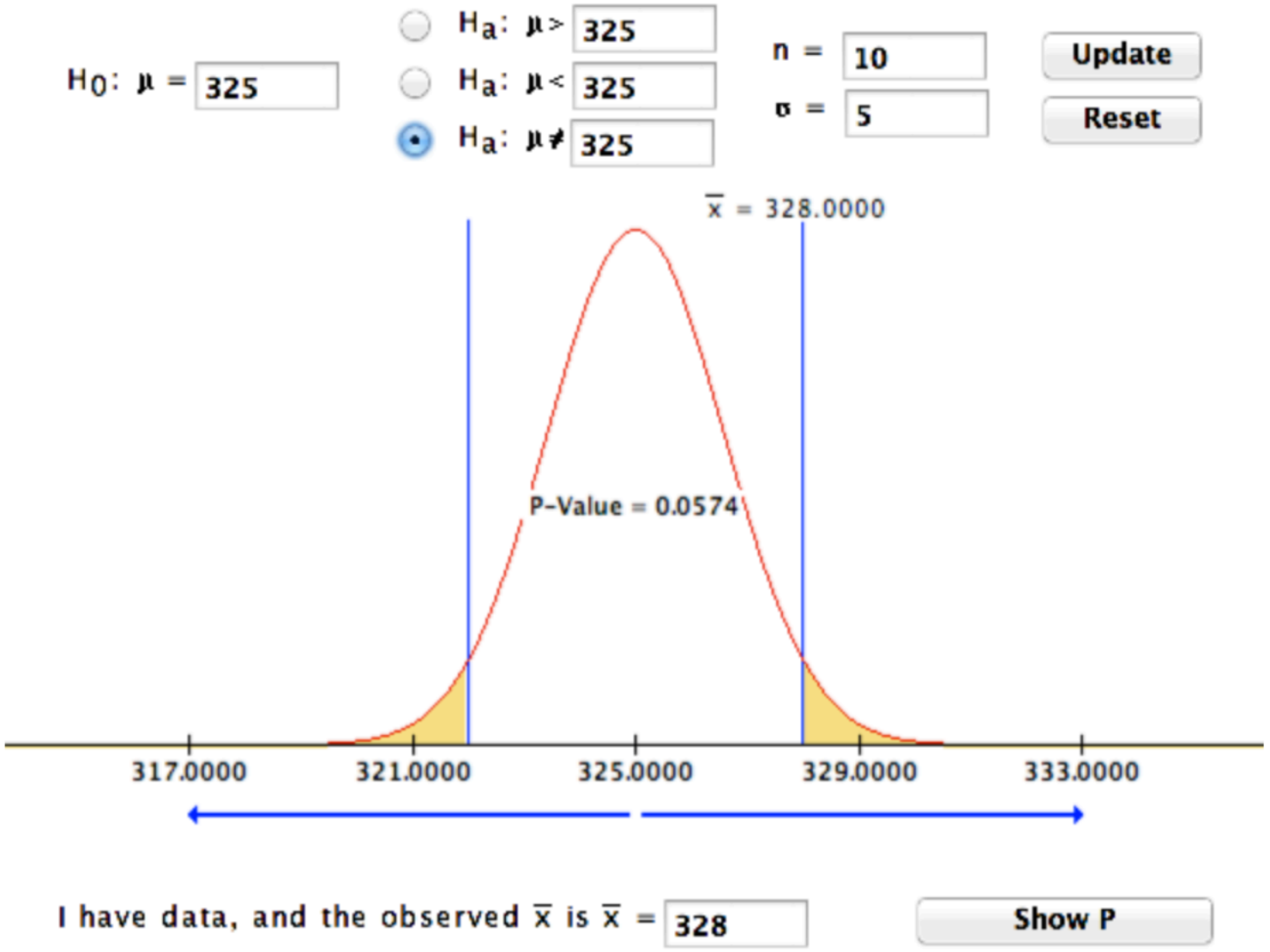
How different is different?

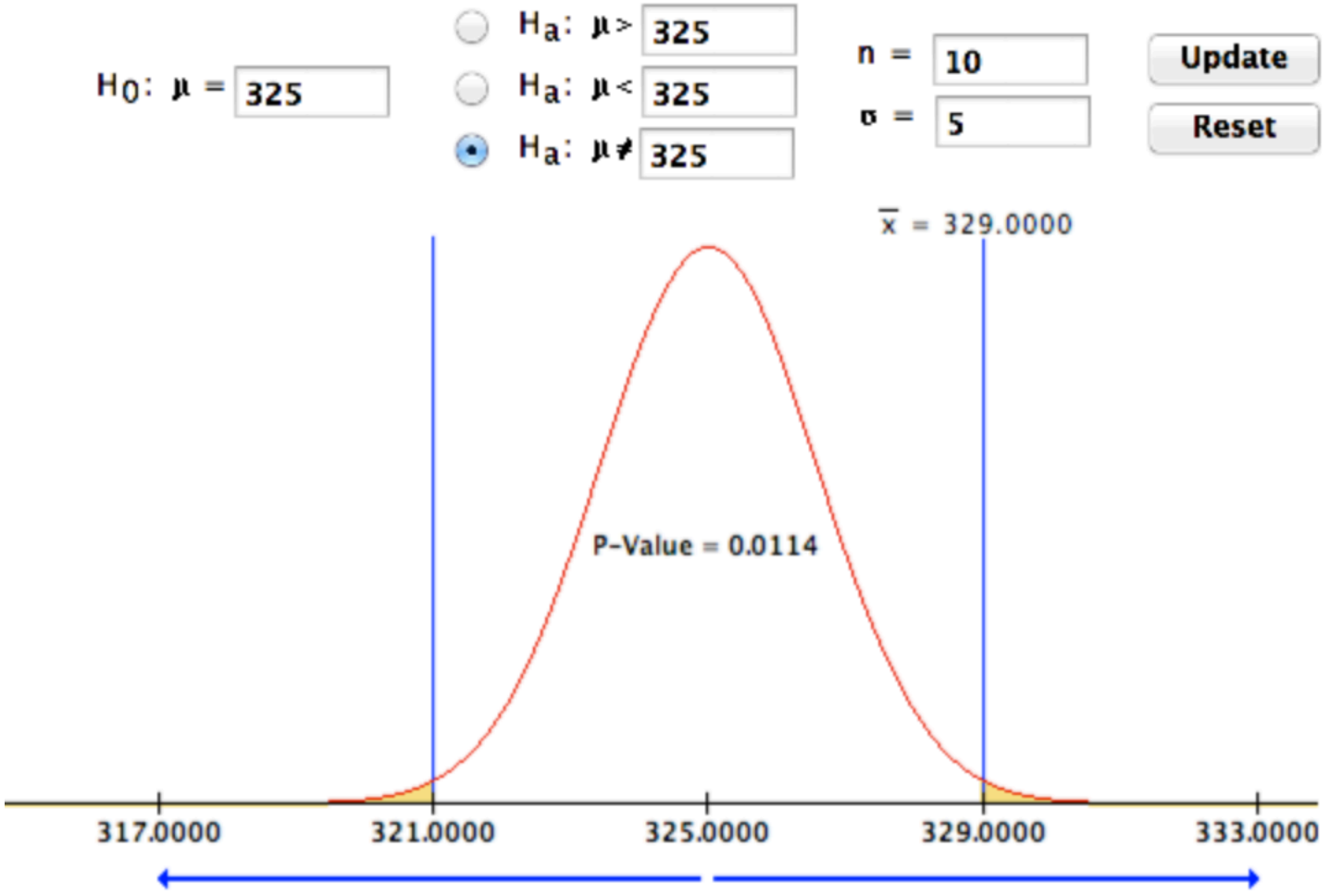
- How do we know the difference in two treatments is not just due to chance?
- We don't. But we can calculate the odds that it is.
- This is the *p-value*
- In repeated experiments at this sample size, how often would you see a result at least this extreme assuming the null hypothesis?

P-value

- If the test is two-sided:
  - P-value = 2 * P(X > |observed value|)
- If the test is one-sided:
  - H_A is $\mu > \mu_0$
  - P-value = P(X > observed value)

  - H_A is $\mu < \mu_0$
  - P-value = P(X < observed value)


H_0: μ = 325
H_a: μ > 325
H_a: μ < 325
H_a: $\mu \neq$ 325
n = 10
σ = 5
Update
Reset
$\bar{x}$ = 328.0000
P-Value = 0.0574
317.0000
321.0000
325.0000
329.0000
333.0000
I have data, and the observed $\bar{x}$ is $\bar{x}$ = 328
Show P



I have data, and the observed $\bar{x}$ is $\bar{x}$ = 329 Show P

UNIVERSITY of WASHINGTON

Jonah Lehrer, 2010, The New Yorker
The Truth Wears off

John Davis, University of Illinois
"Davis has a forthcoming analysis demonstrating that the efficacy of antidepressants has gone down as much as threefold in recent decades."

Anders Pape Møller, 1991
"female barn swallows were far more likely to mate with male birds that had long, symmetrical feathers"
"Between 1992 and 1997, the average effect size shrank by eighty per cent."

Jonathan Schooler, 1990
"subjects shown a face and asked to describe it were much less likely to recognize the face when shown it later than those who had simply looked at it."
The effect became increasingly difficult to measure.

Joseph Rhine, 1930s, coiner of the term extrasensory perception
Tested individuals with card-guessing experiments. A few students achieved multiple low-probability streaks. But there was a "decline effect" – their performance became worse over time.

http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer

Reason 1: Publication Bias

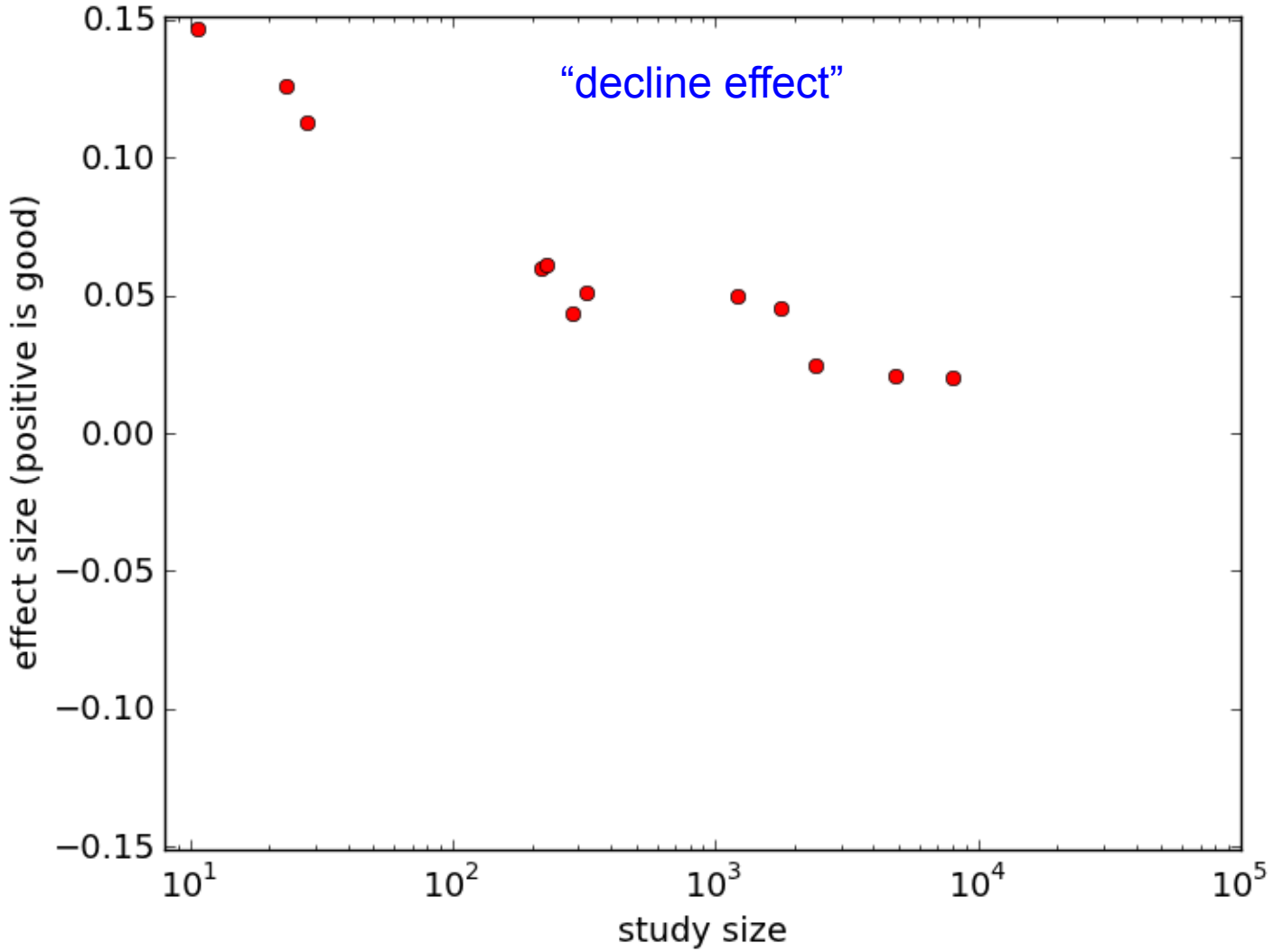
"In the last few years, several meta-analyses have reappraised the efficacy and safety of antidepressants and concluded that the therapeutic value of these drugs may have been significantly overestimated."

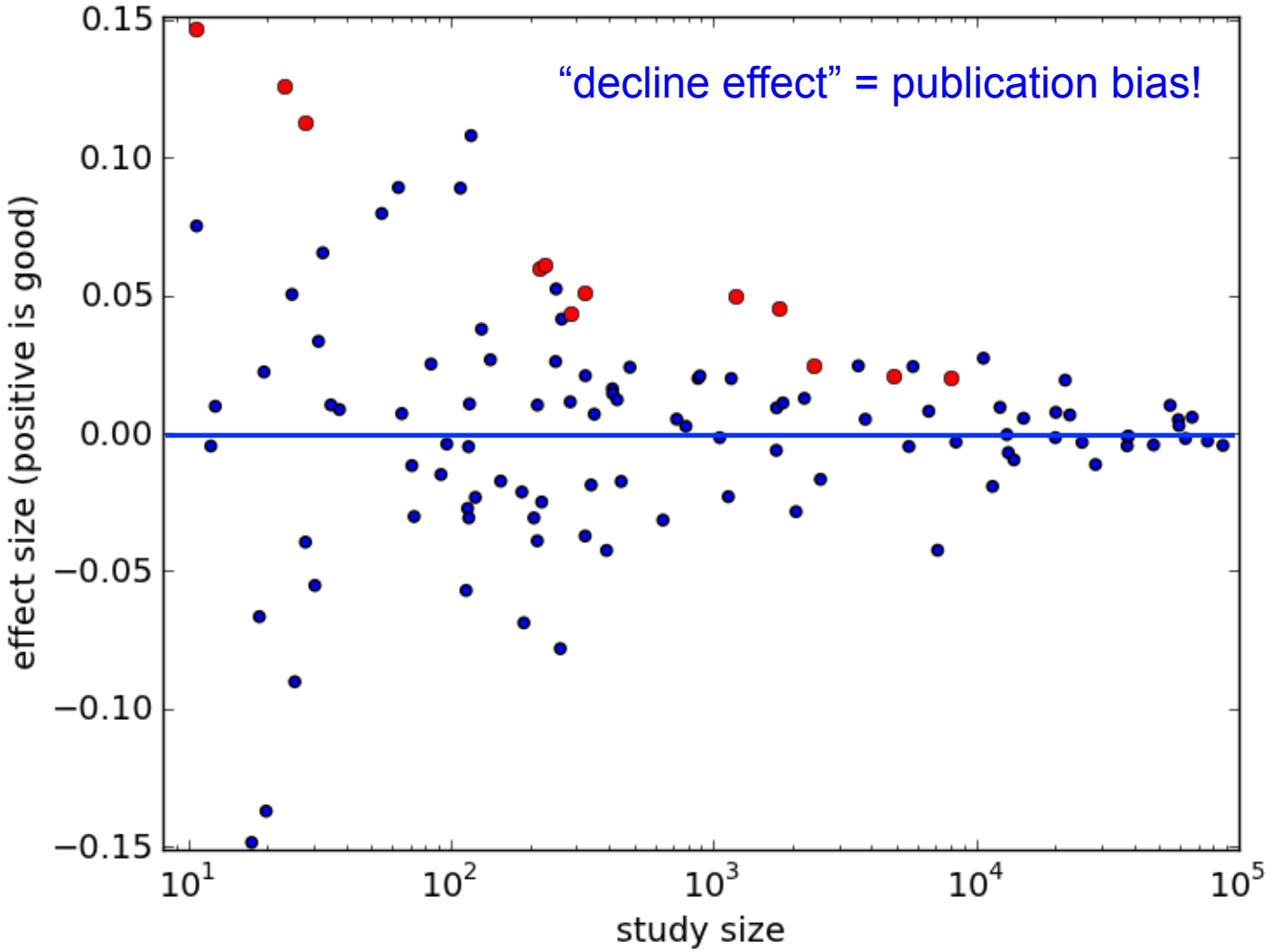
"Although publication bias has been documented in the literature for decades and its origins and consequences debated extensively, there is evidence suggesting that this bias is increasing."

"A case in point is the field of biomedical research in autism spectrum disorder (ASD), which suggests that **in some areas negative results are completely absent**" *(emphasis mine)*

"… a highly significant correlation (R^2= 0.13, $p < 0.001$) between impact factor and overestimation of effect sizes has been reported."

Publication bias: What are the challenges and can they be overcome?
Ridha Joober, Norbert Schmitz, Lawrence Annable, and Patricia Boksa
J Psychiatry Neurosci. 2012 May; 37(3): 149–152. doi: 10.1503/jpn.120065


"decline effect"
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
effect size (positive is good)
10^1
10^2
10^3
10^4
10^5
study size


"decline effect" = publication bias!
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
effect size (positive is good)
10^1
10^2
10^3
10^4
10^5
study size