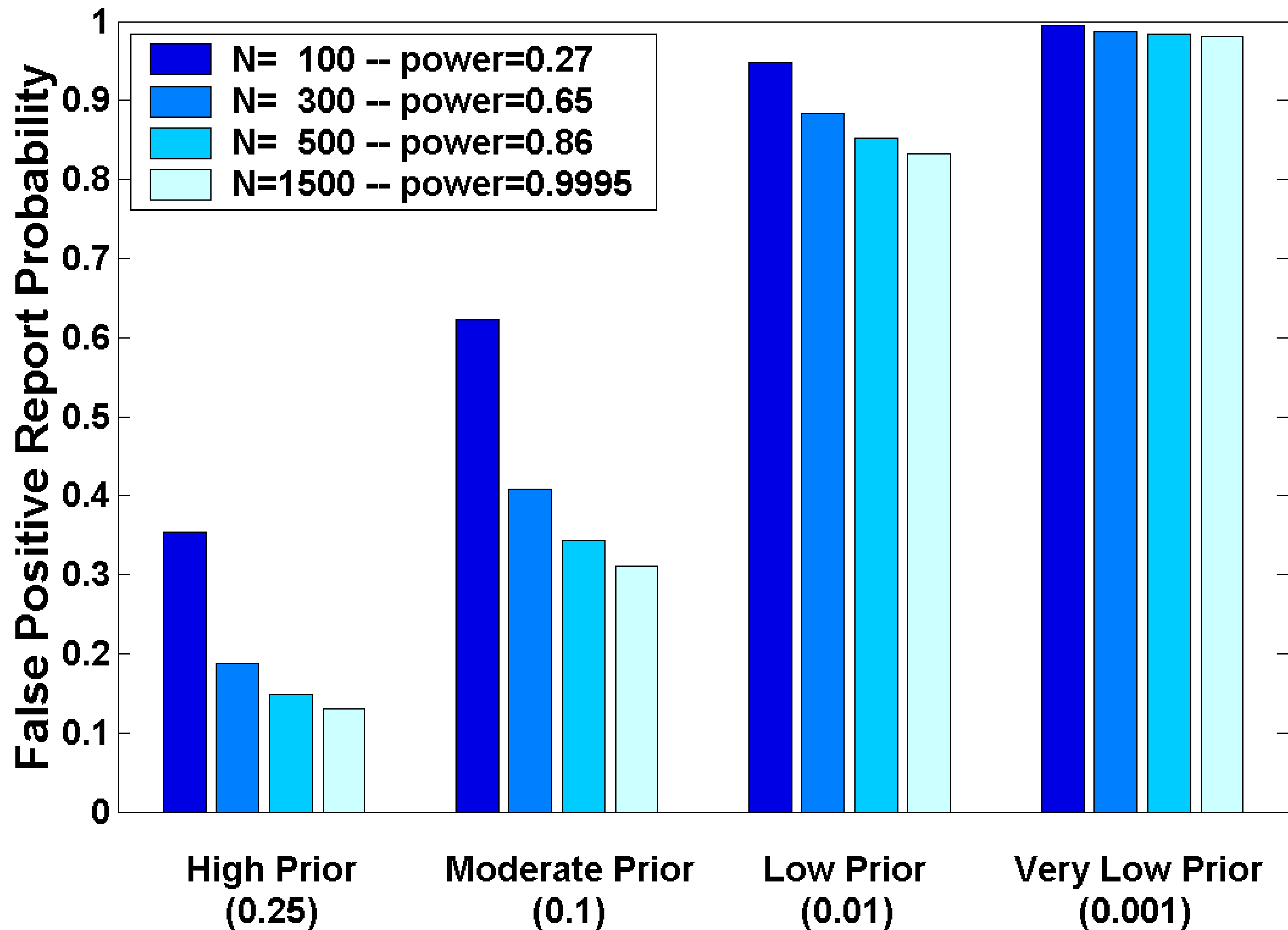
$$FPRP = \frac{\alpha(1-\pi)}{\alpha(1-\pi)+(1-\beta)\pi} = \frac{1}{1+\frac{(1-\beta)}{\alpha}\frac{\pi}{(1-\pi)}}$$

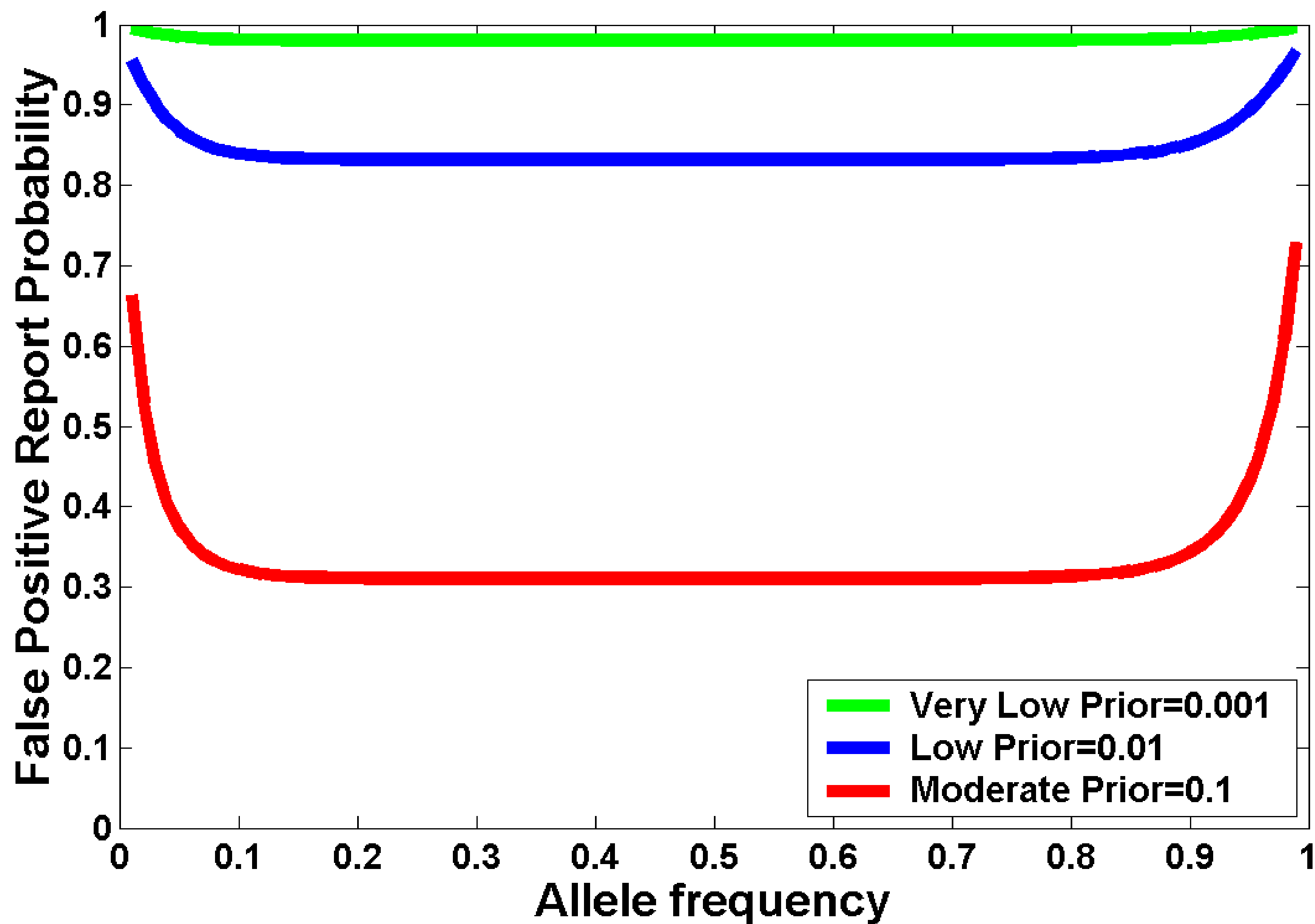
Example of algebra of false positives for speculative H_A

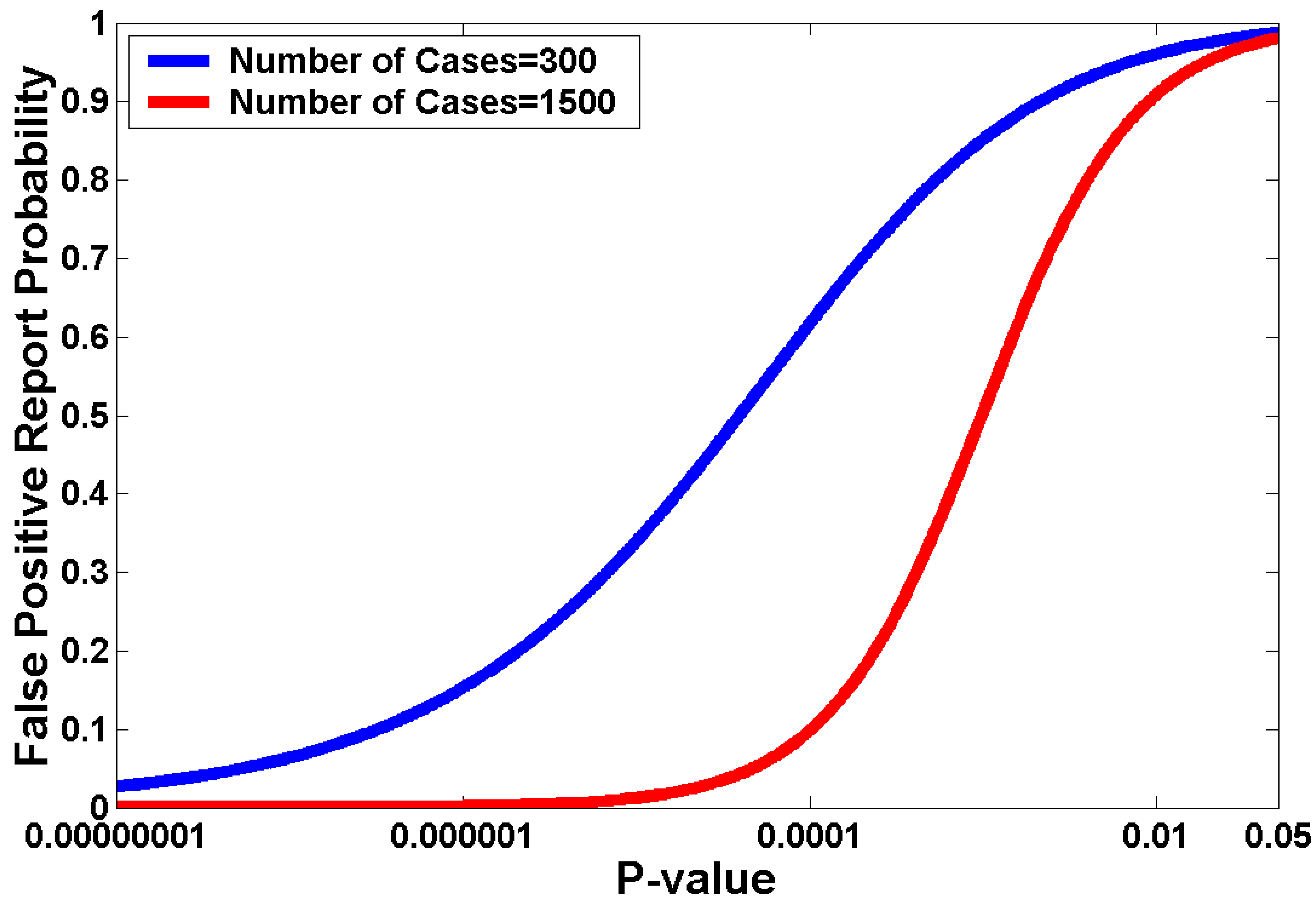
- Chance alternative hypothesis H_A is true = 0.1%=1/1,000=0.001
- If H_A false → 5% chance of rejection
- If H_A true → 100% chance of rejection
- Pr(reject & H_A false)=0.999*0.05 ≈ 0.050
- Pr(reject & H_A true) =0.001*1.00 = 0.001
- FPRP = Pr(H_A true|rejection)

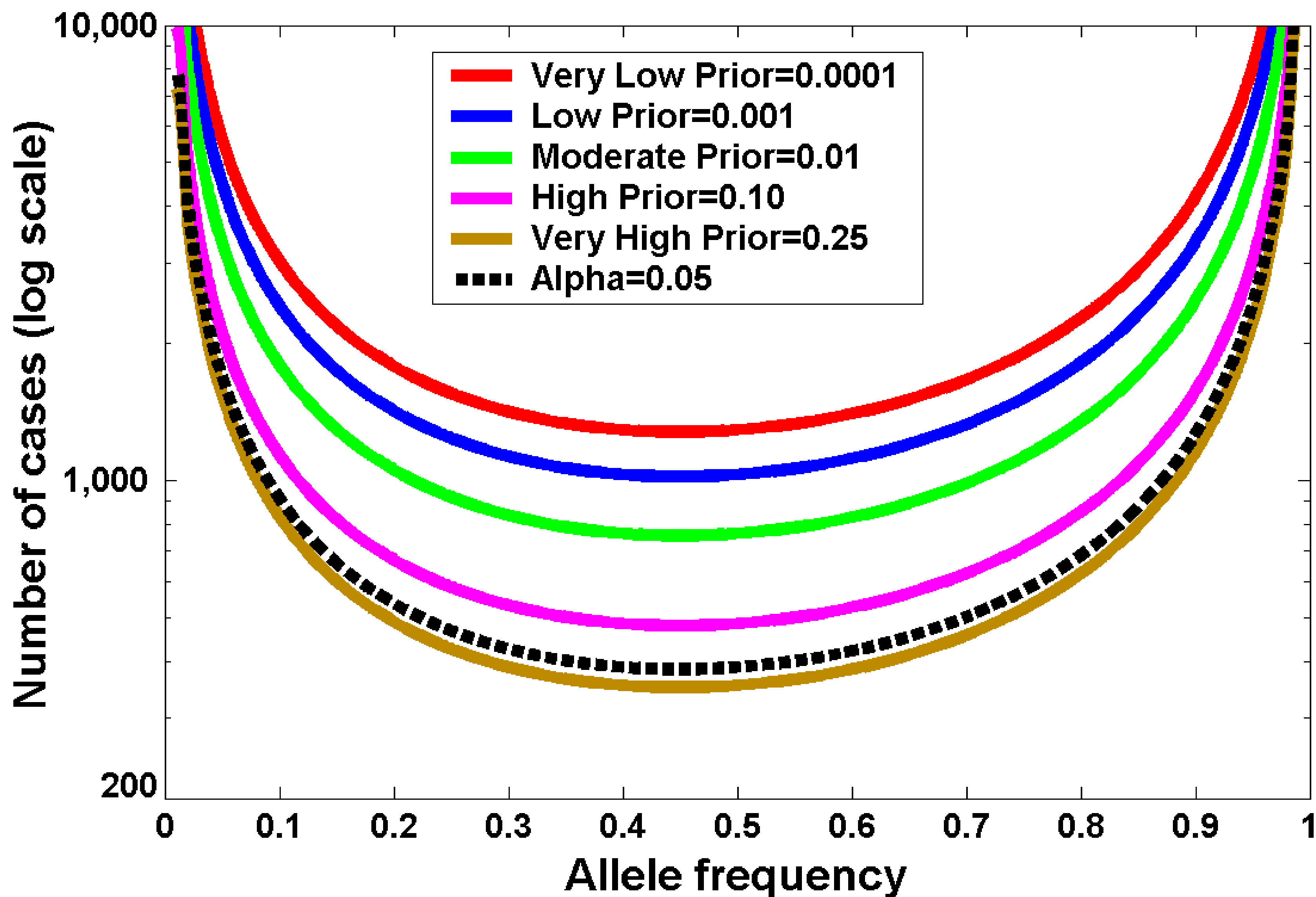
  ≈ 0.001/(0.001+0.050) ≈ 2%


Effect of prior and power on FPRP,
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
Prior
1
0.25
0.1
0.01
0.001
0.0001
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
N= 100 -- power=0.27
N= 300 -- power=0.65
N= 500 -- power=0.86
N=1500 -- power=0.9995
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
High Prior
(0.25)
Moderate Prior
(0.1)
Low Prior
(0.01)
Very Low Prior
(0.001)


FPRP as function of q for three priors,
alpha=0.05, N=1500
False Positive Report Probability
Allele frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Very Low Prior=0.001
Low Prior=0.01
Moderate Prior=0.1


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
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
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

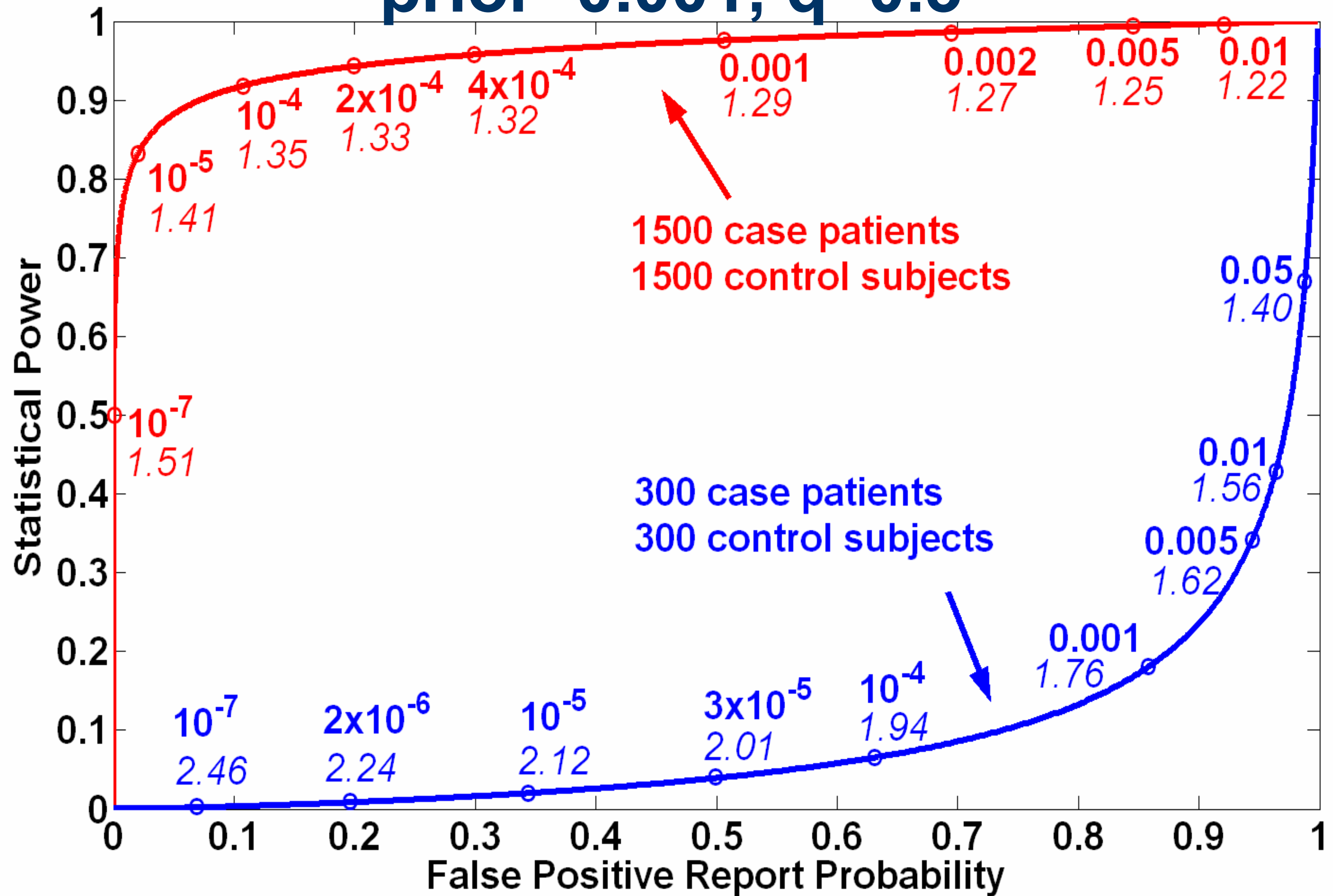
Key message

- What is optimal tradeoff beween power and protection from false positives?
  - Universal 95% CI, p<0.05 equally inappropriate for low prior probabilities
  - Bonferroni is insidious incentive
- Show FPRP-FNRP tradeoff with p-values and ORs

Implication

- Vary the alpha level depending on how likely X is to cause D
  - Bayes kind of approach
  - FPRP: 4-step program
    - JNCI, Wacholder, 2004
    - Can be done in spreadsheet by reader

FPRP vs power for OR=1.5 prior=0.001, q=0.3



Advantages

- Each hypothesis evaluated on own merit
  - Cf. Bonferroni, False discovery rate, empirical Bayes
- Explicit tradeoff between power and false positive report probability
- Investigators and readers can decide
  - prior probability
  - Investigators must consider all possible outcomes
    - E.g., “positive” findings seen in single subgroup only
      - “interaction?”
      - Random variation?

Disadvantages of FPRP

- Can be misinterpreted
  - Like p-value, CI
- Simple minded prior probability
  - Prior distribution is very hard to develop
- Uses area to right of parameter values that specify null and alternative hypothesis

Other sources of false positives

- Poor epidemiologic methods
  - Morton; Potter
- Poor epidemiologic practice

Poor epi practice: hypothetical example

- G and D=breast ca
  - OR=2 in premenopausal women
  - OR=1 in postmenopausal women
  - OR=2.5 in men
- How to integrate the evidence?

When is a finding in molecular epidemiology ready for prime time?

- When the FPRP is low for realistic low prior probabilities
- When the design is appropriate
  - Cf. other results from the same study for a clue
- When the analysis is rigorous
  - Don't change test statistic after seeing the data
    - Wrong for Bayesian or frequentist
    - Cf. FDA evaluating pharma

When is a finding in molecular epidemiology ready for prime time?

- When alternative explanations of the finding are far less likely than a real association
  - Poor design
  - Poor analysis
- When positive evidence for finding overwhelms random variation as explanation
  - How much evidence needed for “overwhelming”?
  - Rational decision must consider outside information, e.g., via prior

Final thoughts

- Be self critical
  - Randomized and observational studies
  - Qx based, molecular studies
  - Studies of genes, environment or both
- Don’t be overly cautious either
- Evaluate *all* the evidence
  - Formally and informally
  - From lab, genomics
  - Evidence for bias in epi studies
    - From other analyses of same studies
    - From sampling
- Molecular epi requires changes in design and analysis