A GENERALIZED EXCESS SIGNIFICANCE TEST FOR SELECTIVE OUTCOME REPORTING WITH DEPENDENT EFFECT SIZES

James E. Pustejovsky
& Melissa A. Rodgers
UT Austin

SRSM, Chicago, IL
July 22, 2019
pusto@austin.utexas.edu

BEDIOU ET AL. (2018). META-ANALYSIS OF ACTION VIDEO GAME IMPACT ON PERCEPTUAL, ATTENTIONAL, AND COGNITIVE SKILLS.

70 studies (88 samples), 194 effect size estimates (standardized mean differences) measuring differences between video gamers and non-gamers.

- ✓ Multiple outcomes
- ✓ Multiple treatment groups
- ✓ Multiple comparison groups
- ✓ Multiple follow-up times
- ✓ 1-28 effect size estimates per study (median = 2)

WE NEED METHODS TO DETECT SELECTIVE OUTCOME REPORTING WITH DEPENDENT EFFECTS

- Many methods available to detect selective outcome reporting, publication bias, small-study effects (funnel plot asymmetry).

- But nearly all available methods assume effect size estimates are independent.
  - Exception: cluster-robust Egger's regression ("Egger sandwich")

- Aim: Develop an Excess Significance Test so that it can be used in syntheses of dependent effect sizes.

TEST OF EXCESS SIGNIFICANCE (TES) (IOANNIDIS & TRIKALINOS, 2007)

- TES assesses selective reporting using the statistic:

$$\frac{O - \hat{E}}{\sqrt{V^{Binom}}}$$

where

O = observed number of statistically significant effects
$\hat{E}$ = expected number of statistically significant effects
$V^{Binom} = \hat{E}\left(k - \hat{E}\right)/k$

- $\hat{E}$ is estimated power under a fixed or random effects model.

GENERALIZED EXCESS SIGNIFICANCE TEST

- Suppose that we have k studies, each with one or more ES.
- O_i = observed number of significant ES from study *i*
- $\hat{E}_i$ = expected number of significant ES from study *i*
- Define the score statistic

$$S_\pi = \frac{1}{k}\sum_{i=1}^{k}\left(O_i - \hat{E}_i\right).$$

- In the absence of publication bias, $\mathbf{E}[S_\pi] = 0$.

GENERALIZED EXCESS SIGNIFICANCE TEST (CONTINUED)

- A cluster-robust score test statistic (Rotnitzky & Jewell, 1990):

$$Z^{GEST} = \frac{S_\pi}{\sqrt{V^{CR}}}$$

where V^{CR} is a cluster-robust estimate of the variance of S_π.

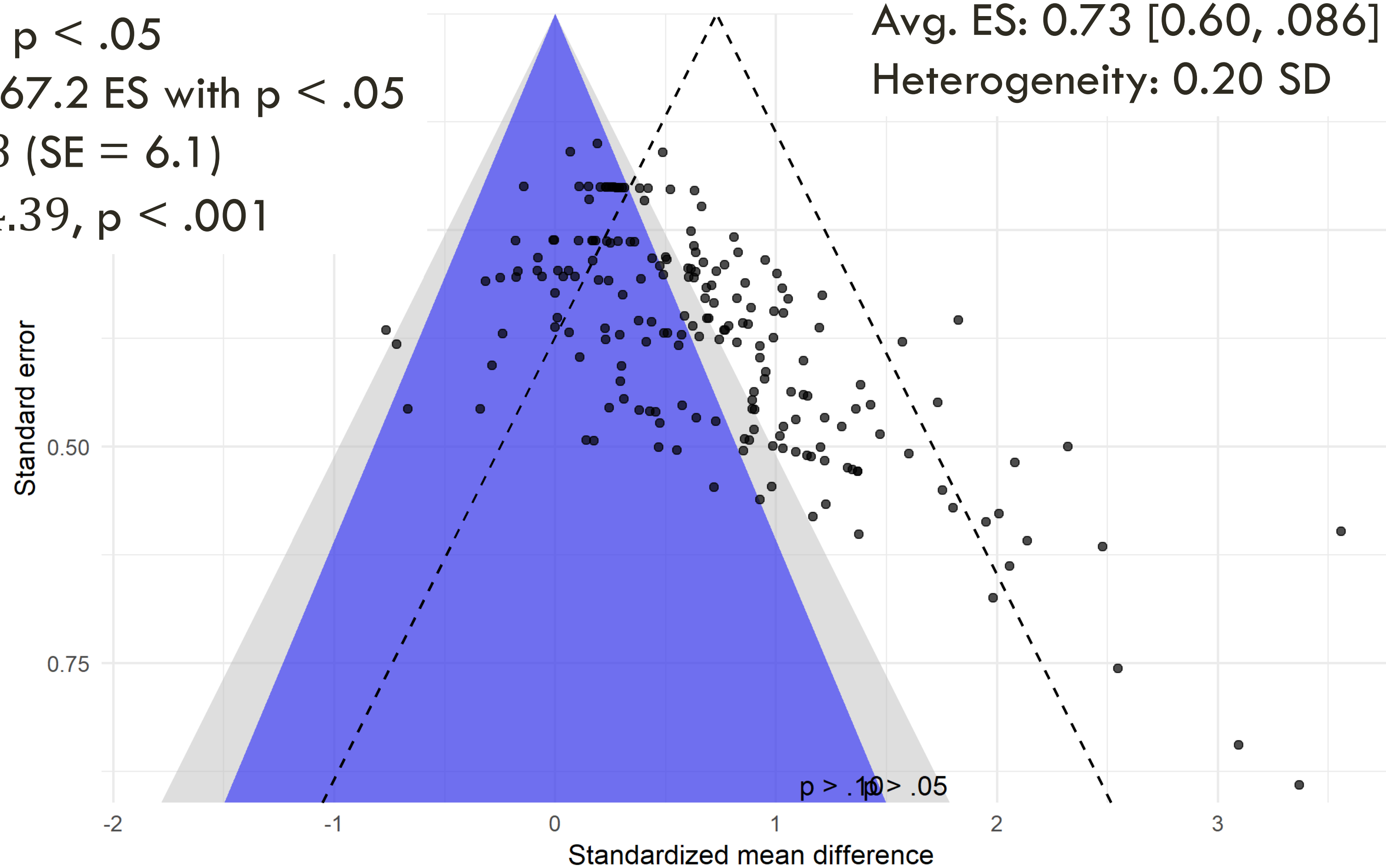
- In the absence of selective reporting, $Z^{GEST} \dot{\sim} N(0,1)$ for large *k*.
- Selective outcome reporting is indicated if

$$Z^{GEST} > \Phi^{-1}(1-\alpha).$$

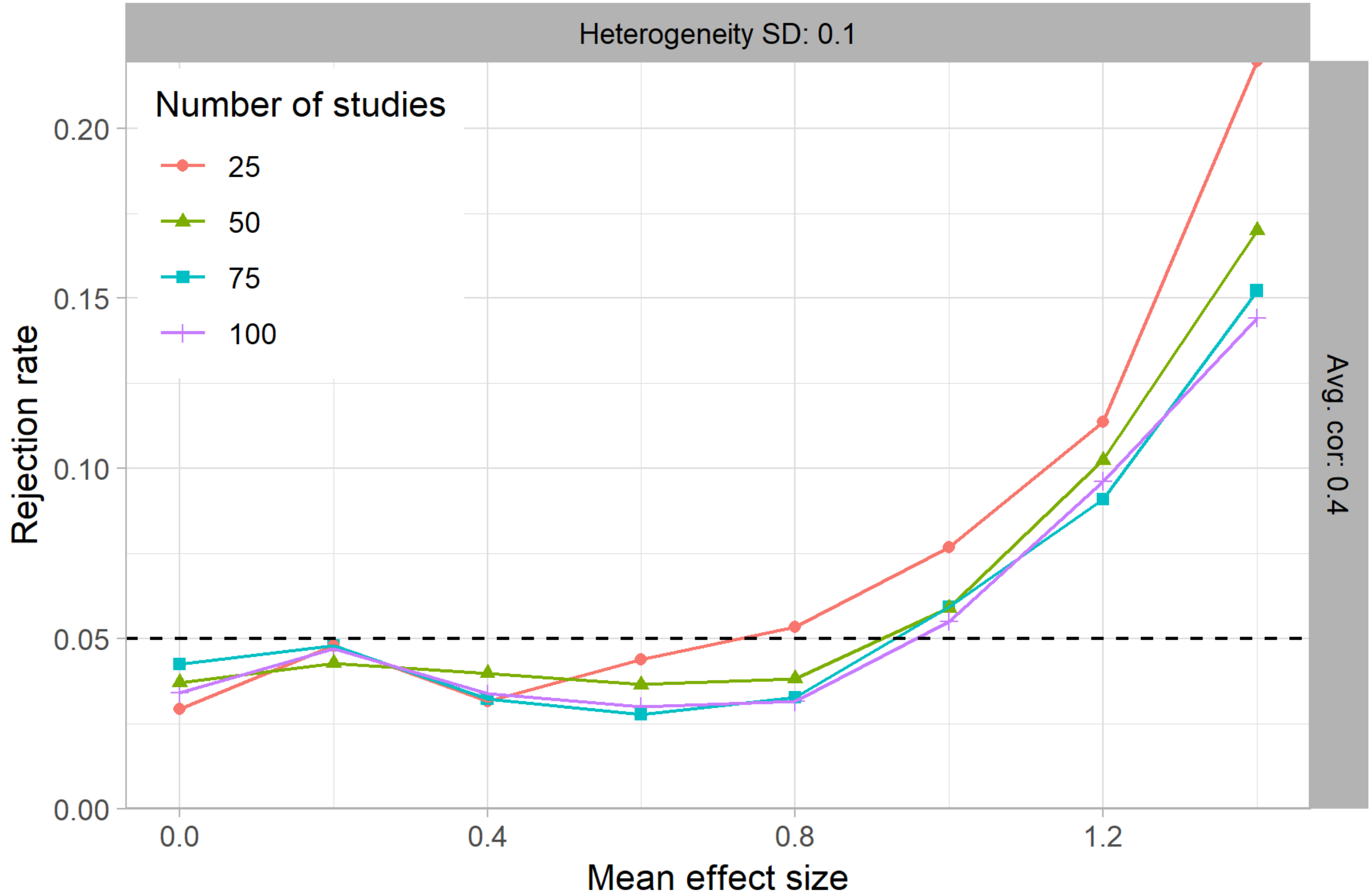
BEDIOU ET AL. (2018). META-ANALYSIS OF ACTION VIDEO GAME IMPACT ON PERCEPTUAL, ATTENTIONAL, AND COGNITIVE SKILLS.

- 94 ES with p < .05
- Expected: 67.2 ES with p < .05
- $S_{pi} = 26.8$ (SE = 6.1)
- $Z^{GEST} = 4.39$, p < .001

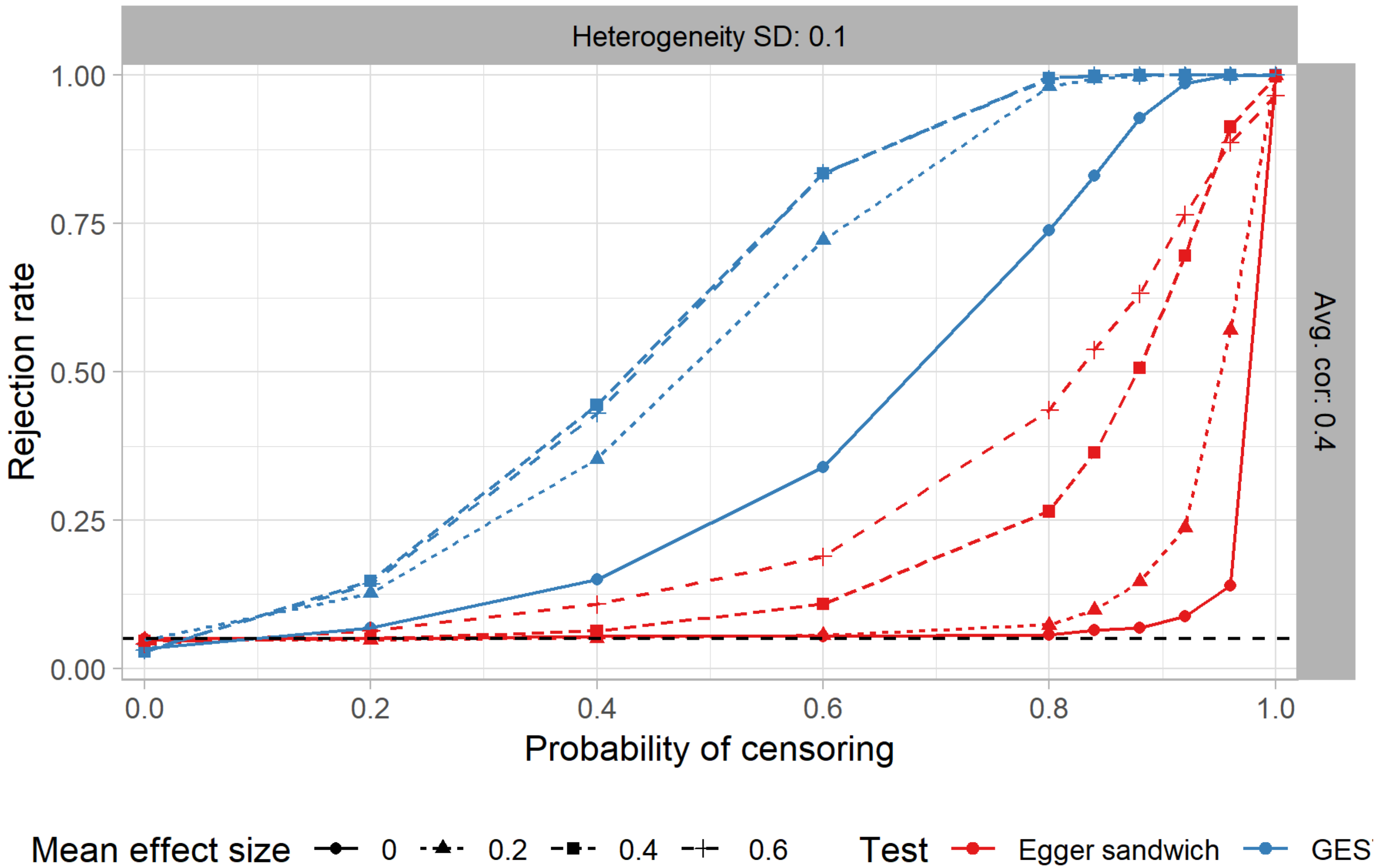
Avg. ES: 0.73 [0.60, .086]
Heterogeneity: 0.20 SD



SIMULATIONS: TYPE-I ERROR RATES (CORRELATED STANDARDIZED MEAN DIFFERENCES)



SIMULATIONS: POWER COMPARISON (K = 50)



DISCUSSION

- GEST requires consistent estimates of mean and variance of ES distribution in the absence of selection.
  - Can accommodate meta-regression models.
  - Can use weighting schemes that are not inverse-variance (e.g., Henmi & Copas, 2010).

- GEST involves estimating expected power *marginally* for each ES.
  - Does not consider the joint pattern of statistically significance.

- Type-I error rates are inflated when average effects are large and homogeneous (i.e., all studies have high power).
  - Small-sample refinements need further work.

Primary Investigator: “I’m not really concerned about selective outcome reporting.”

Statistician: “Surely you GEST?”

James E. Pustejovsky
pusto@austin.utexas.edu
https://jepusto.com

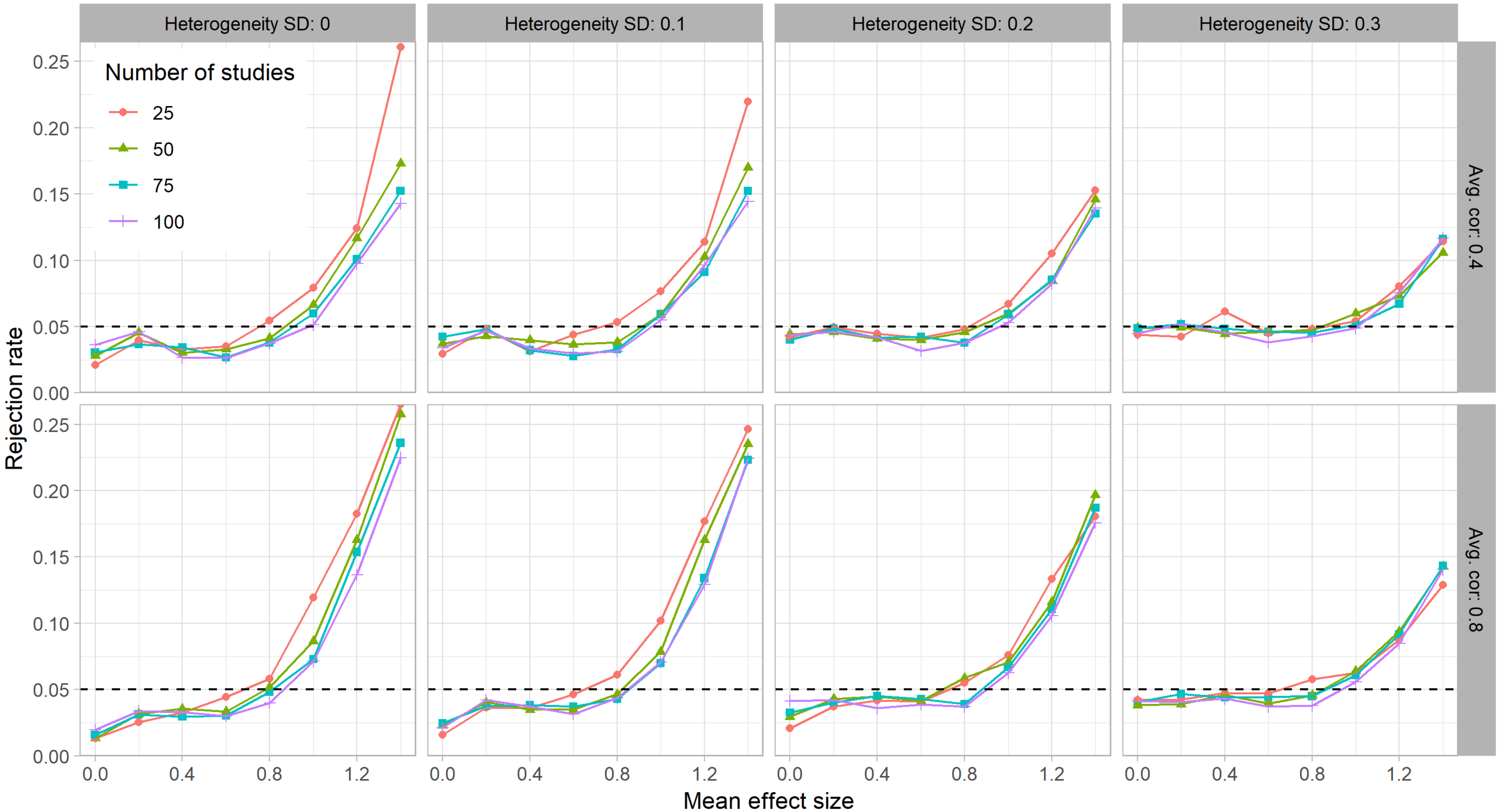
REFERENCES

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*(1), *77*.

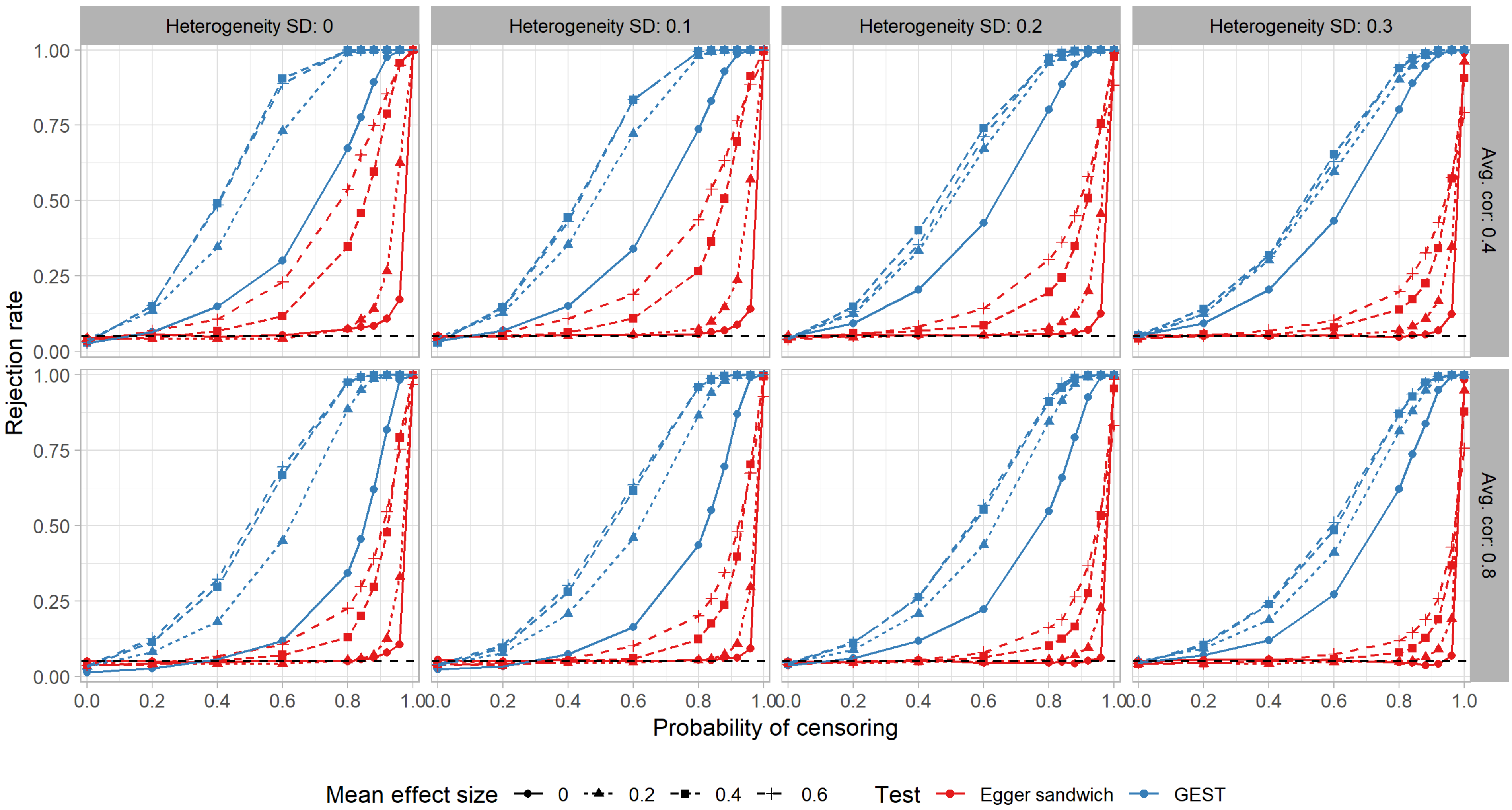
Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*(3), 245-253.

Rotnitzky, A., & Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, *77*(3), 485-497.

SIMULATIONS: TYPE-I ERROR RATES (CONTINUED)



SIMULATIONS: POWER COMPARISON (K = 50)



SIMULATIONS: POWER COMPARISON (K = 100)