Do We Need “Big” Or Rather Consecutive Studies In Epidemiology?

Presented (20 June 2007) at the SER, Annual Meeting, Boston, USA, 19 – 22 June 2007

Gerd Kallischnigg, Etienne Kaelin, Hans-Jörg Urban & Rolf Weitkunat

Philip Morris Products SA, PMI R&D, Epidemiology, Biostatistics and Modeling
Quai Jeanrenaud 56, 2000 Neuchâtel, Switzerland

Background

Conzin & Kaiser, Science 11 May 2007:

"Huge data sets and lower cost analytical methods are speeding up the search for DNA variations that confer an increased risk for diabetes, heart diesease, cancer, and other common ailments."

"What sets these studies apart from earlier gene discoveries…is that the new associations are statistically far more powerful and highly unlikely to be due to chance."

In the same paper it is reported that studies showing an association between genes and different diseases (macular degeneration, memory, Chron's disease) are replicated.

It will be shown that **replication is necessary** to establish an association.

What's the problem?

"Too many reports of associations between genetic variants and common cancer sites and other complex diseases are false positives"
(*Wacholder et al., 2004, p.434*)

"Most associations reported in the literature have
not been confirmed by subsequent studies.
The most likely explanation is that most initial reports
are false positives, and the most common reason for this
is simply chance (type I error), exacerbated by publication bias."
(Pharoah et al., 2004, p. 852)

"…an alarming proportion of reported associations between genetic variants and diseases are not replicated"
(Thomas & Clayton, 2004, p. 421)

"Why most published research findings are false"
(Ioannidis JPA, 2005)

What's the problem?

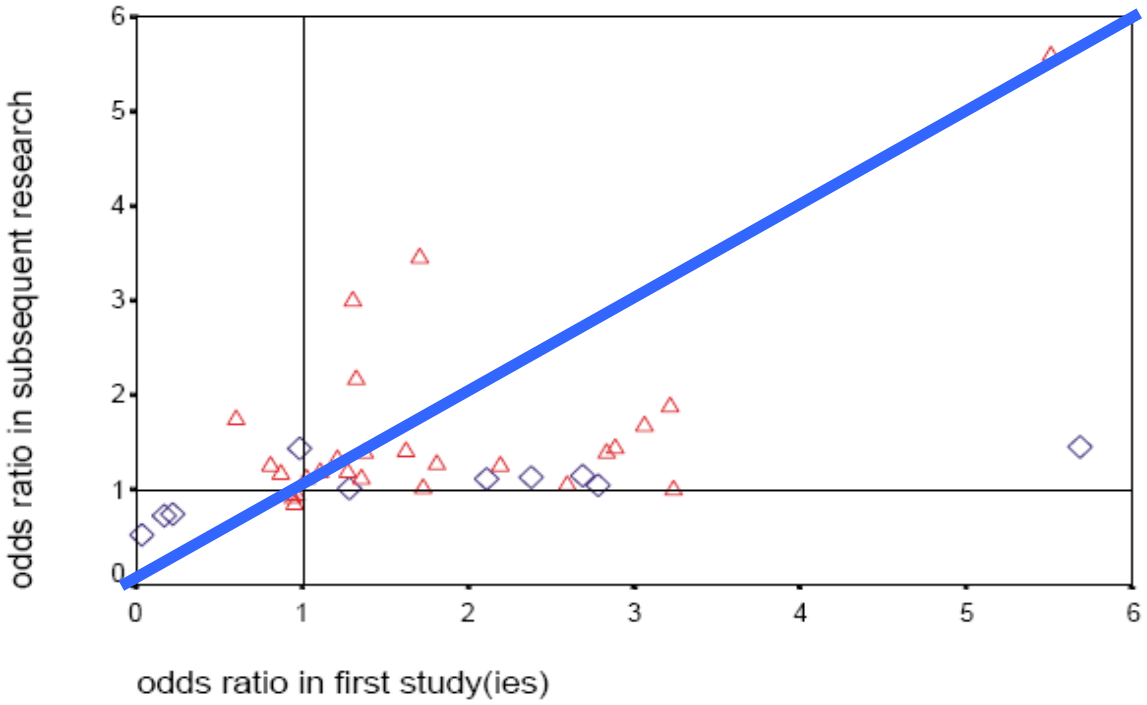


Fig.1 OR's in the first and subsequent studies. Blue diamonds denote statistically significant discrepencies between first and subsequent studies (fixed effects).

Figure from Ionnidis et al. Reprinted by permission from Macmillan Publishers Ltd: Nature genetics ,John P.A. Ioannidis, Evangelia E. Ntzani, Thomas A. Trikalinos, Despina G. Contopoulos-Ioannidis, Replication validity of genetic association studies,Copyright (2004)

False positive?

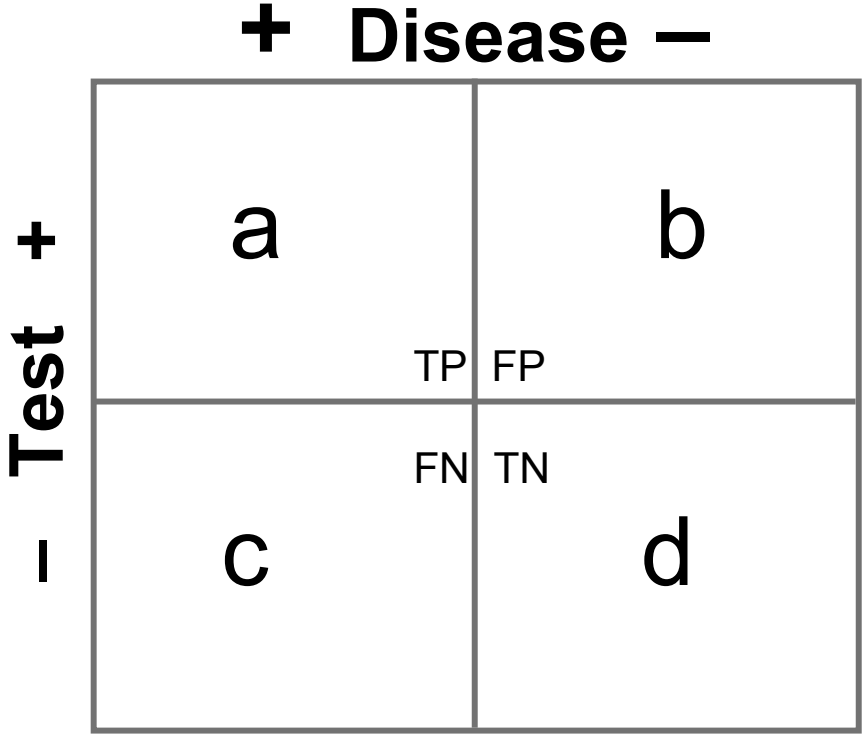
- If a statistical test identifies an association as being "significant“, it is possible that this finding is false.
- This “type I error” equals in size the selected level of significance, α (usually 5%).
- Making a type I error means that an association is identified where in reality there is none (false positive finding)

Question : How large is the true report probability (TRP)?

TRP is the probability of a true association if the report gives a positive result

TRP = Pr(T+|R+) = ?

Diagnostic test as a model for the estimation of the probability of TRP



Sensitivity (true positive rate)
Pr(T+|D+) = a / (a+c)

Specificity (true negative rate)
Pr(T-|D-) = d / (b+d)

Positive predicted value of a diagnostic test:
PPV = Pr(D+|T+) = a / (a+b)

How large is the PPV?

$$PPV = Pr(D+|T+) = \underbrace{Pr(D+)}_{\text{Prevalence}} \cdot \underbrace{Pr(T+|D+)}_{\text{Sens}} / Pr(T+)$$

Test	Disease +	Disease –
+	a (TP)	b (FP)
–	c (FN)	d (TN)

$$Pr(T+|D+) \cdot Pr(D+) + Pr(T+|D-) \cdot (1-Pr(D+))$$

$$1-Pr(T-|D-) = 1\text{-Spec}$$

$$PPV = \frac{prev \cdot sens}{prev \cdot sens + (1-prev) \cdot (1-spec)} = \frac{1}{1 + \frac{(1-spec)}{sens} \cdot \frac{(1-prev)}{prev}}$$

Example: HIV screening test

1. Test:

ELISA-Test

sensitivity = 0.95

specificity = 0.95

disease / 1. test	+	-
T+	950	9900
T-	50	989100
	1000	999000

p = 0.001

PPV = 0.088

2. Test:

Western-Blot-Test

sensitivity = 0.999 (*0.95*)

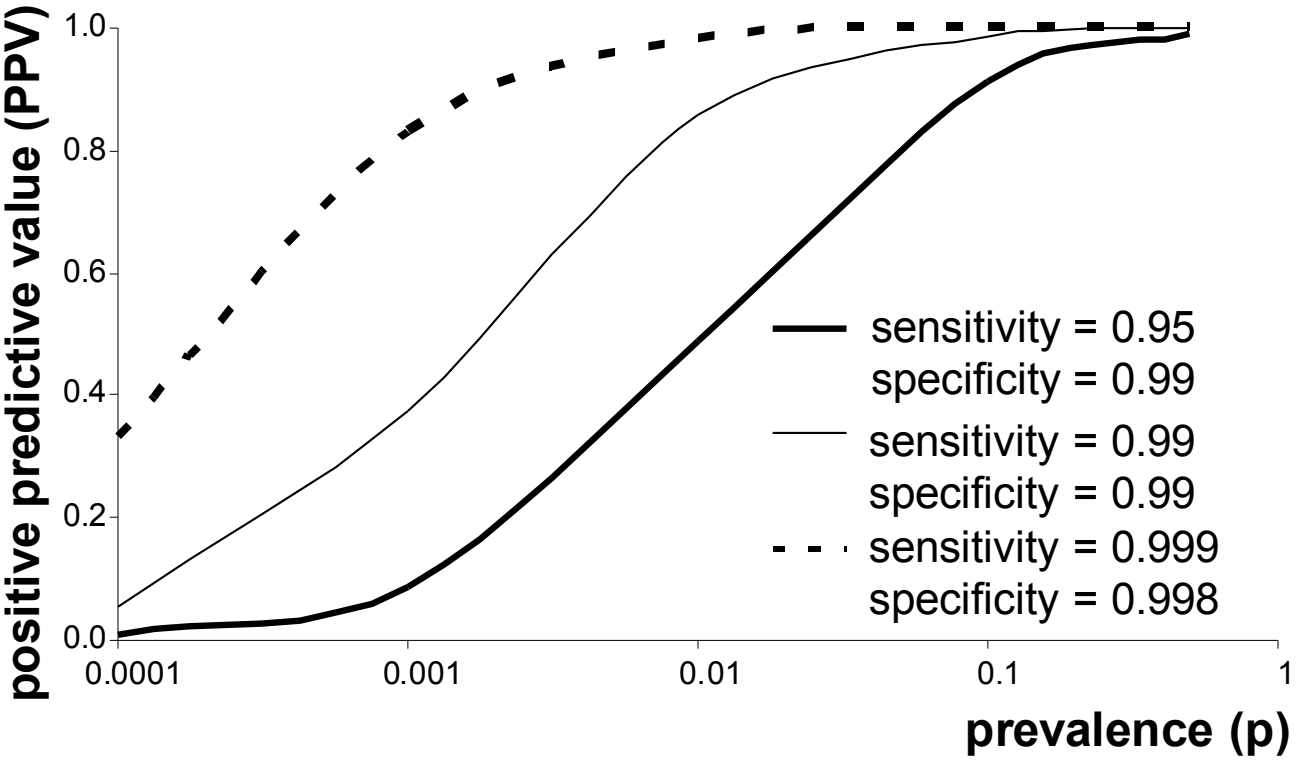
specificity = 0.999 (*0.95*)

disease / 2. test	+	-
T+	949	10
T-	1	9890
	950	9900

p = 0.088

PPV = 0.999 (*0.95*)

Effect of sensitivity and specificity on PPV



Correspondence of definitions in diagnostic and statistical tests

Diagnostic test	Statistical Test
presence of disease	association (H_1 is true)*
absence of disease	no association (H_0 is true)
positive result	reject H_0
negative result	fail to reject H_0
sensitivity	power, 1-β = 1 - type II error
1- specificity	type I error, α
prevalence of disease	prior probability of an association
positive predictive value (PPV)	true report probability (TRP)

*research hypothesis : association between a genetic variant and a disease

TRP: True Report Probability

$$TRP = Pr(\text{Association} \mid \text{Rejection of } H_0) = \frac{1}{1 + \frac{\alpha}{1-\beta} \cdot \frac{(1-p)}{p}}$$

p = prior probability of $H_1 \triangleq$ prior probability of an association

Remark:

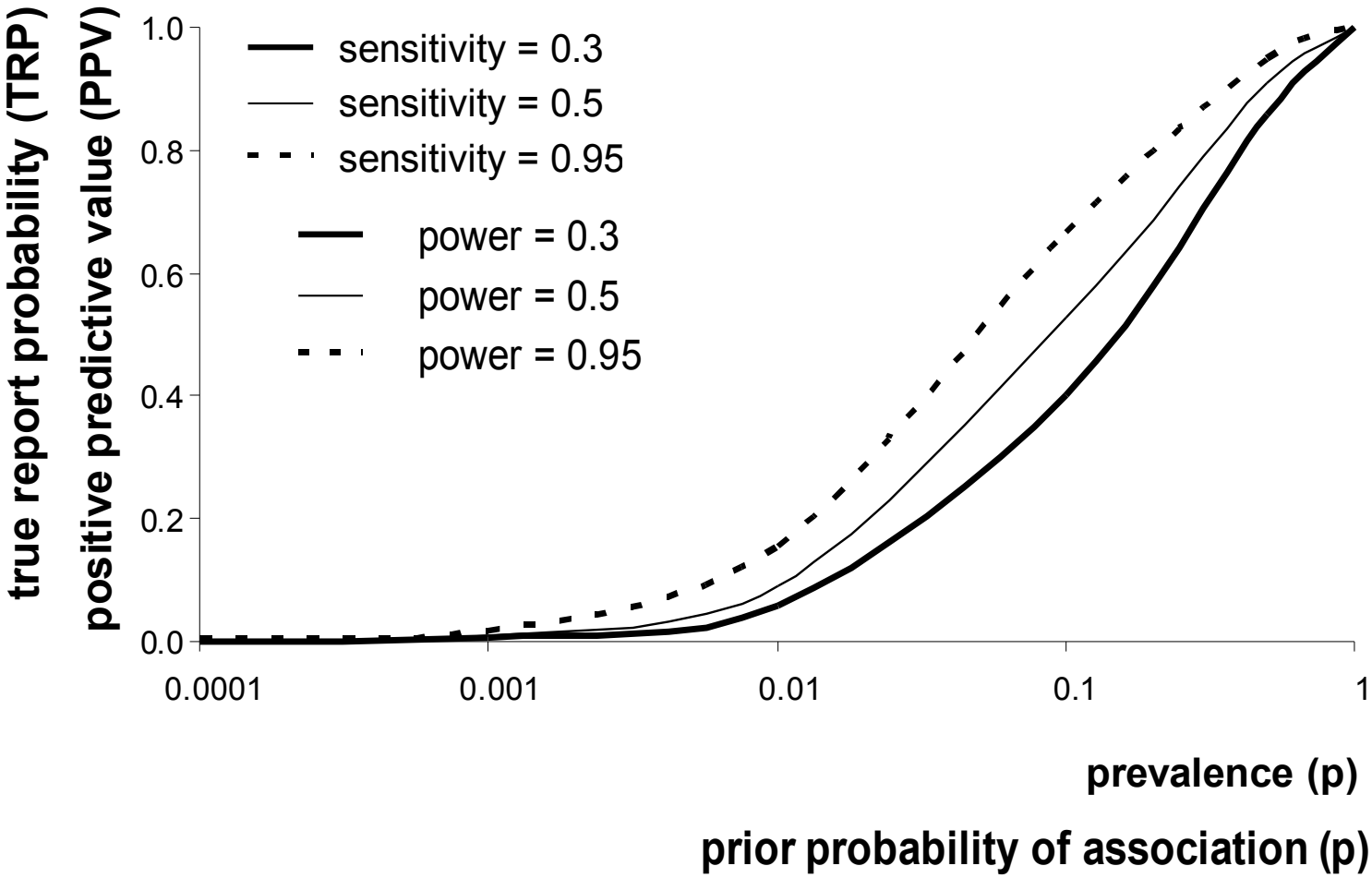
If the power is maximal (≈ 1) and $\alpha = 0.05$ then:

TRP > 0.5 if p > 0.048!

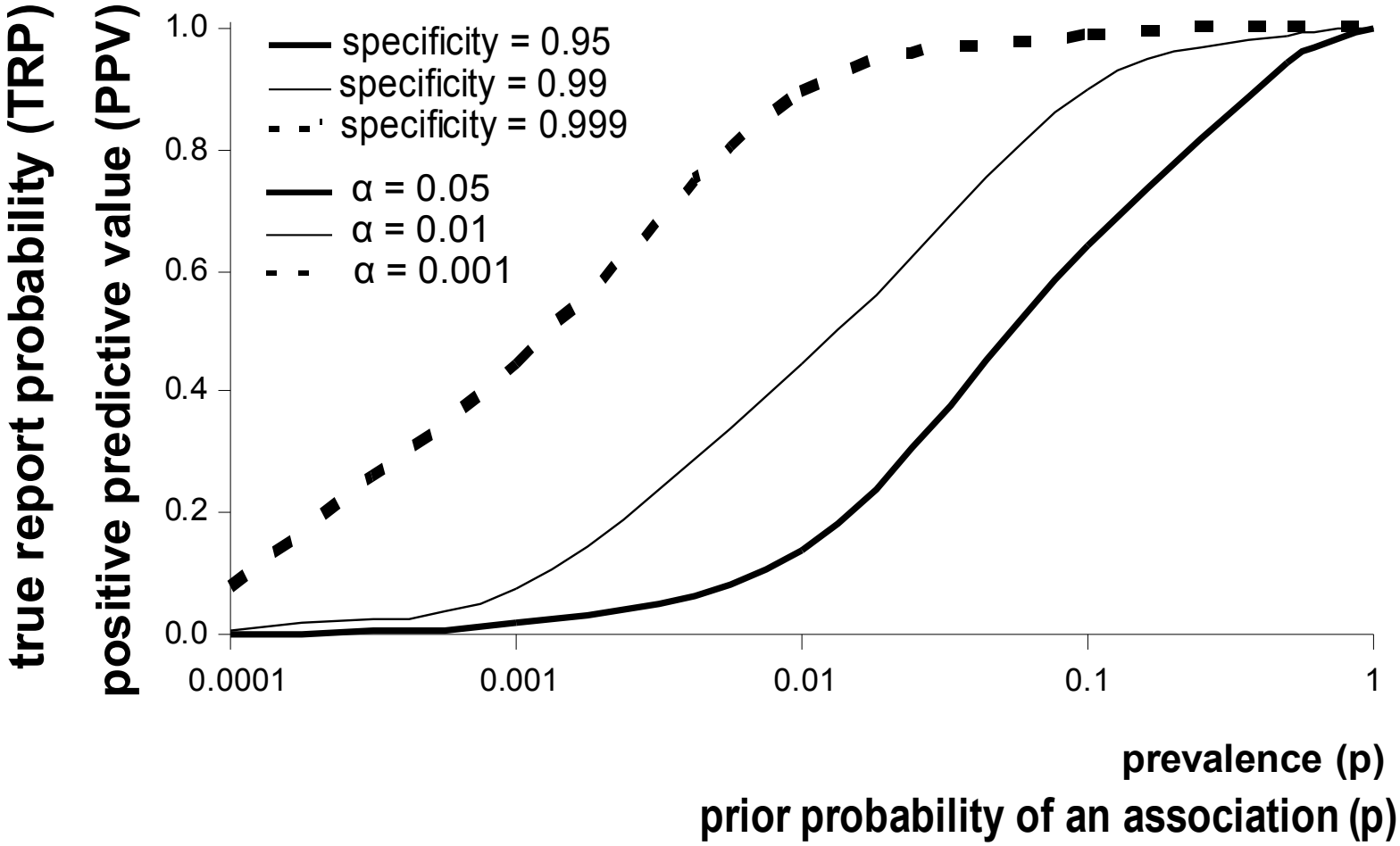
This means: it is not possible to reach a TRP > 0.5 with a positive result of one study if the prior probability of an association is less than 0.048.

- See: Wacholder et al.,2004 (using False Positive Report Probabilities)

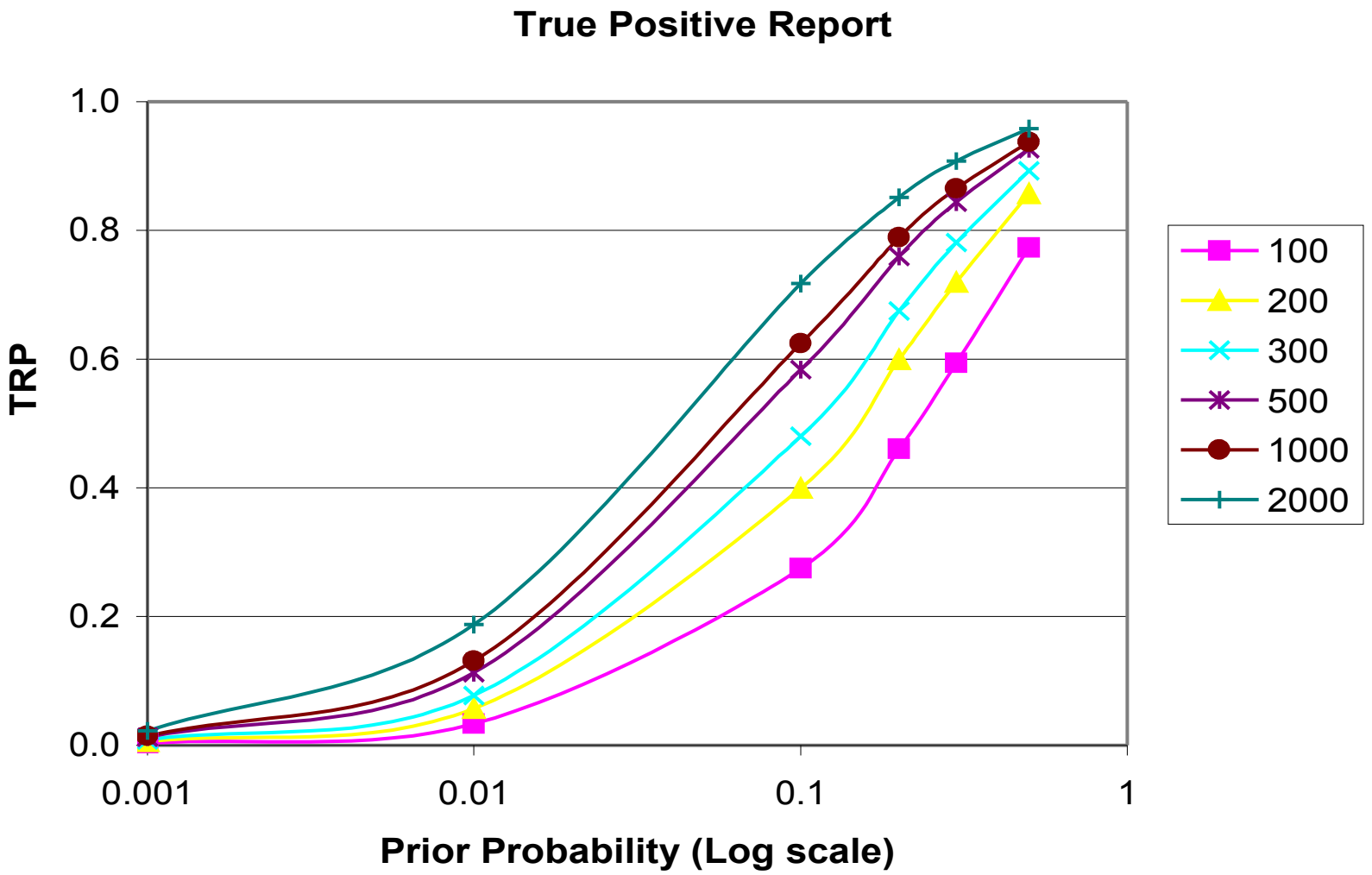
Epidemiologic Study:
Effect of Power on TRP (α = 0.05)



Epidemiologic Study:
Effect of α on TRP (1 – β = 0.8)



Effect of sample size on TRP



TRP after k replications (k+1 studies)

$$TRP(k) = \frac{(1-\beta)\,TRP(k-1)}{(1-\beta)TRP(k-1) + \alpha\,(1-TRP(k-1))}$$

$$= \frac{1}{1+\left(\frac{\alpha}{1-\beta}\right)\left(\frac{1-TRP(k-1)}{TRP(k-1)}\right)}$$

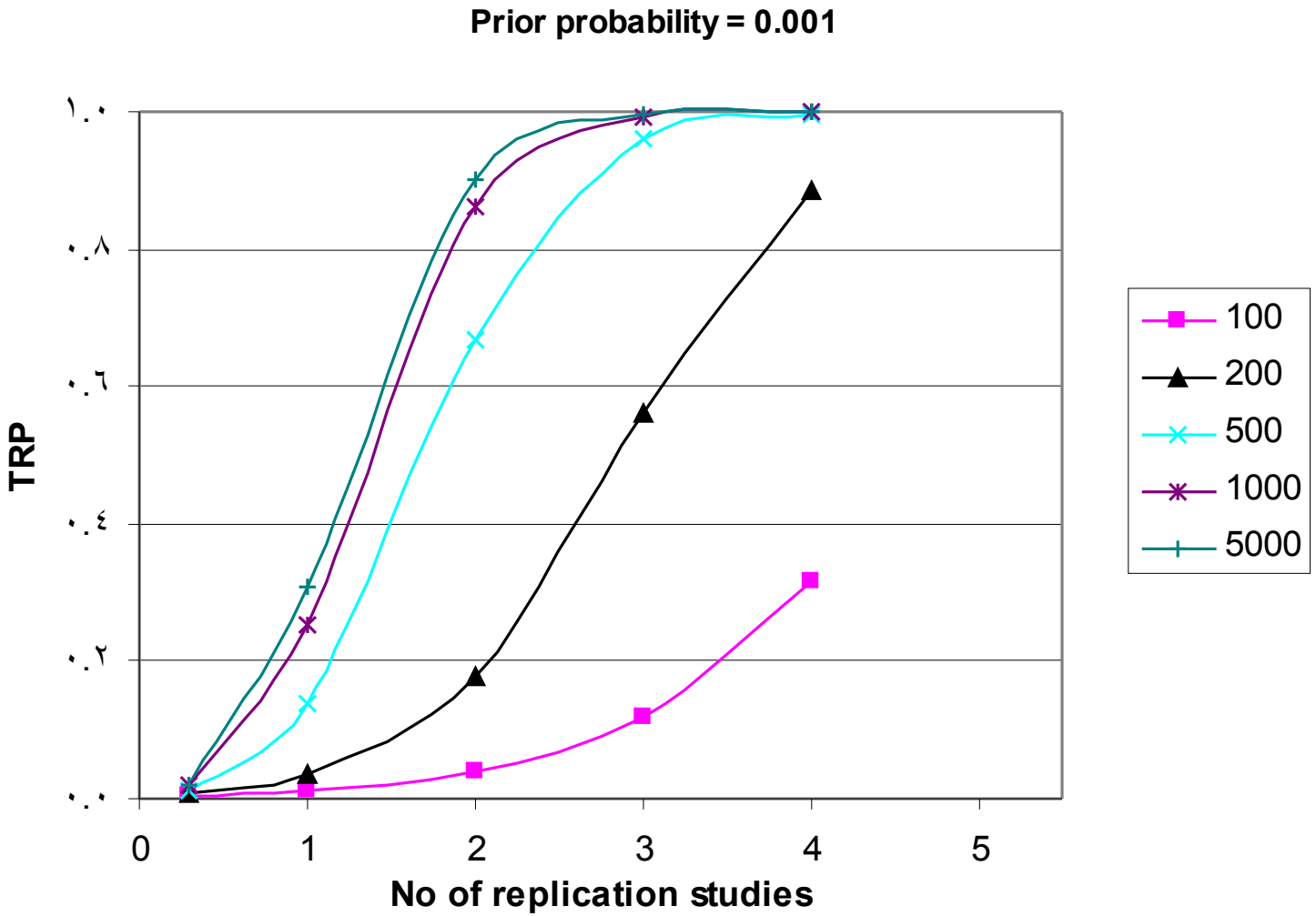
$$= \frac{1}{1+\left(\frac{\alpha}{1-\beta}\right)^{k+1}\frac{(1-p)}{p}}$$

Assumptions of the further analyses of TRP

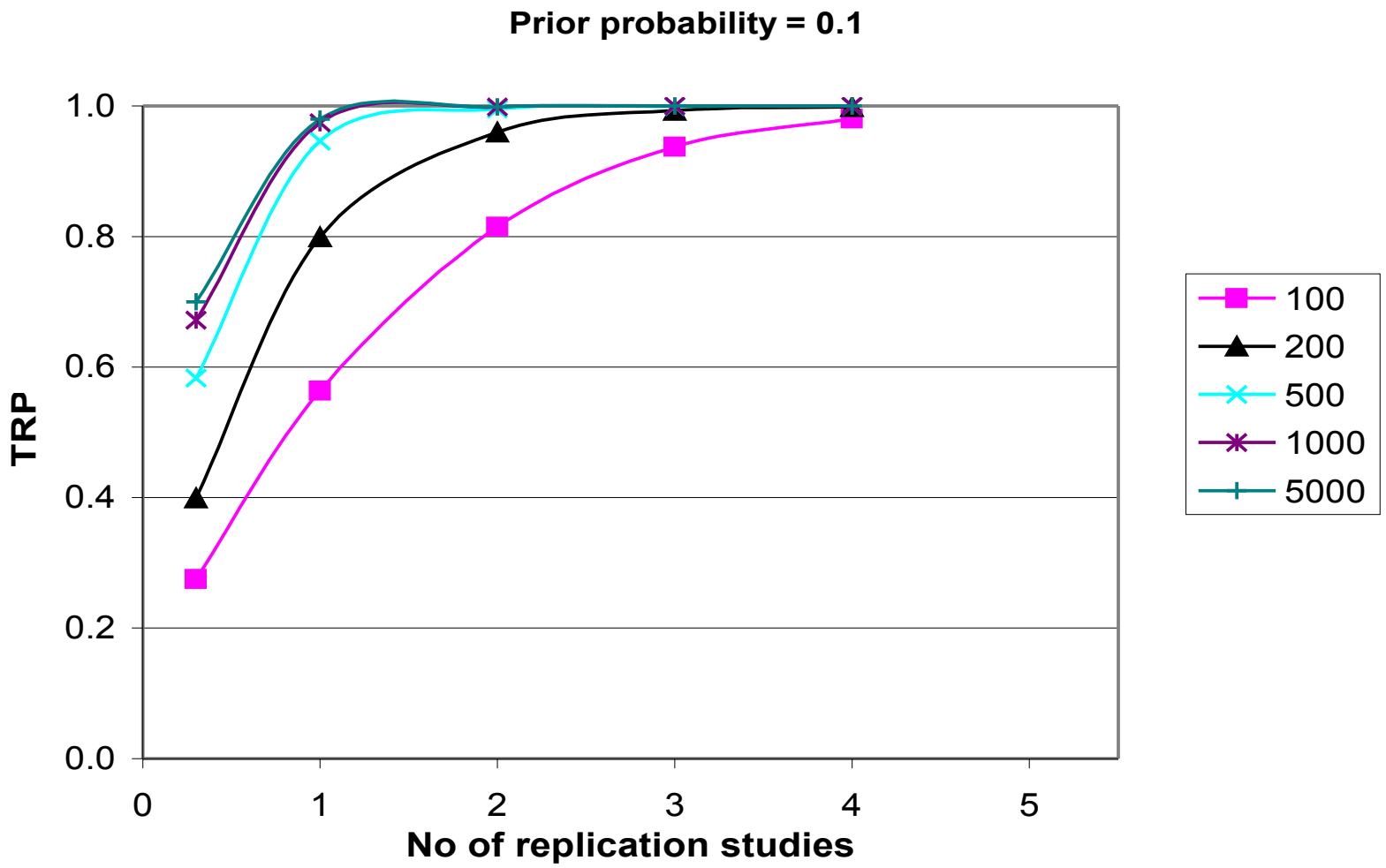
For the following analyses it is assumed that an OR ≥ 1.5 should be detected with the power of 1 – β, α =0.05.

For most calculations it is assumed that the proportion of diseased subjects with a genetic variant (true positives) is $p_a = 0.3$ (it will be shown that the effect of this assumption is relatively small).

Effect of replication on TRP



Effect of replication on TRP



Maximal asymptotic value of TRP

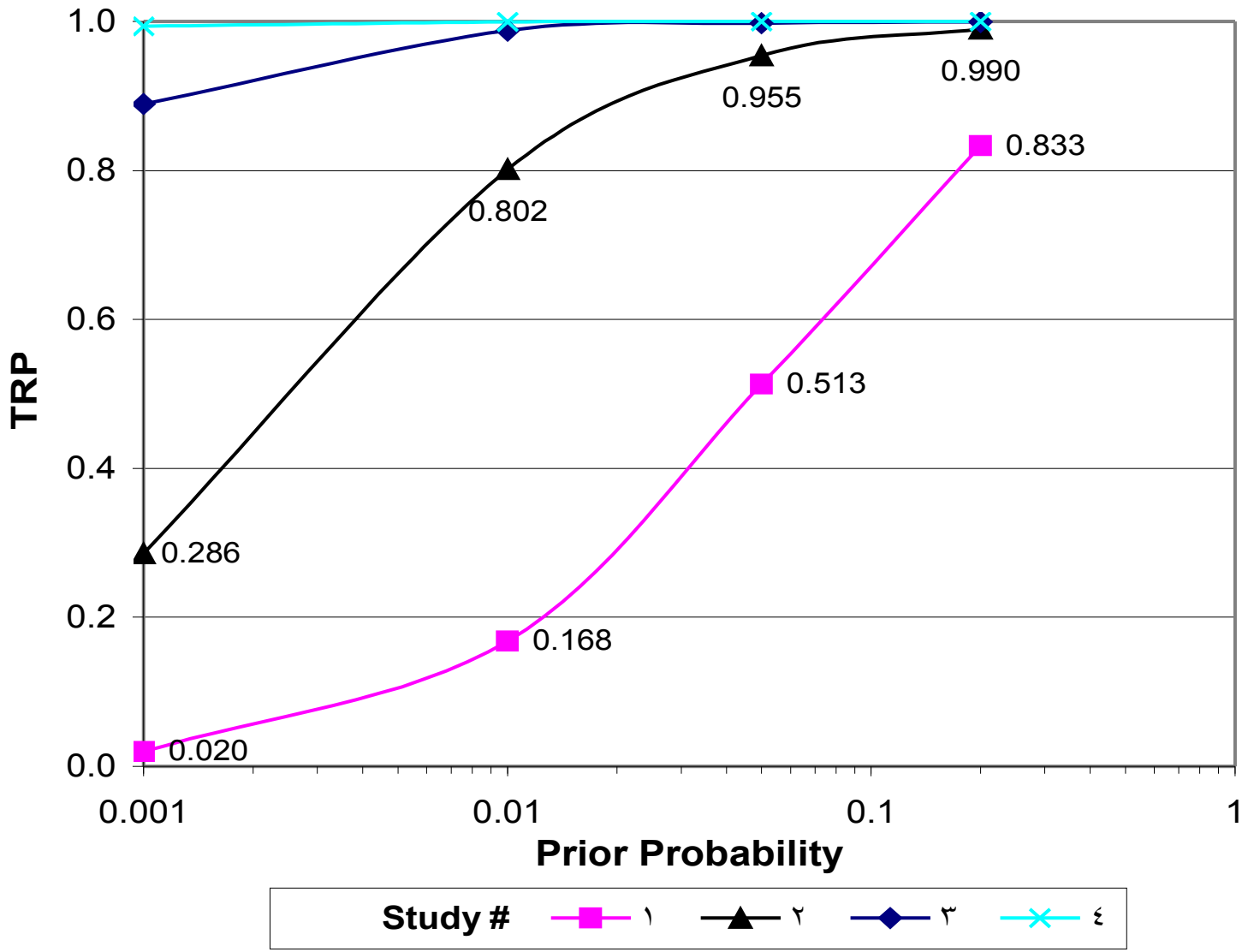
What is the maximal value of TRP?

With (1-β) ~ 1 and α = 0.05 :

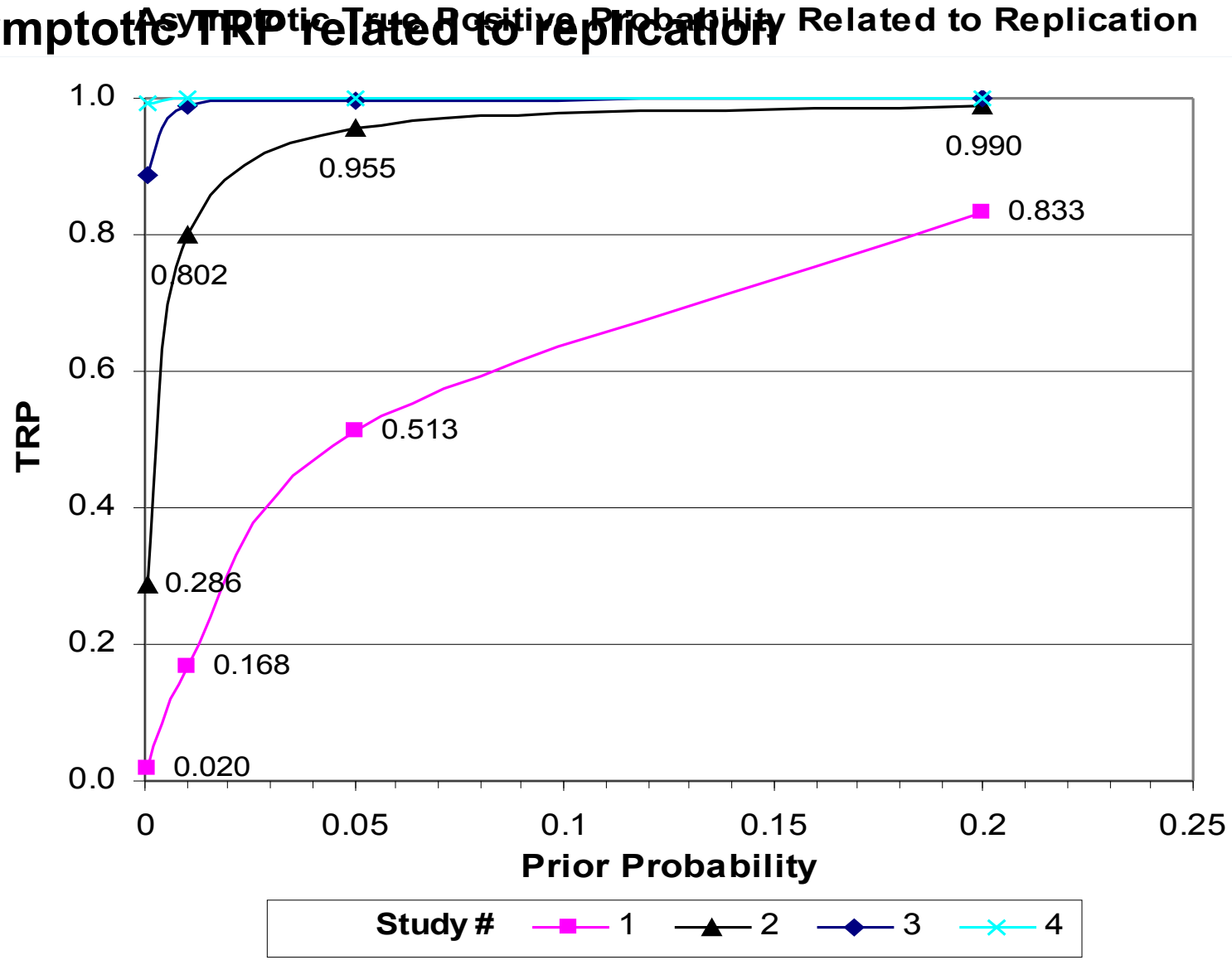
$$TRP_{max}(k) = \frac{1}{1 + 0.05^{k+1}\,\frac{(1-p)}{p}}$$

Asymptotic TRP related to replication

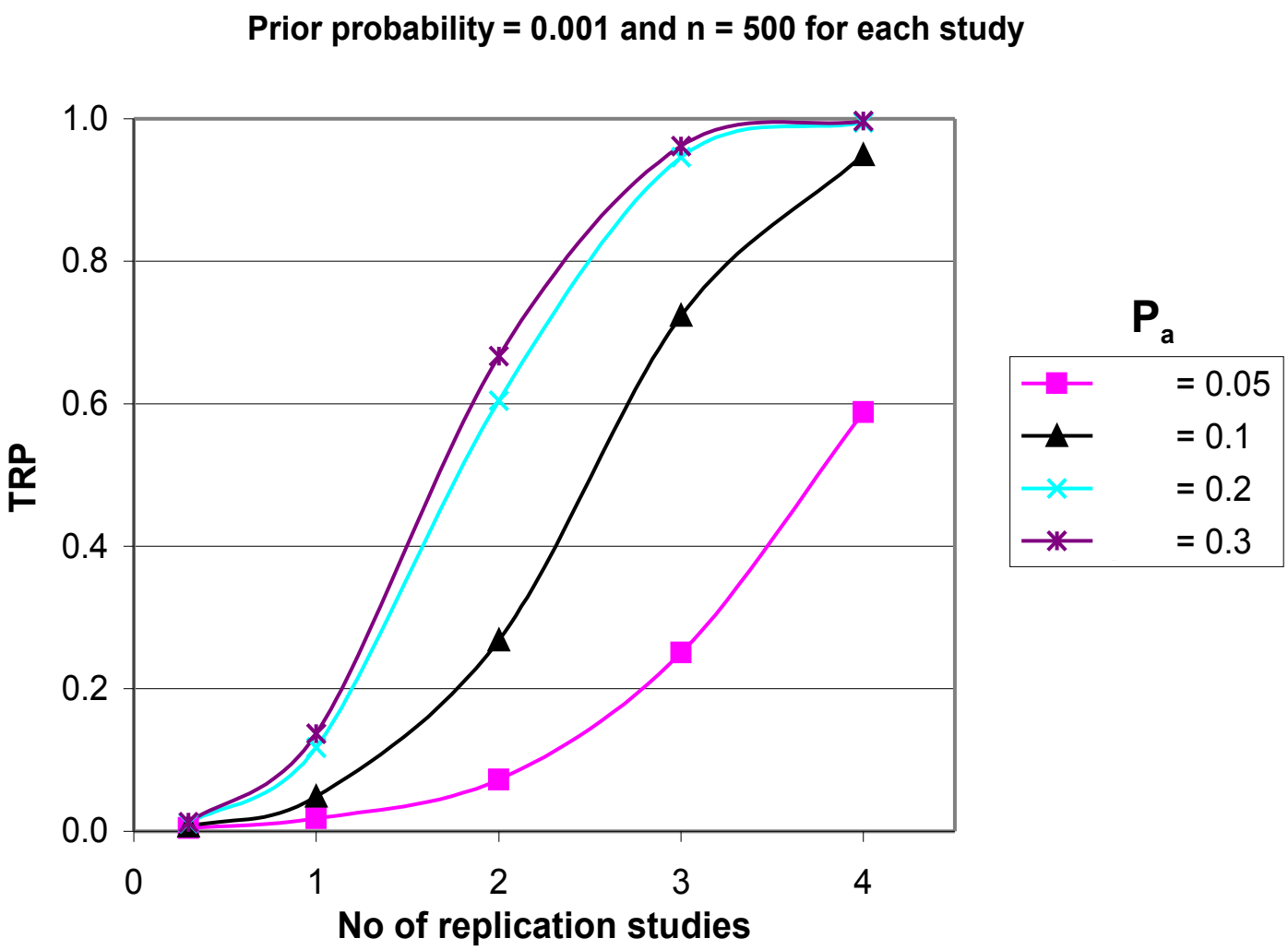
Asymptotic TRP related to replication



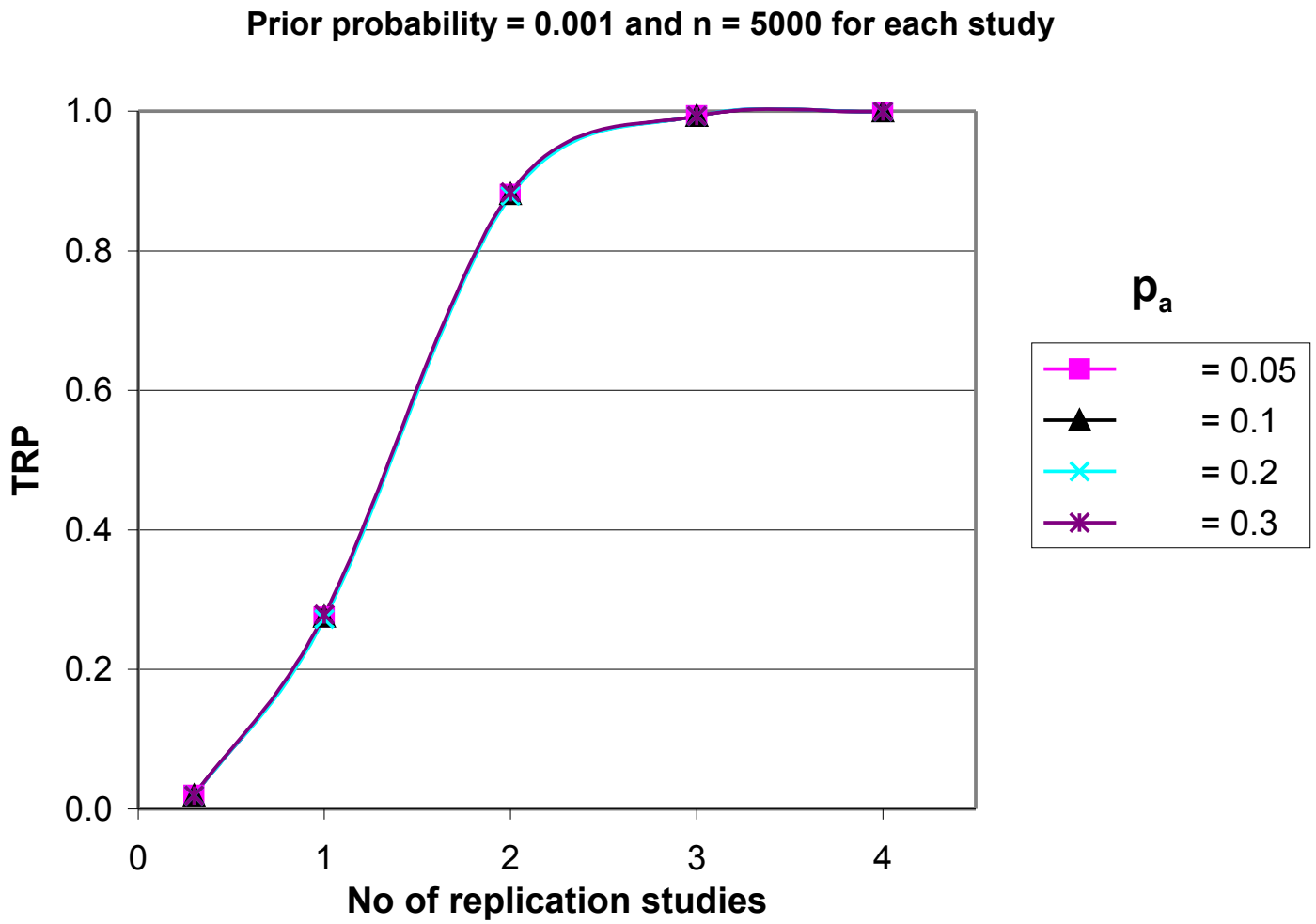
Asymptotic TRP related to replication



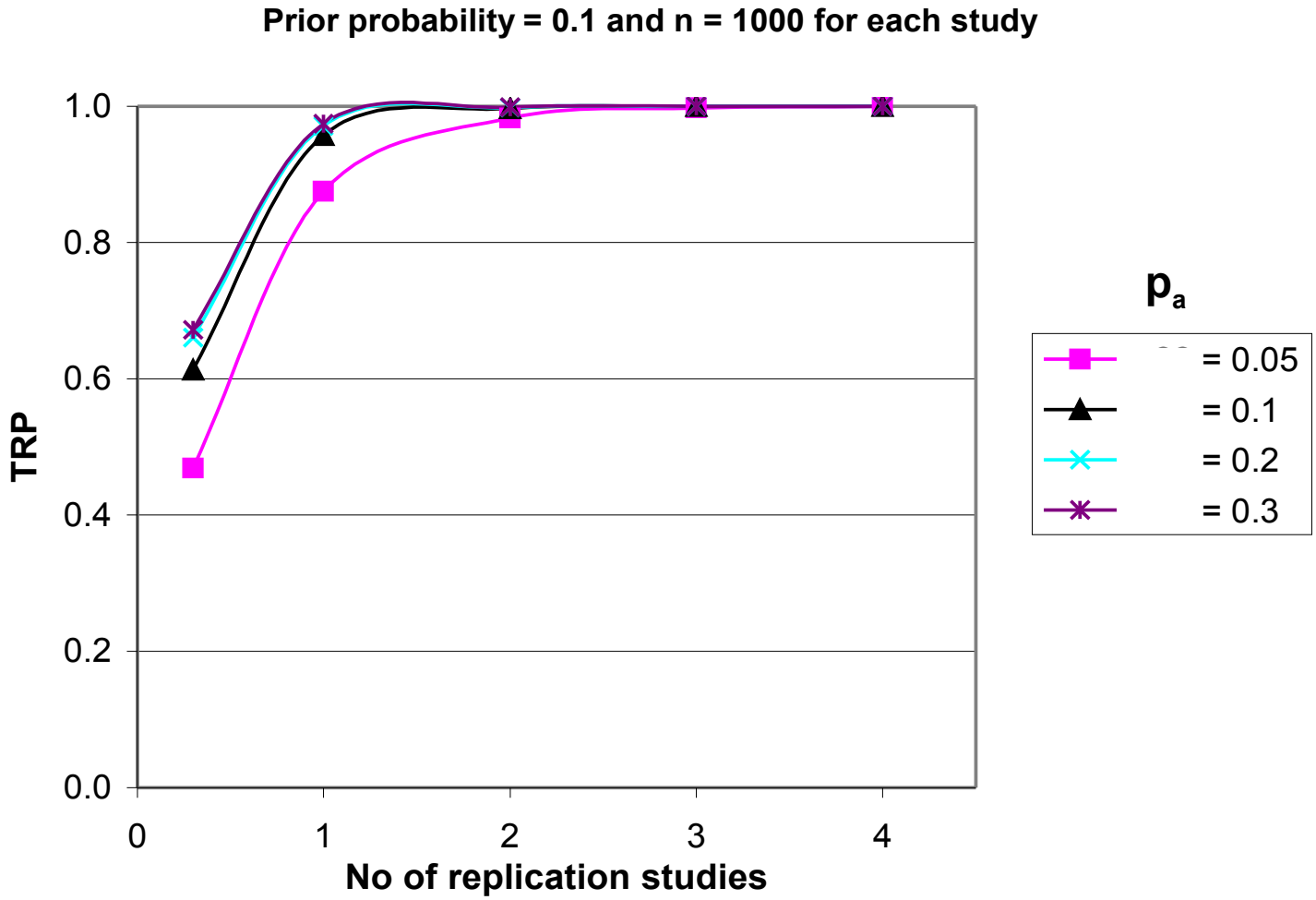
Effect of prop. of diseased subjects with genetic variants on TRP



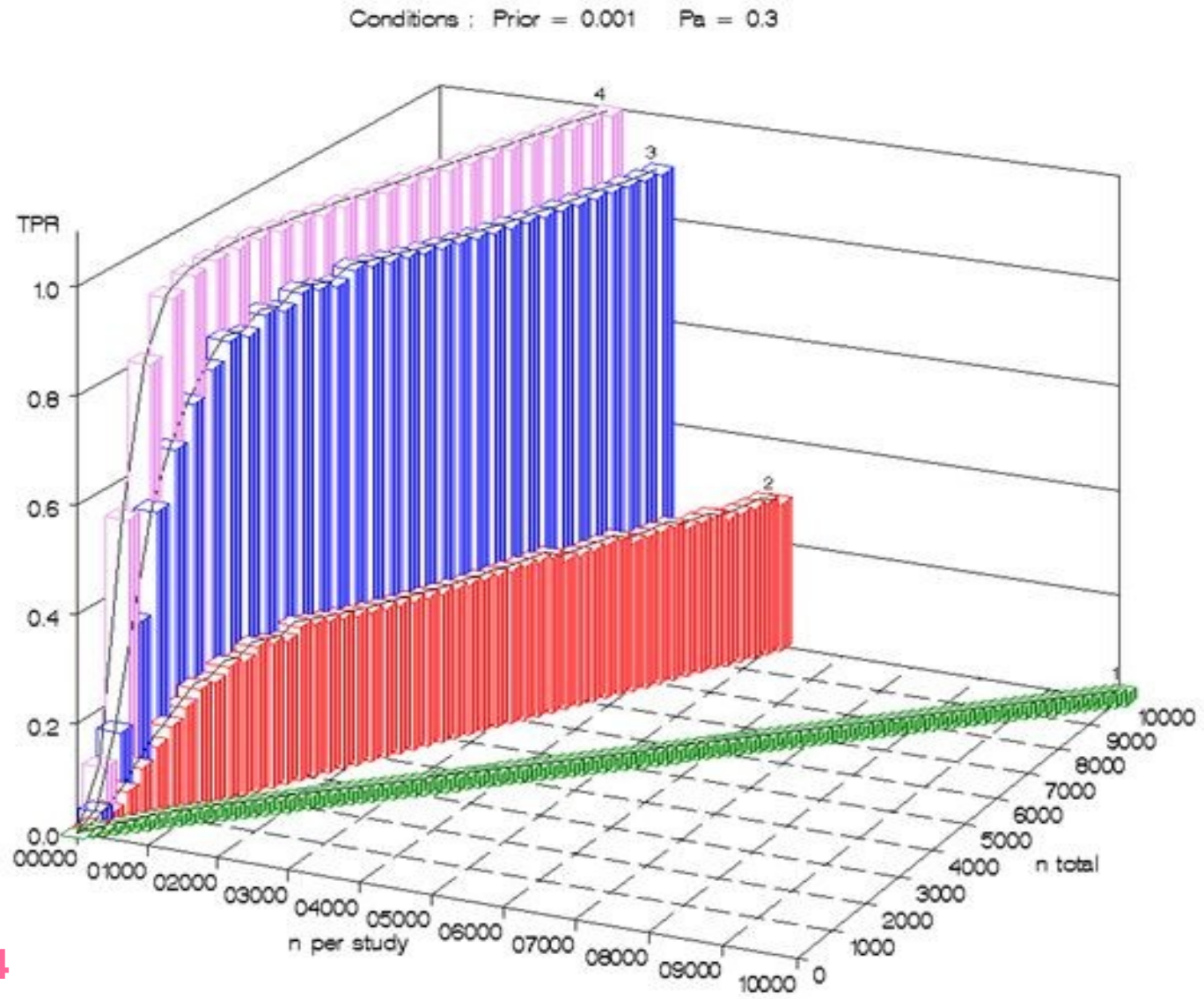
Effect of proportion of diseased subjects with genetic variants on TRP



Effect of proportion of diseased subjects with genetic variants on TRP



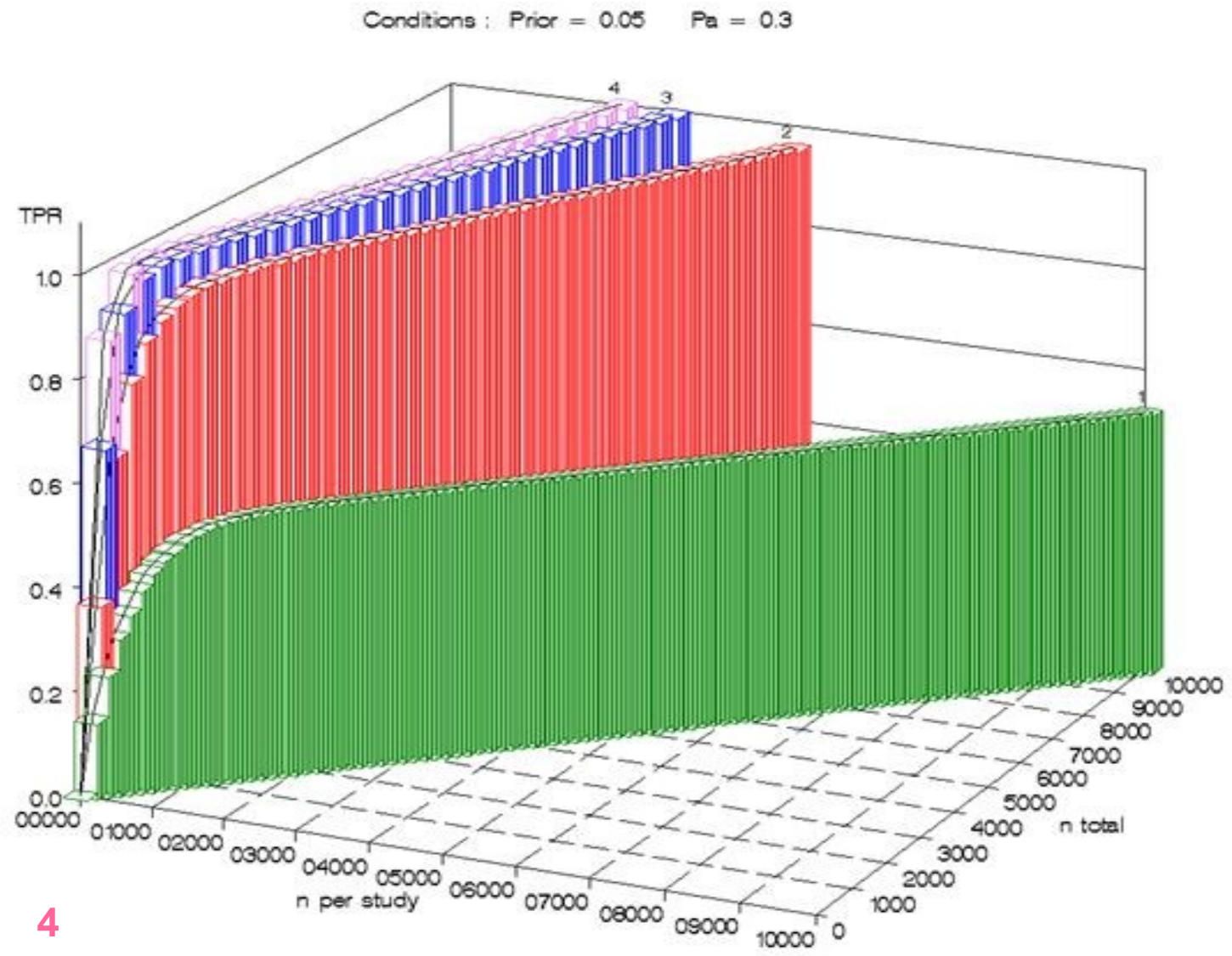
TRP after replication



Studies:

1 2 3 4

TRP after replication



Studies:

1 2 3 4

Power/sample size needed for a specified TRP?

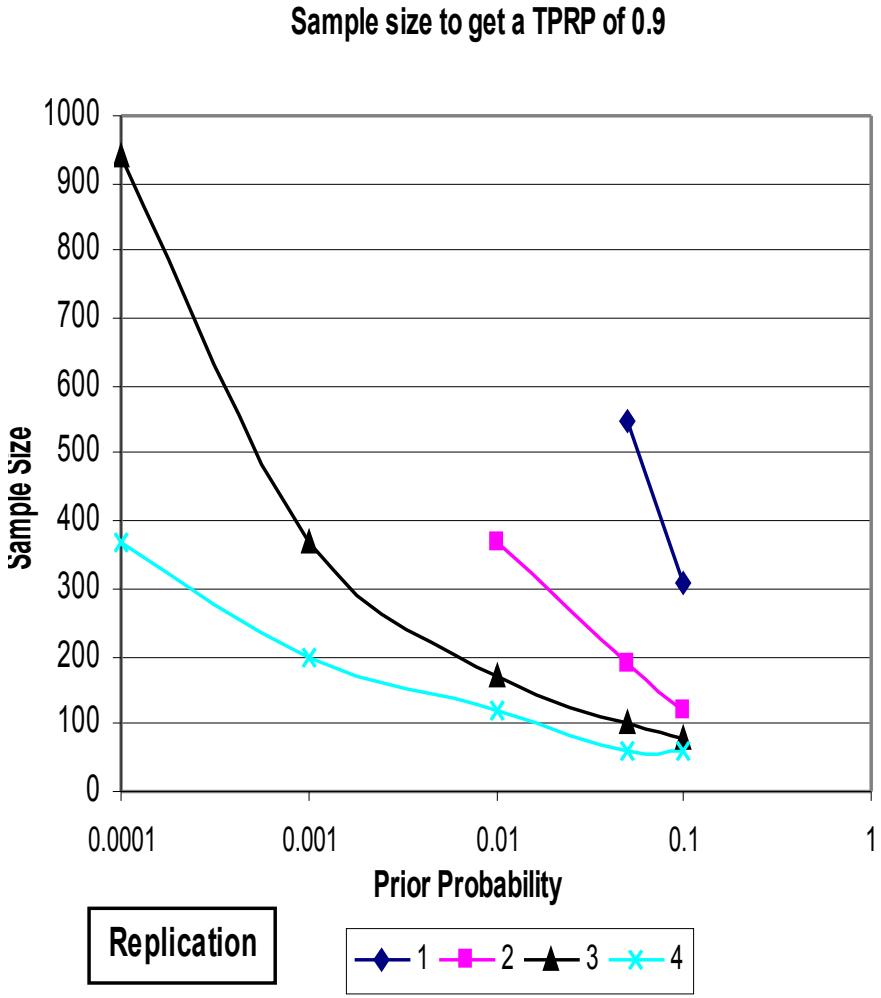
For a fixed TRP it is possible to calculate the power:

$$(1-\beta) = \left(\frac{TRP}{(1-TRP)} \cdot \frac{(1-p)}{p} \cdot \alpha^{k+1}\right)^{\frac{1}{k+1}}$$

The necessary sample size per study can be calculated according to the power.

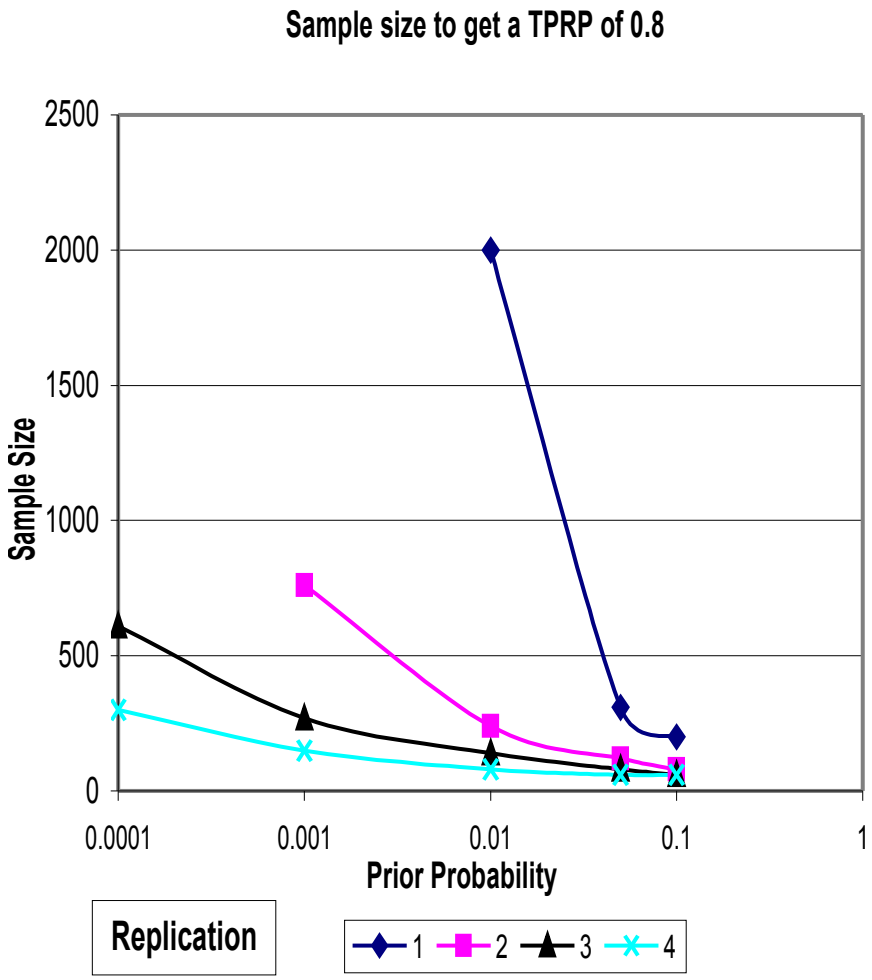
Sample size of each study to get a TRP of 0.9

Replic. p	1	2	3	4
0.0001	---	---	940	370
0.001	--	--	370	200
0.01	---	370	170	120
0.05	550	190	100	60
0.1	310	120	80	60



Sample size to get a TRP of 0.8

Replic. p	1	2	3	4
0.0001	---	---	610	300
0.001	--	760	270	150
0.01	2000	240	140	80
0.05	310	120	80	60
0.1	200	80	60	60



Conclusions

- It is important to have sufficient power to detect the effects of interest
- Increasing study sample size beyond that level does not do much to reduce false positive rates
- Replication, in contrast, is very efficient to sort-out false-positive findings*
- Replications should be part of the planning of a project, especially if the knowledge about a possible association is low
- Should new positive findings (of genome-wide association studies) be published only after replication?

  Readers should not believe or disbelieve the research hypothesis/association of a study alone on the basis of whether the results were statistically significant. It also depends on the power, the p-value, and the prior probability of the research hypothesis

* Assuming random rather than systematic error being responsible for false positive findings

References

Browner WS, Newman TB (1987). Are all significant p values created equal? JAMA,257,2459-2463.

Colhoun HM, McKeigue PM, Smith GD (2003). Problems of reporting genetic associations with complex outcomes. The Lancet, 361, 865-872.

Couzin J, Kaiser J (2007). Closing the Net on Common Disease Genes. Science, 316, 820-822.

Ioannidis JPA, Trikalinos TA, & Khoury MJ (2006). Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. American Journal of Epidemiology, 164, 609-614.

Ioannidis JPA, Ntzani EE, Trikalinos TA, & Contopoulos-Ioannidis DG (2001). Replication validity of genetic association studies. Nature genetics, 29, 306-309.

Ioannidis JPA (2005). Why most published research findings are false. PLoS Med, 2(8): e124.

Ioannidis JPA (2005). Commentary: Grading the credibility of molecular evidence for complex diseases. International Journal of Epidemiology, 35, 572-573.

Herrington D (2007). Eliminating the Improbable: Sherlock Holmes and Standards of Evidence in the Genomic Age. Circulation, 112, 2081-2084

Hoover RN (2007). The evolution of epidemiological research. From cottage industry to „big“ science. Epidemiology, 18, 13-17

Horn M & Vollandt R (1995). Multiple Tests und Auswertungsverfahren. Stuttgart: Fischer.

Kaplan GA (2007). How big is big enough for epidemiology?. Epidemiology, 18, 18-20.

Khoury MJ, Little J, Gwinn M, Ioannidis JPA (2006). On the synthesis and interpretation of consistent but weak gene-disease associations in the area of genome-wide association stidies. International Journal of Epidemiology, doi:10.1093/ije/dyl253

Lehmacher W (1987). Verlaufskurven und Crossover. Berlin: Springer.

Pharoah PDP, Dunning AM, Ponder BAJ, & Easton DF (2004). Association studies for finding cancer-susceptivility genetic variants. Nature Reviews, 4, 850-860.

Thomas DC & Clayton DG (2004). Betting odds and genetic associations. Journal of the National Cancer Institute, 96, 421-423.

Wacholder S, Chanock S, Garcia-Closas M, El ghormli L, & Rothman N (2004). Assessing the probability that a positive report is false: An approach for molecular epidemiology studies. Journal of the National Cancer Institute, 96, 434-443.

Thank you for your attention