Supplementary Material

1 Data generation Directed Acyclic Graph

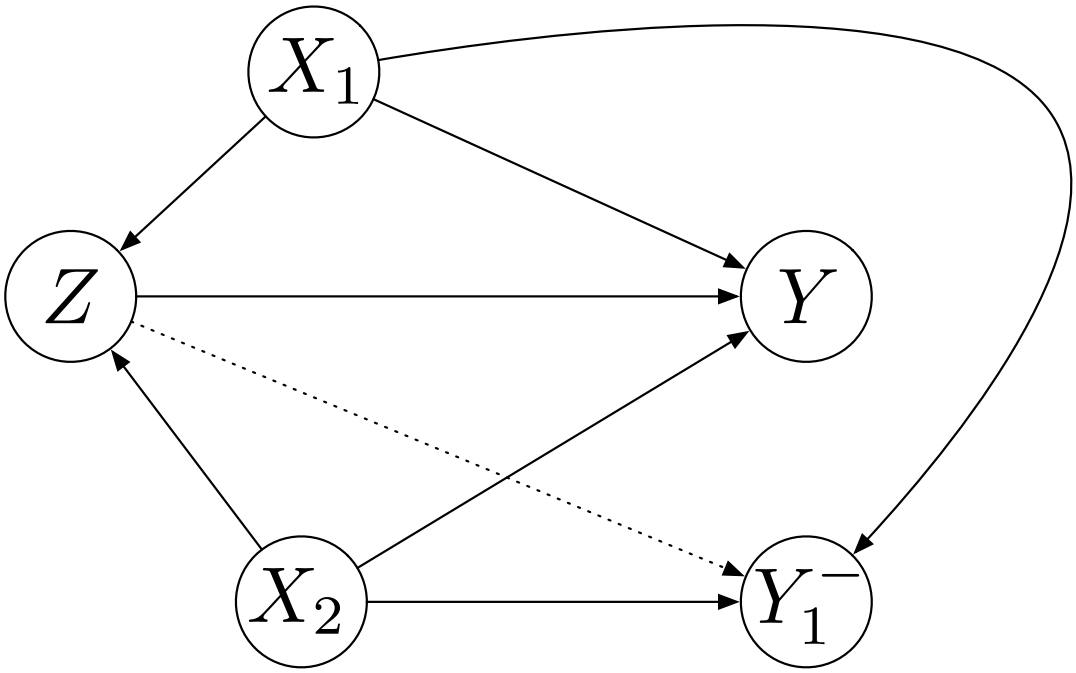


Figure 1. Direct acyclic graph of simulation setup with outcome of interest Y, one negative control outcome Y_1^- and two measured confounders X_1 and X_2. The dashed line between the treatment Z and the negative control indicates a potential association, for an ideal negative control.

2 Modelling the systematic error in empirical calibration

Suppose there are a total of S negative control outcomes, each with V corresponding number of positive controls. Let denote $\theta_i, i \in \{1,2,\dots,S \times V\}$ as the treatment effect of each individual control with $\theta_i = 0$ for negative controls and $\theta_i > 0$ for positive controls. An estimate of the treatment effect can be modelled as

$$\widehat{\theta_i} = \theta_i + \epsilon_i \quad , \quad \epsilon_i \sim \mathcal{N}(0, \tau^2)$$

such that $\widehat{\theta_i} - \theta_i$ is due to the random error term ϵ_i.

Empirical calibration assumes that there is a *systematic bias* component to the error term, (here denoted by ψ_i such that

$$\epsilon_i \sim \mathcal{N}(\psi_i, \tau^2).$$

The bias ψ_i is assumed to come from another Gaussian distribution

$$\psi_i \sim \mathcal{N}\{\mu(\theta_i), \sigma^2(\theta_i)\}$$

with the mean and variance estimated as parametric functions of θ_i

$$\mu(\theta_i) = \alpha + \gamma\theta_i \quad (1)$$

$$\log\{\sigma(\theta_i)\} = \phi + \nu\theta_i \quad (2)$$

The parameters in **Error! Reference source not found.** and **Error! Reference source not found.** α,γ,ϕ,ν are estimated by maximising the likelihood of observing the joint probability distribution of the systematic bias ψ_i and estimates from all controls $\widehat{\theta_i}$

$$\Pr(\widehat{\theta_i}, \psi_i) = \Pr(\widehat{\theta_i} \mid \psi_i)\Pr(\psi_i). \quad (3)$$

Since we do not observe ψ_i, we integrate them out from **Error! Reference source not found.** to obtain the marginal probability distribution of our estimates $\widehat{\theta_i}$ as a function of the known θ_i

$$\Pr(\widehat{\theta_i}) = \int \Pr(\widehat{\theta_i}, \psi_i)\, d\psi_i.$$

The likelihood of this marginal distribution is then given by

$$\mathcal{L}(\alpha, \gamma, \phi, \nu) \sim \propto \prod_{i=1}^{S\times V} \int \Pr(\theta_i \mid \psi_i, \theta_i, \widehat{\tau_i}) \Pr(\psi_i \mid \alpha, \gamma, \phi, \nu, \theta_i)\, d\psi_i .$$

A new estimate is then assumed to come from an empirical Gaussian distribution which incorporates the estimate error. See (1), (2) for more detail.

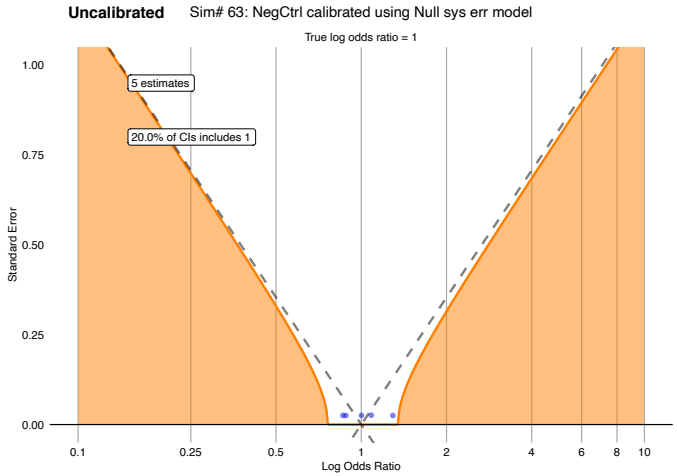
1. Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(11):2571-7.
2. Schuemie MJ, Ryan PB, Pratt N, Chen R, You SC, Krumholz HM, et al. Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study. Journal of the American Medical Informatics Association. 2020;27(8):1268-77.

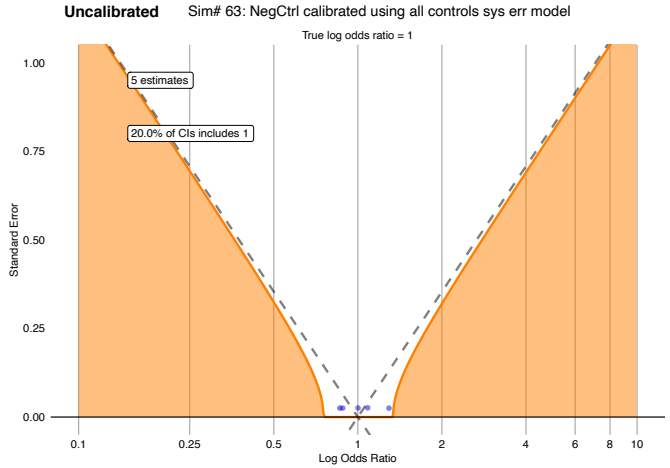
3 Coverage plots of calibration of controls

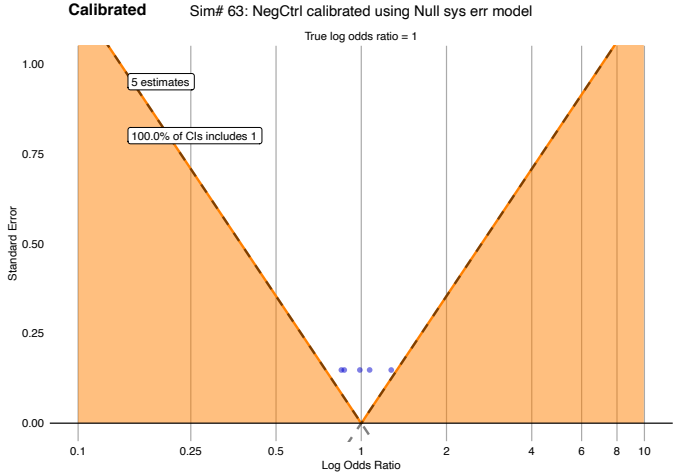
For each bias simulation scenario, coverage plots of uncalibrated and calibrated estimates and corresponding standard errors for positive and negative controls are shown. The estimates are stratified by true effect size.

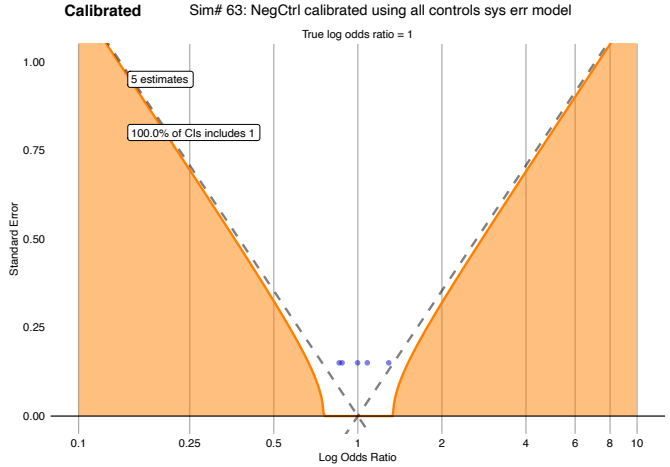
3.1 Unmeasured Confounder Scenario

3.1.1 Calibration of Negative Controls (Five)

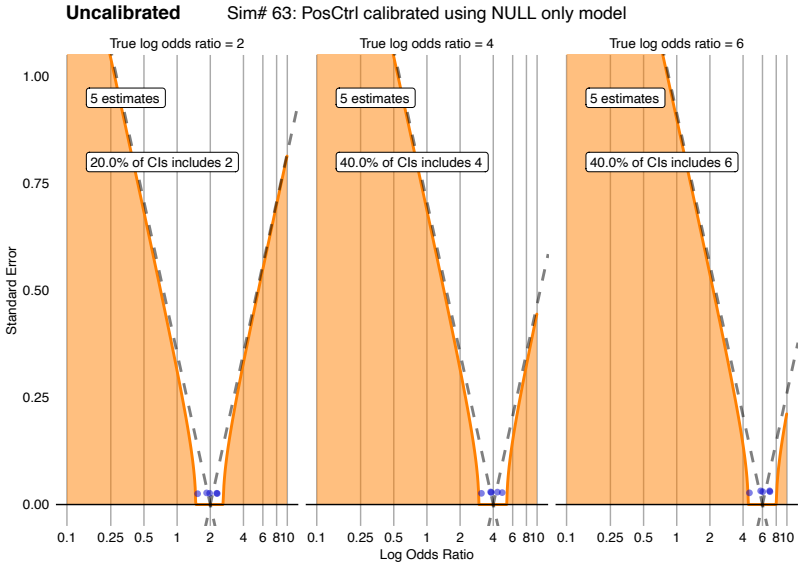


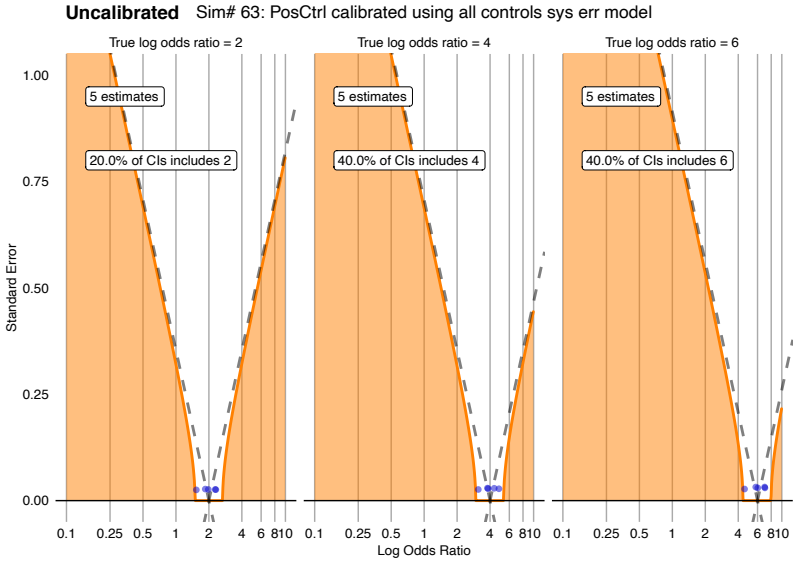


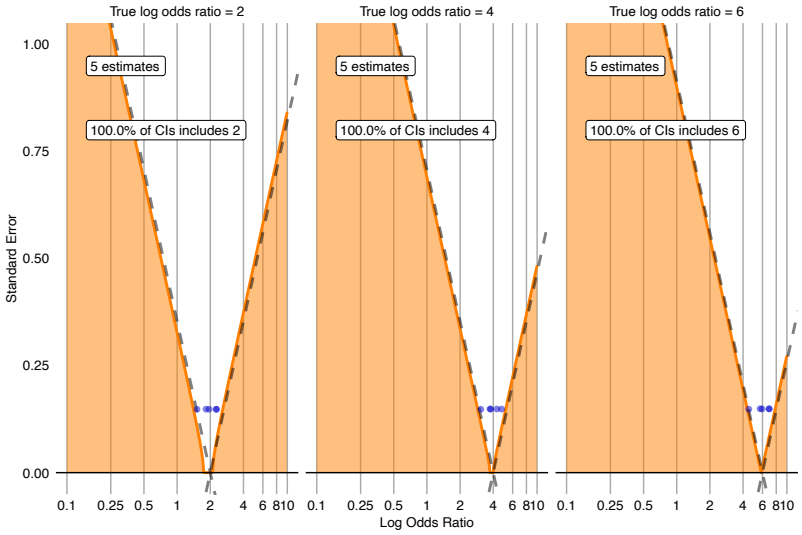


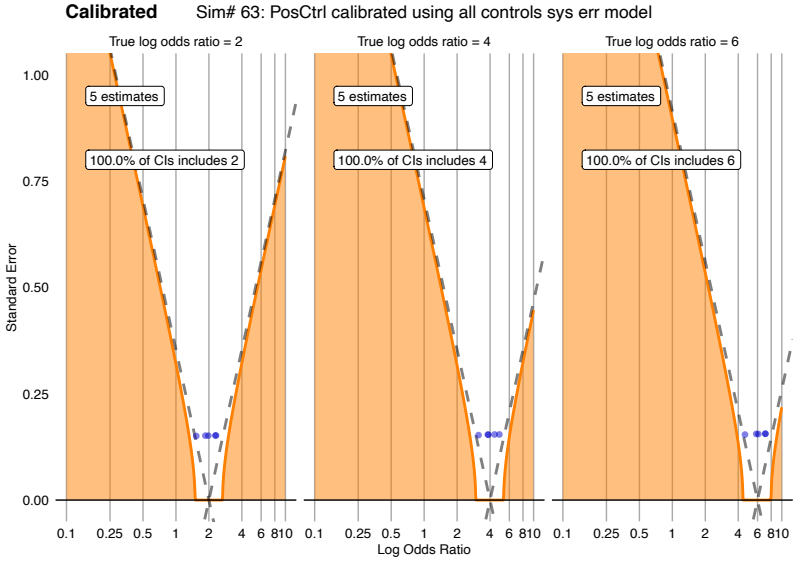


3.1.2 Calibration of Positive Controls (Five)

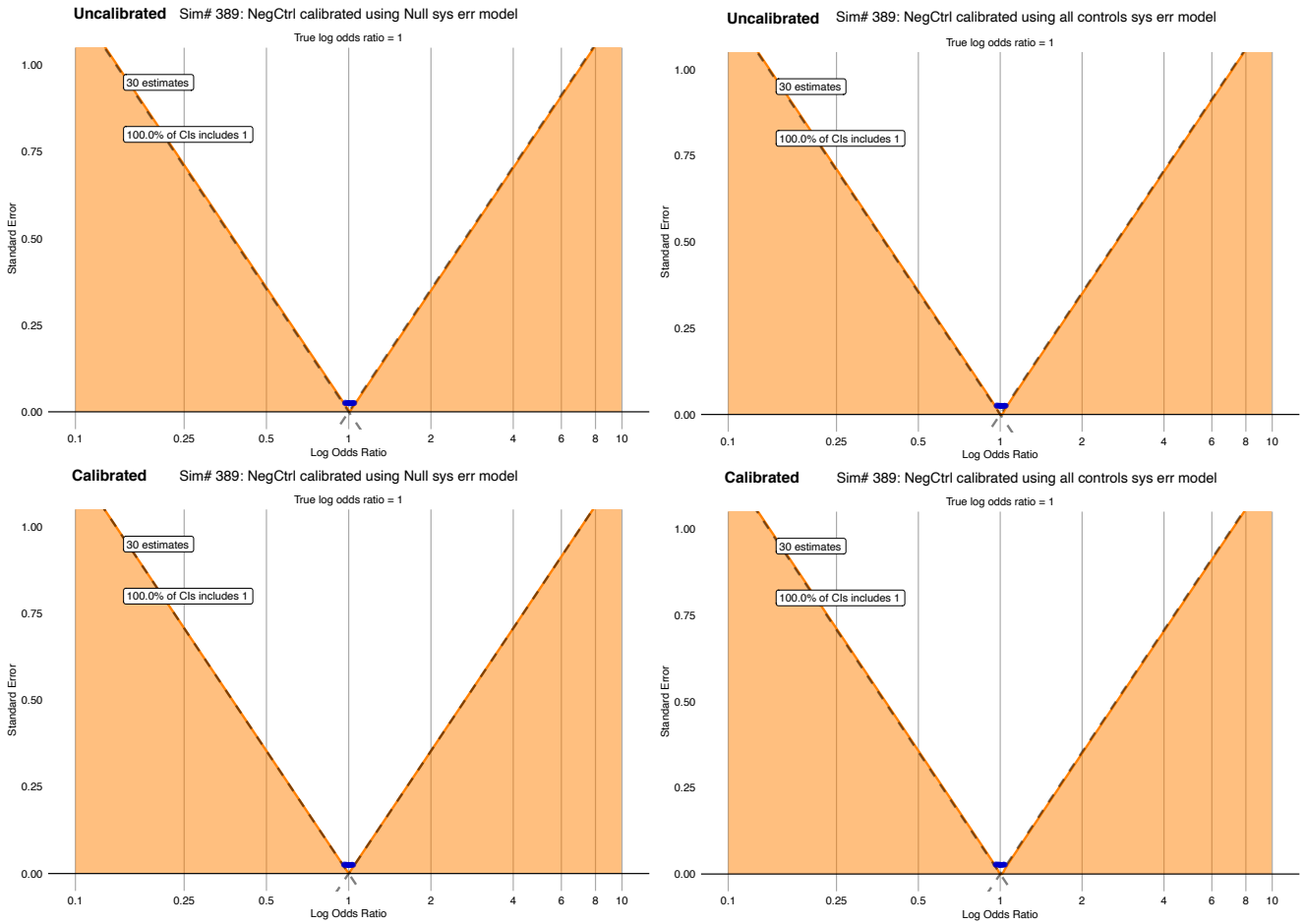




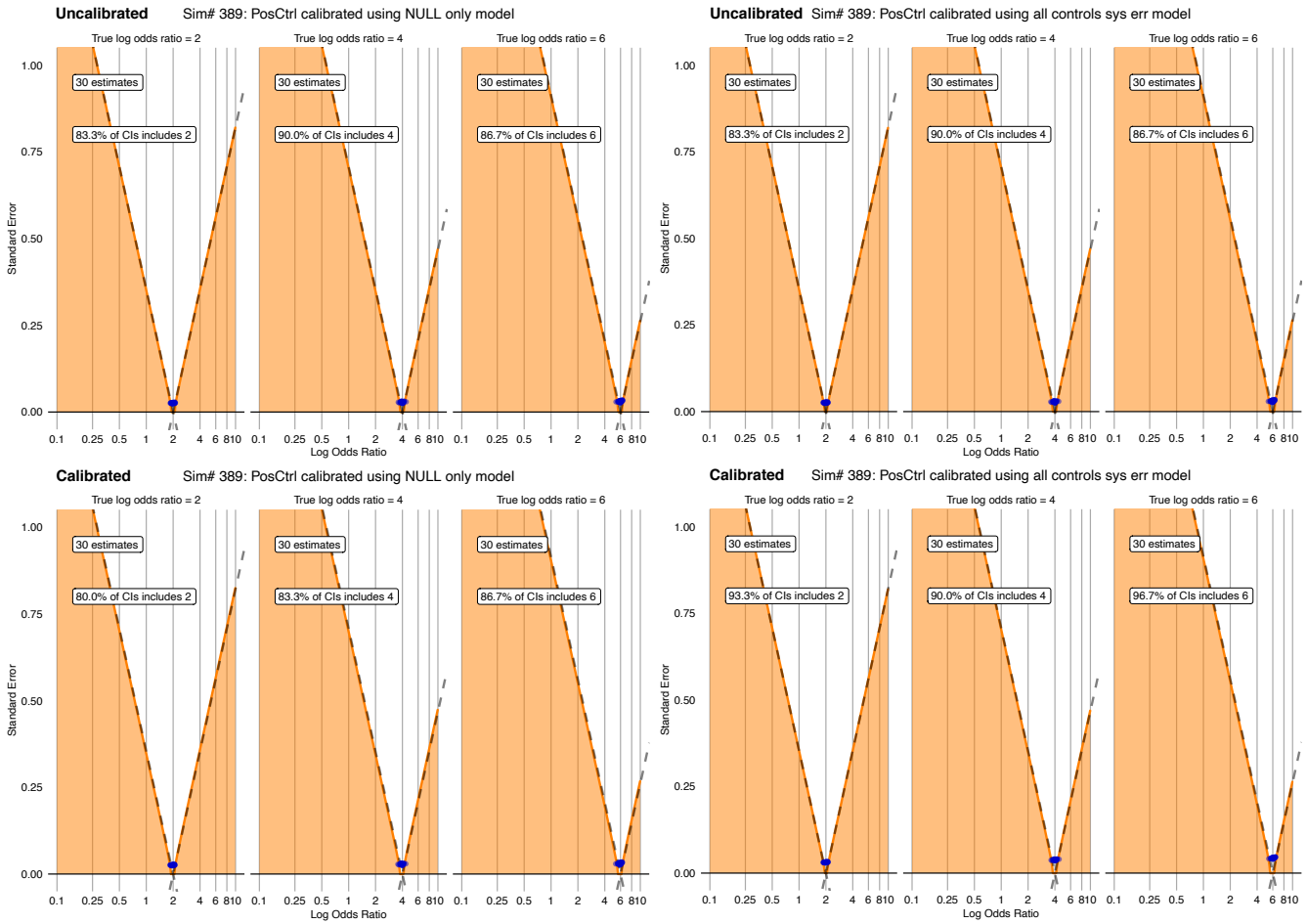




3.1.3 Calibration of Negative Controls (30)

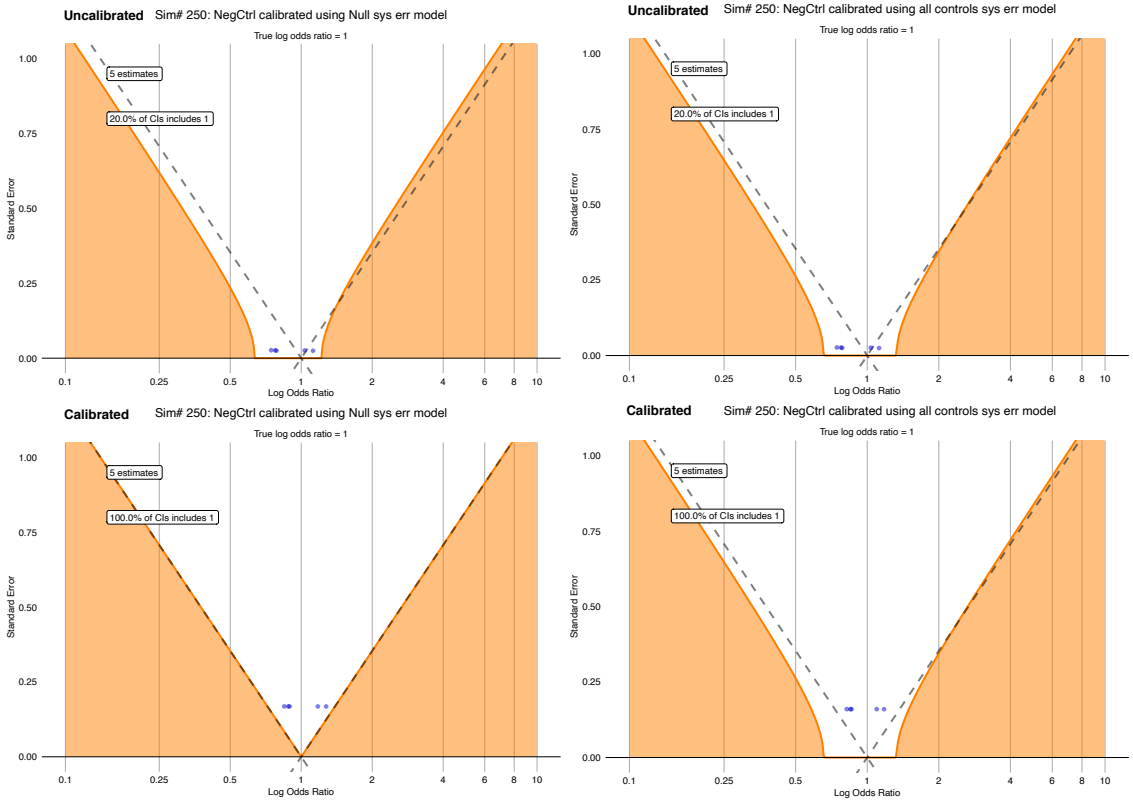


3.1.4 Calibration of Positive Controls (30)

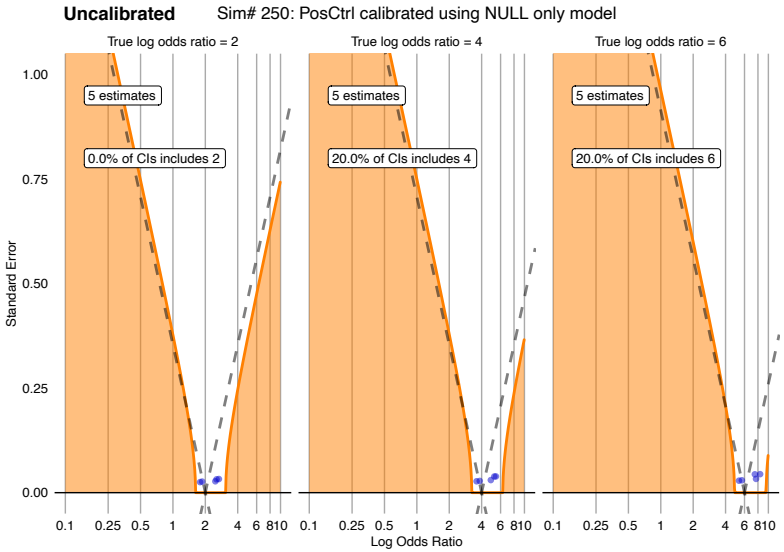


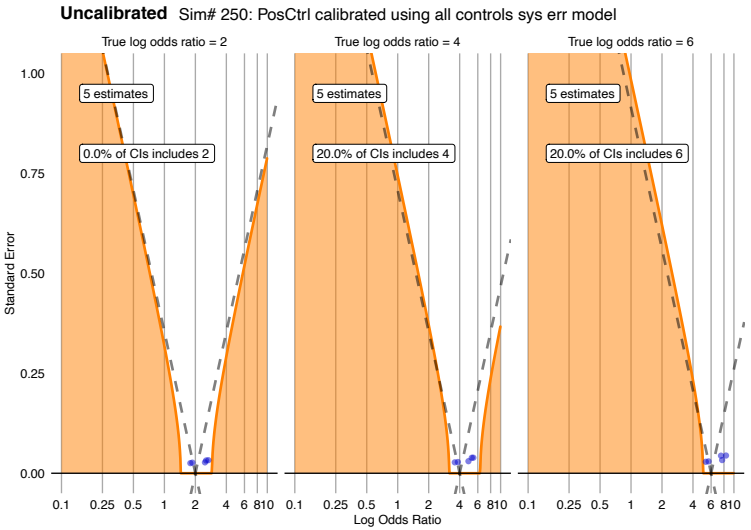
3.2 Quadratic Term Scenario

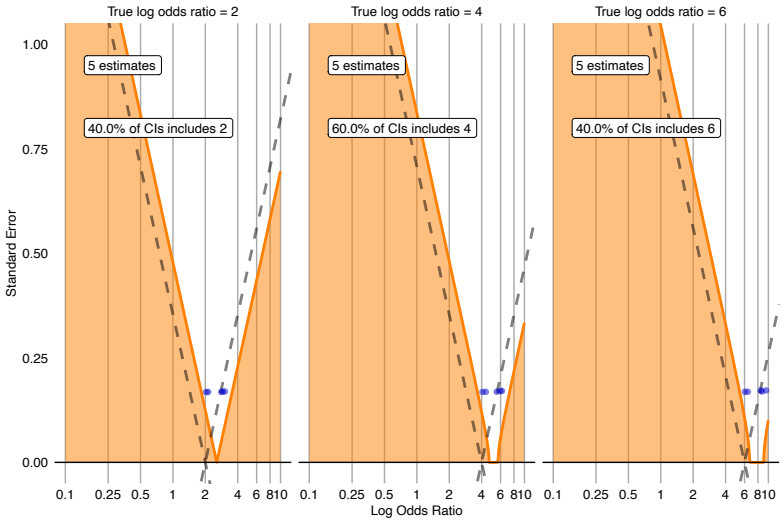
3.2.1 Calibration of Negative Controls (Five)

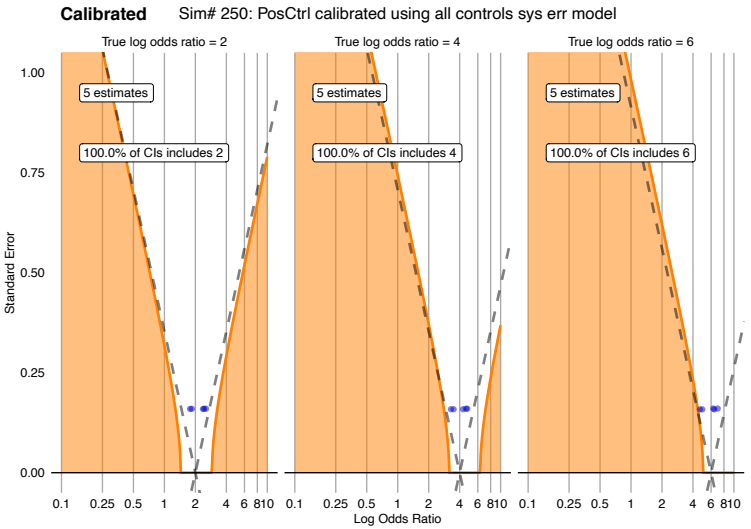


3.2.2 Calibration of Positive Controls (Five)

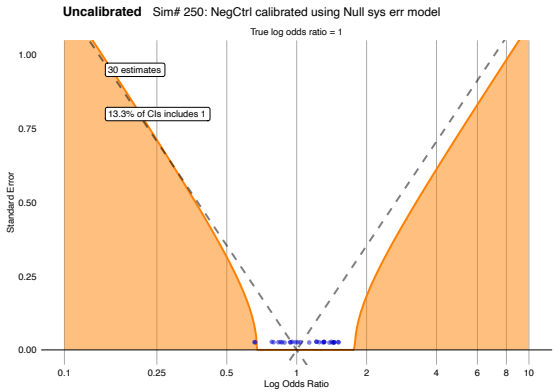


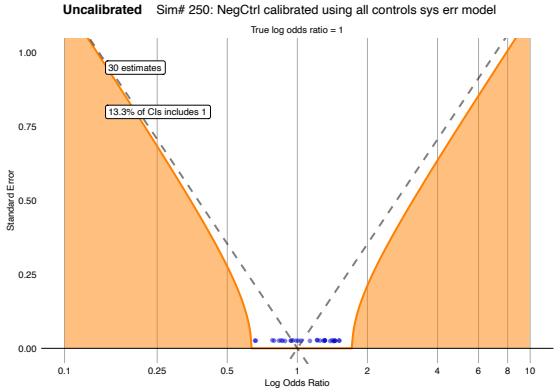


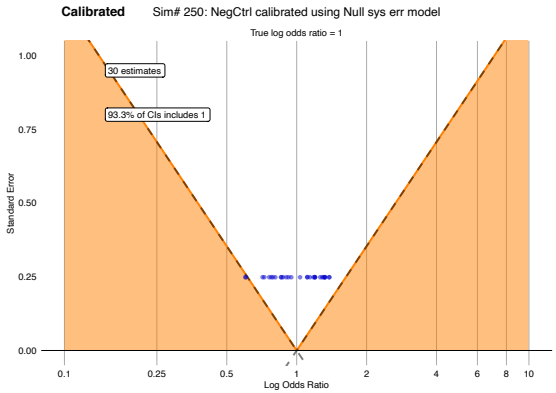


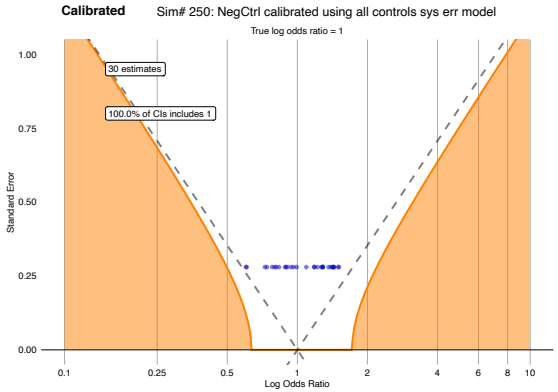


Calibration of Negative Controls (30)

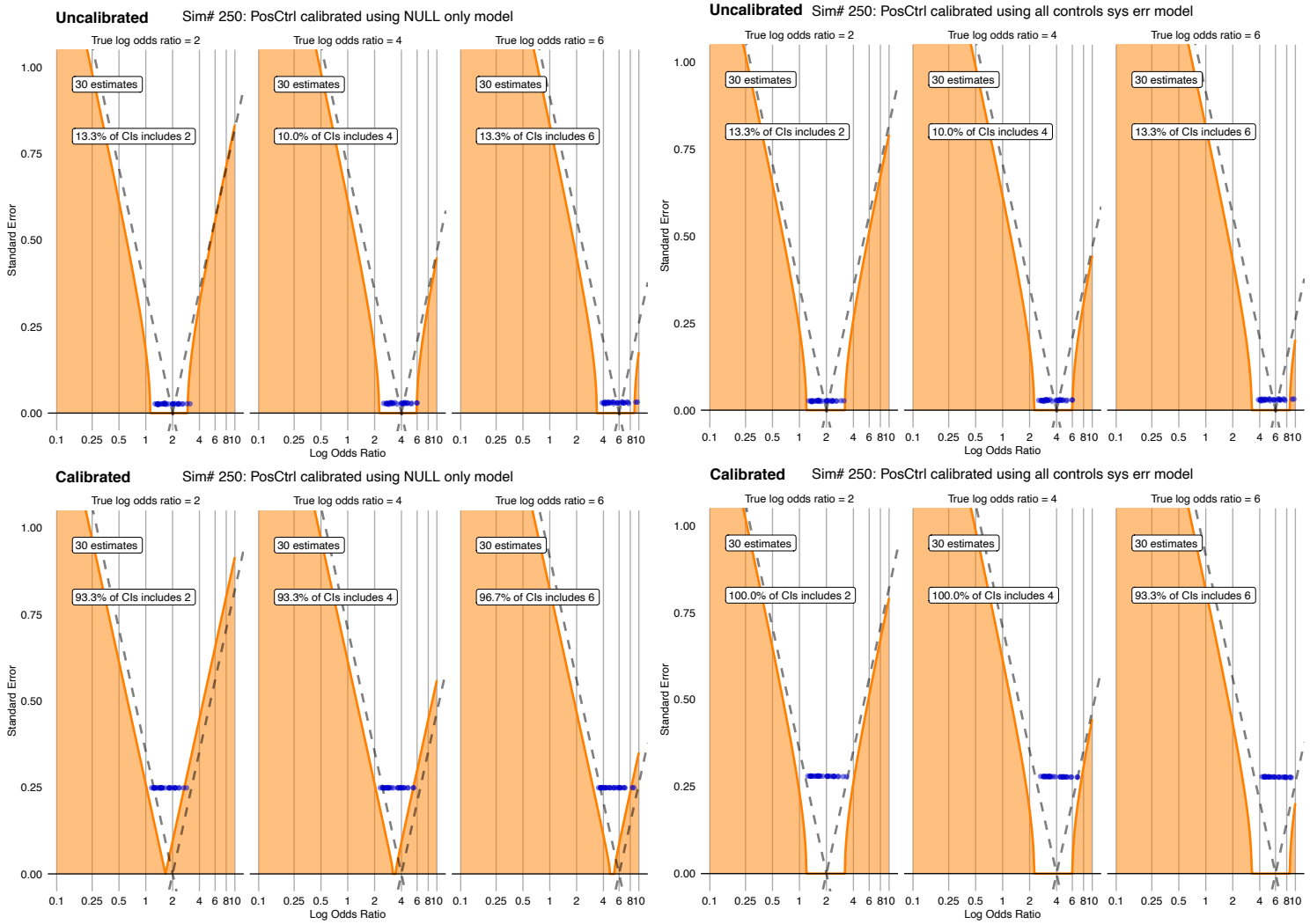






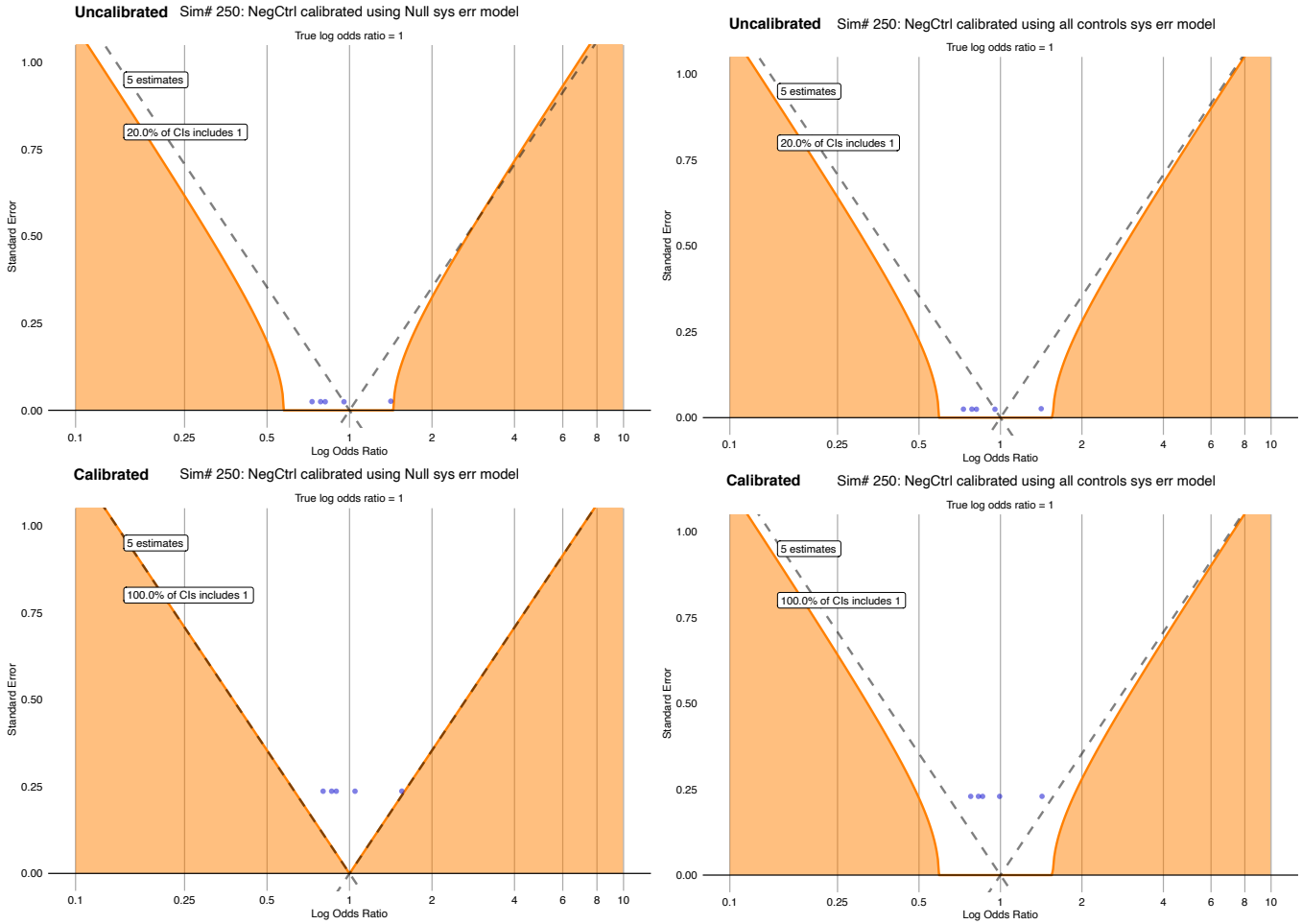


3.2.3 Calibration of Positive Controls (30)

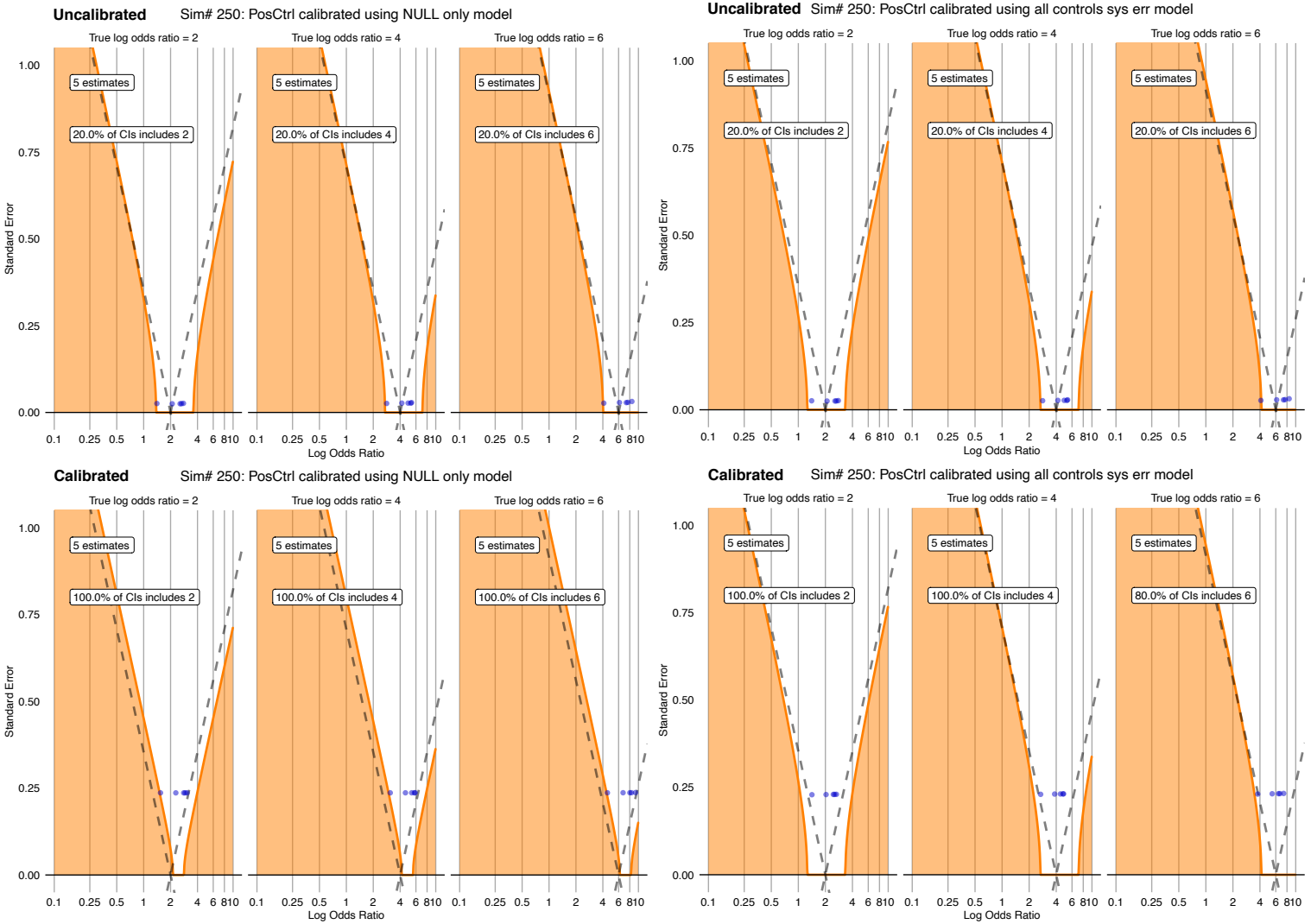


3.3 Interaction between Two Confounders

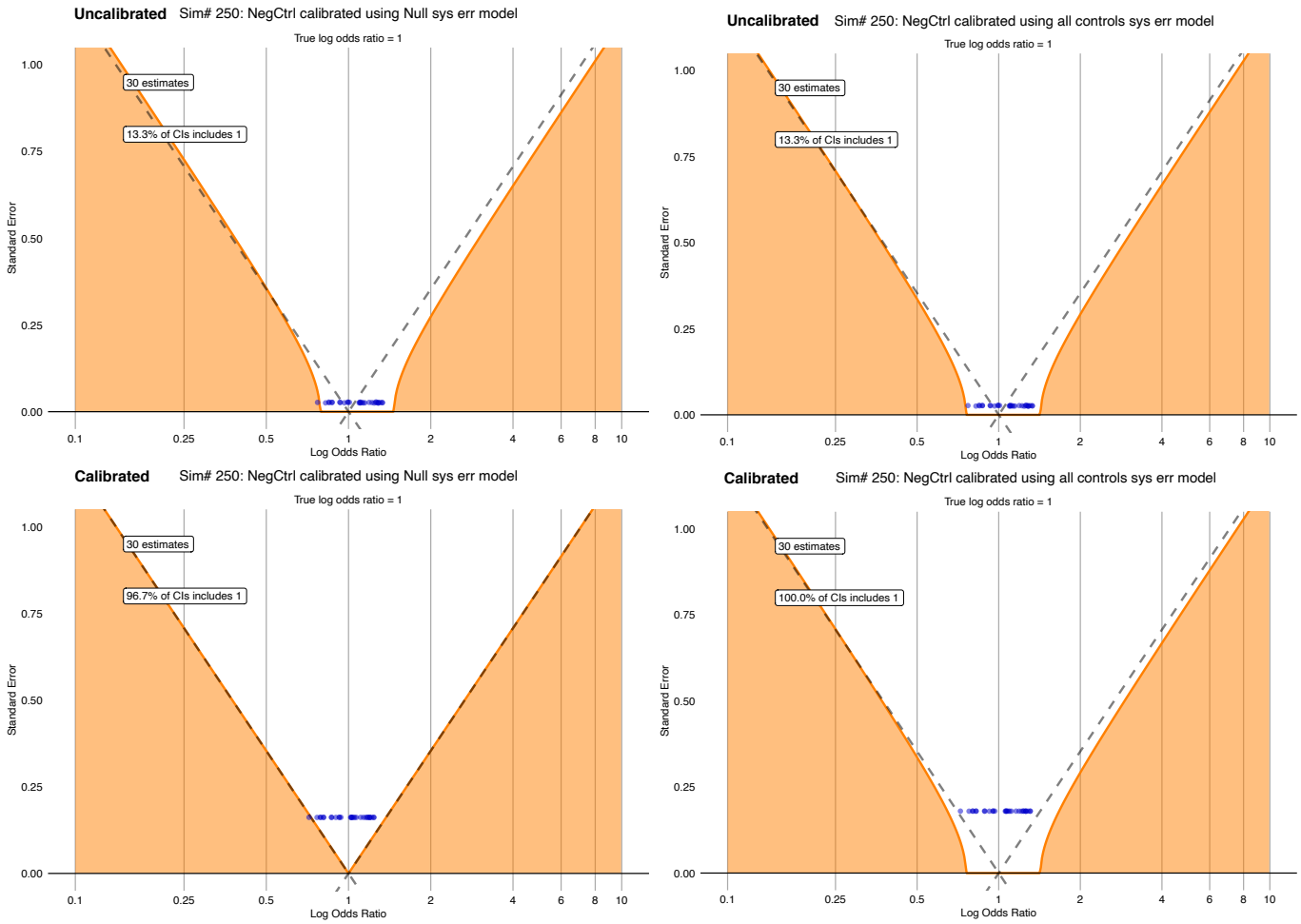
3.3.1 Calibration of Negative Controls (5)



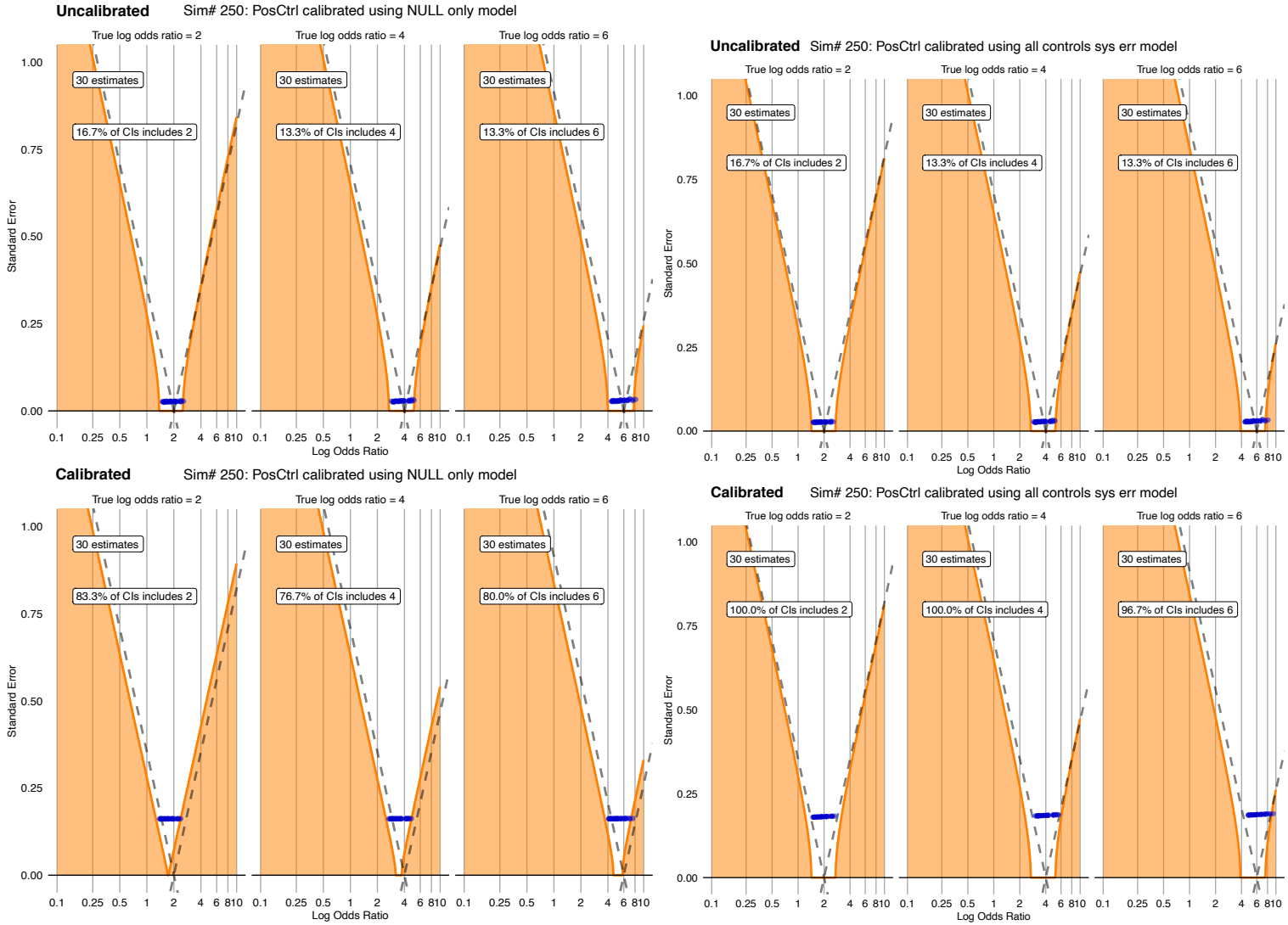
3.3.2 Calibration of Positive Controls (5)



3.3.3 Calibration of Negative Controls (30)

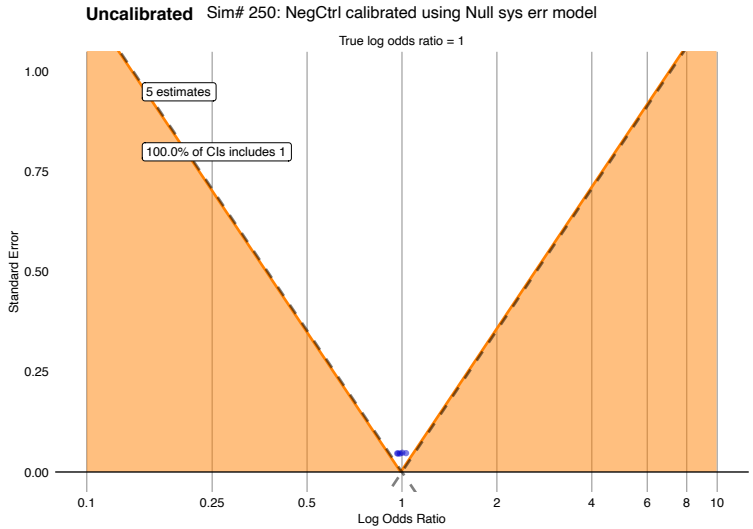


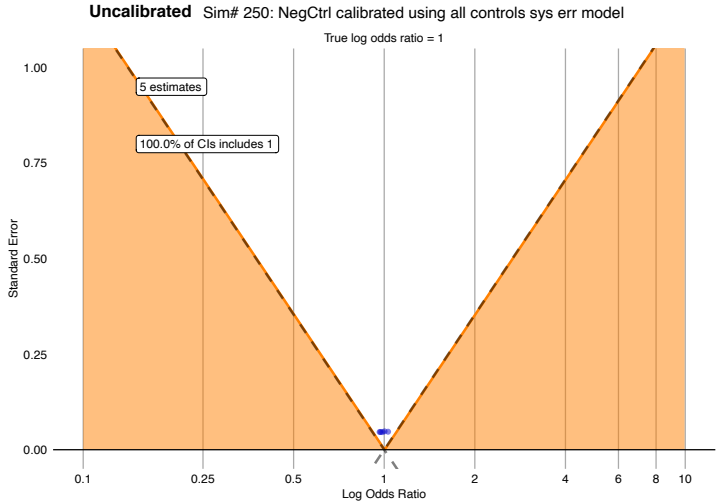
3.3.4 Calibration of Positive Controls (30)

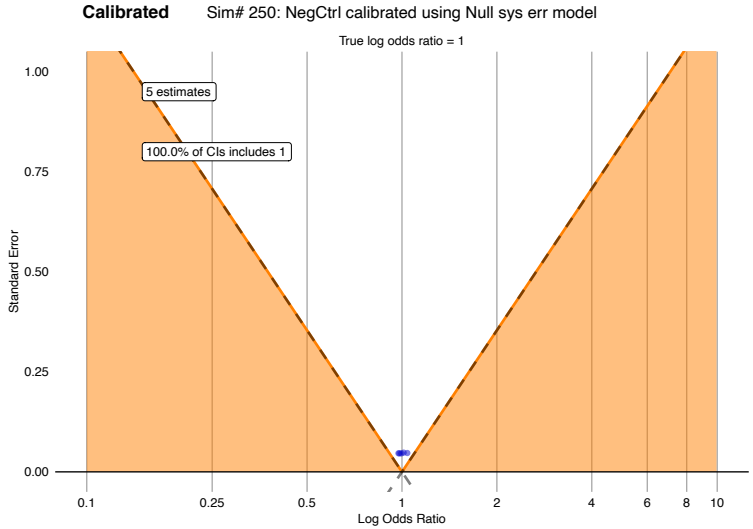


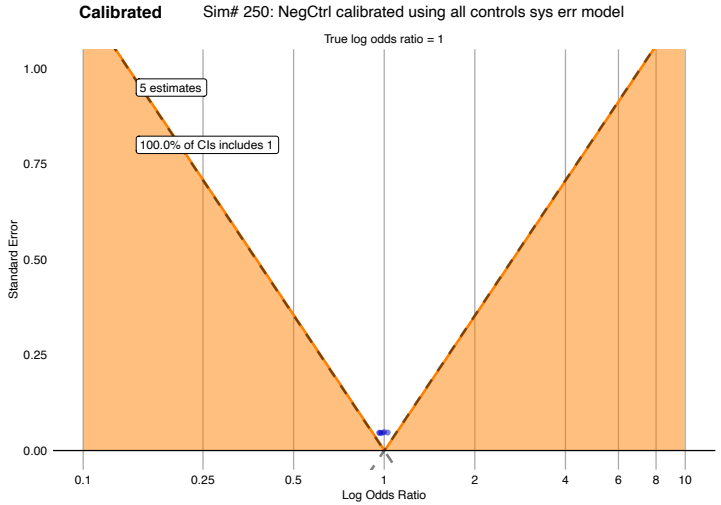
3.4 Lack of positivity scenario

3.4.1 Calibration of Negative Controls (5)

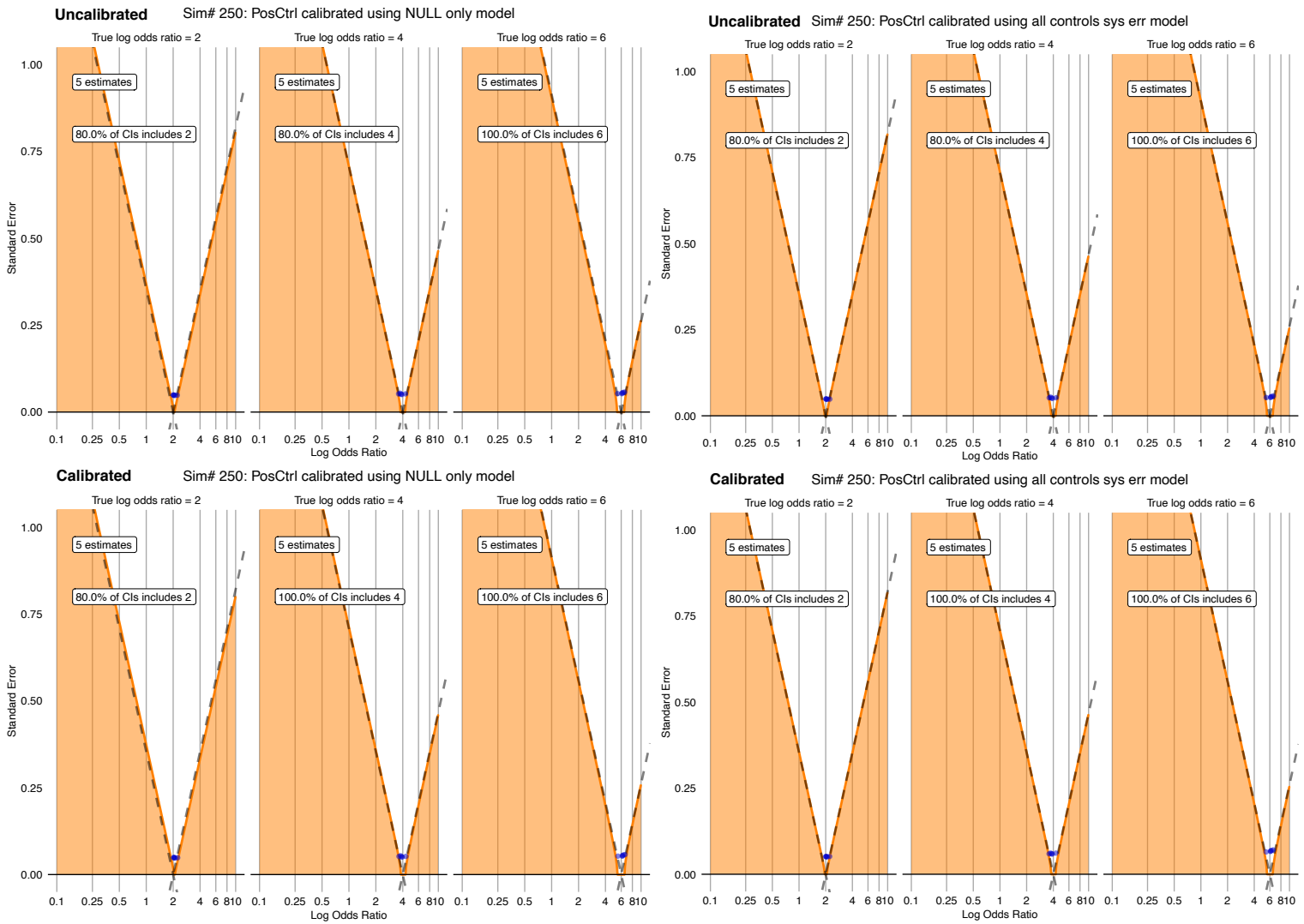




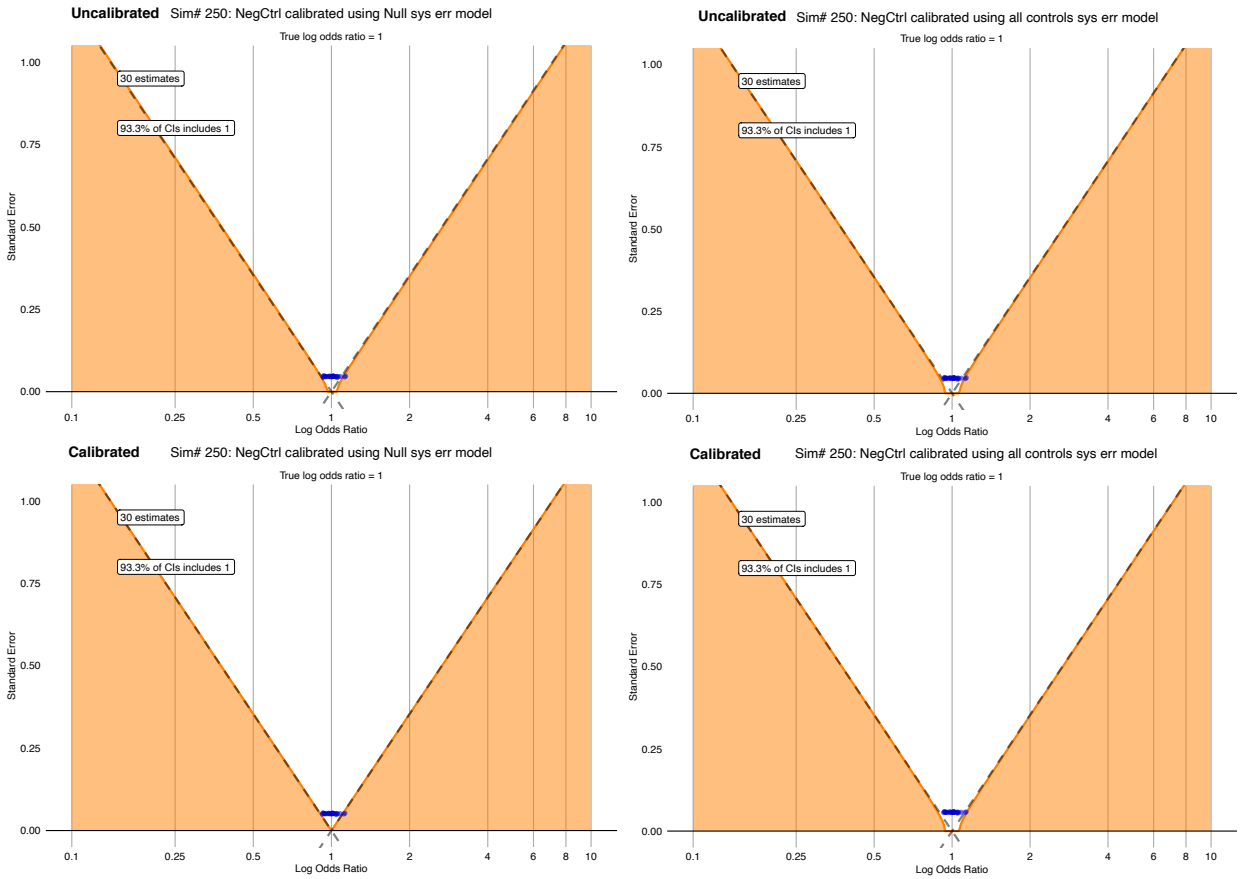




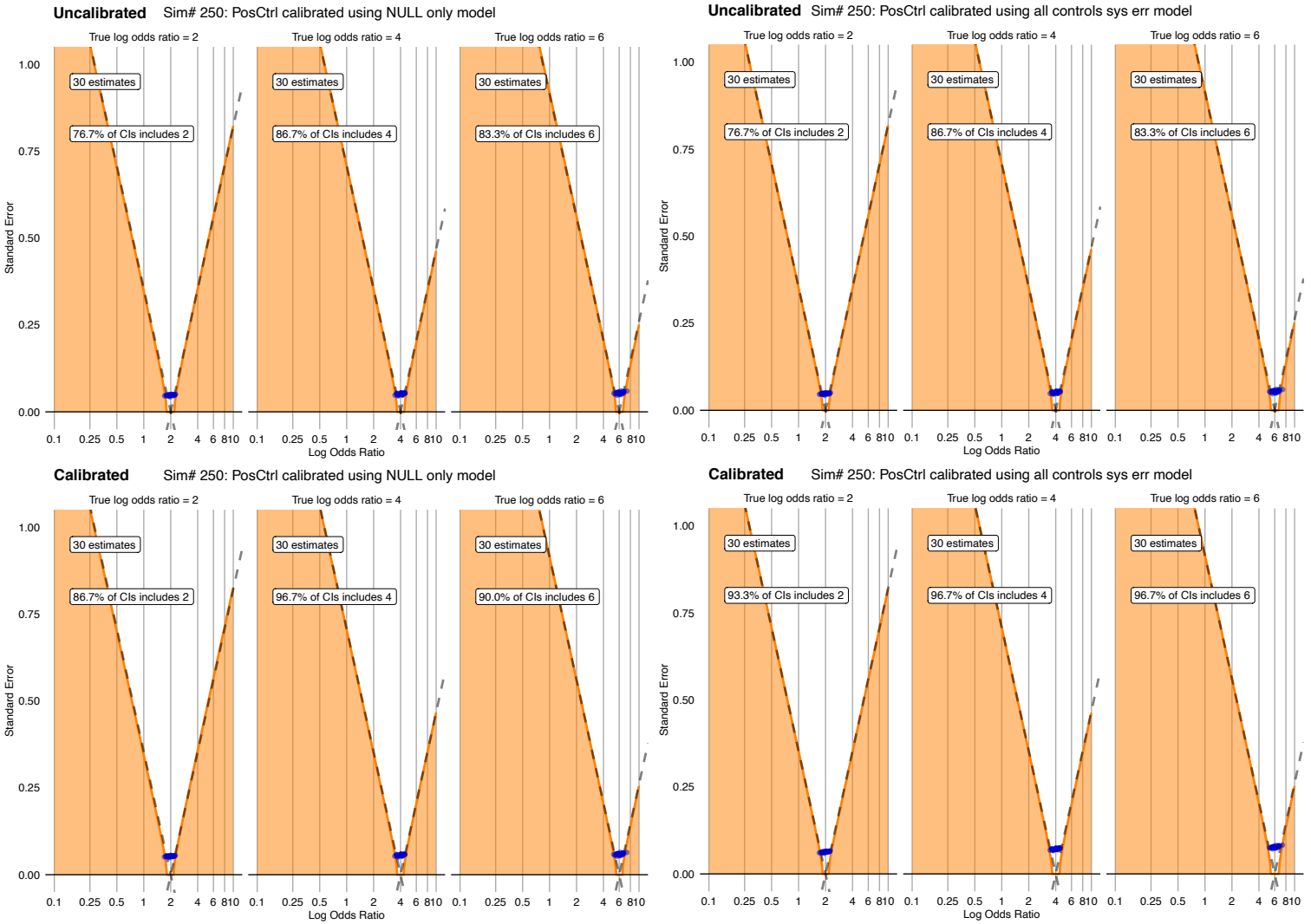
3.4.2 Calibration of Positive Controls (5)



3.4.3 Calibration of Negative Controls (30)

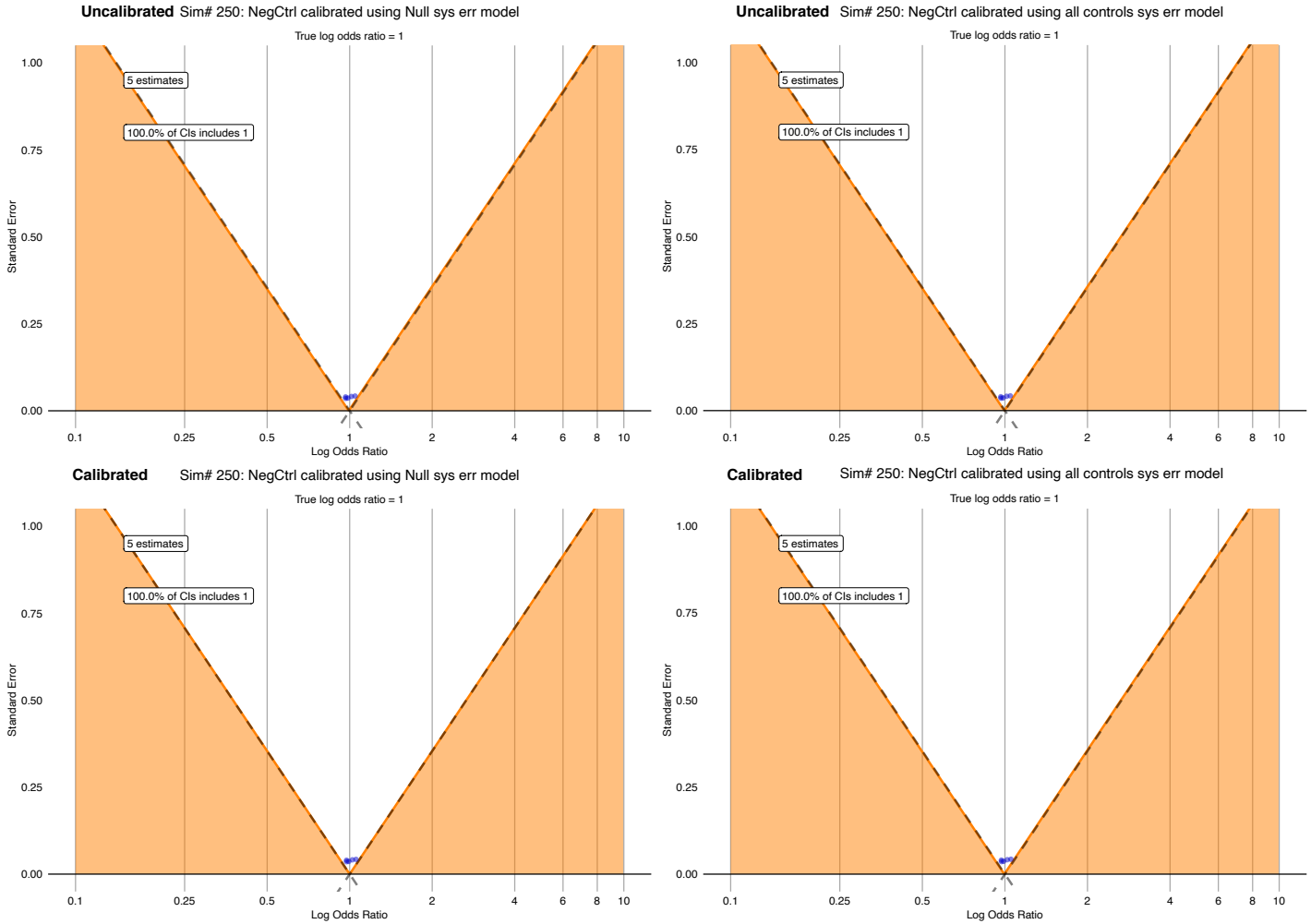


3.4.4 Calibration of Positive Controls (30)

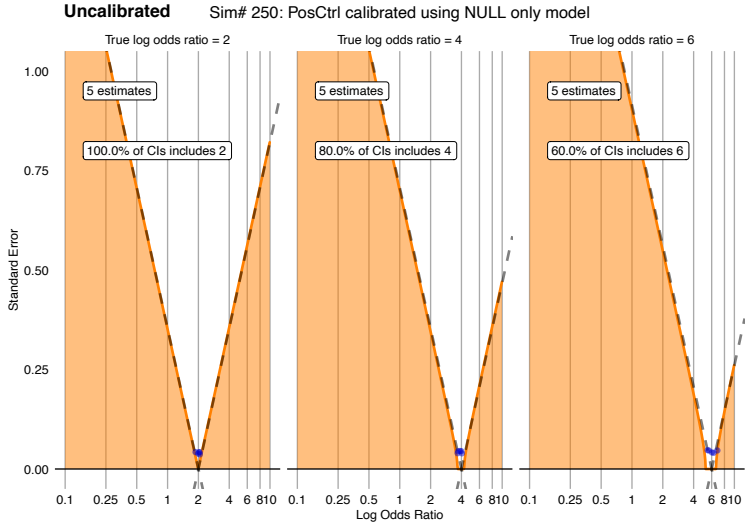


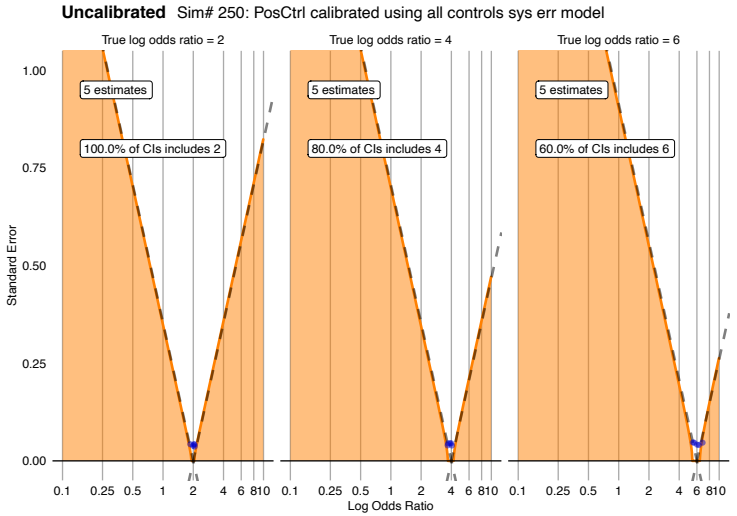
3.5 Measurement Error in confounder scenario

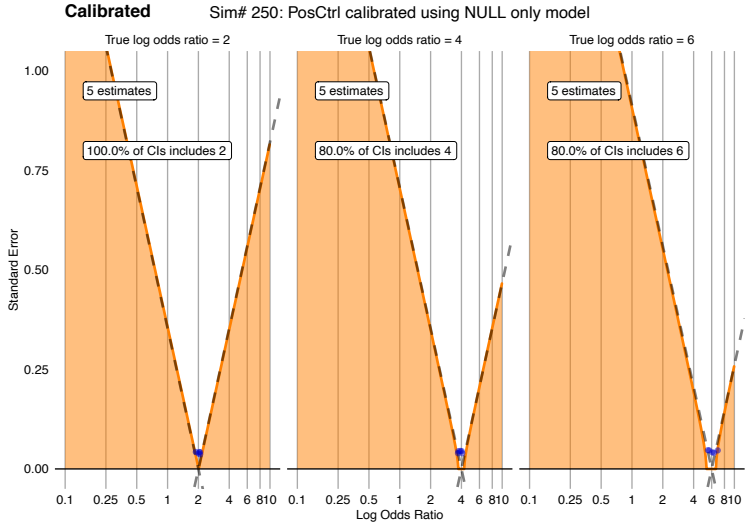
3.5.1 Calibration of Negative Controls (5)

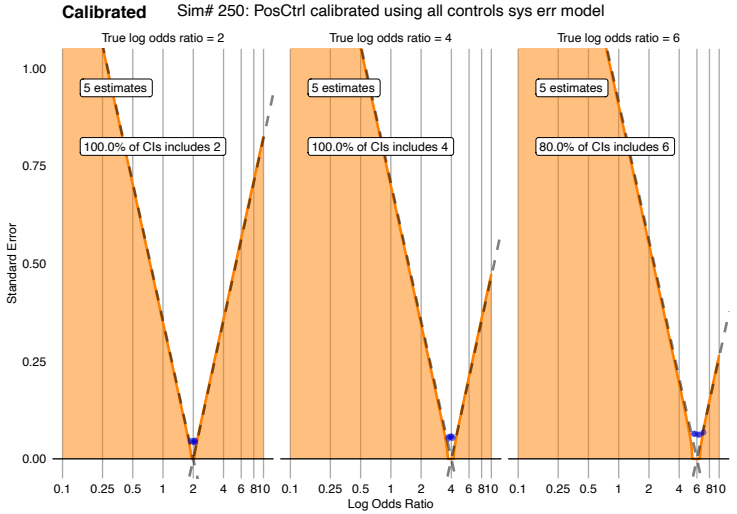


3.5.2 Calibration of Positive Controls (5)

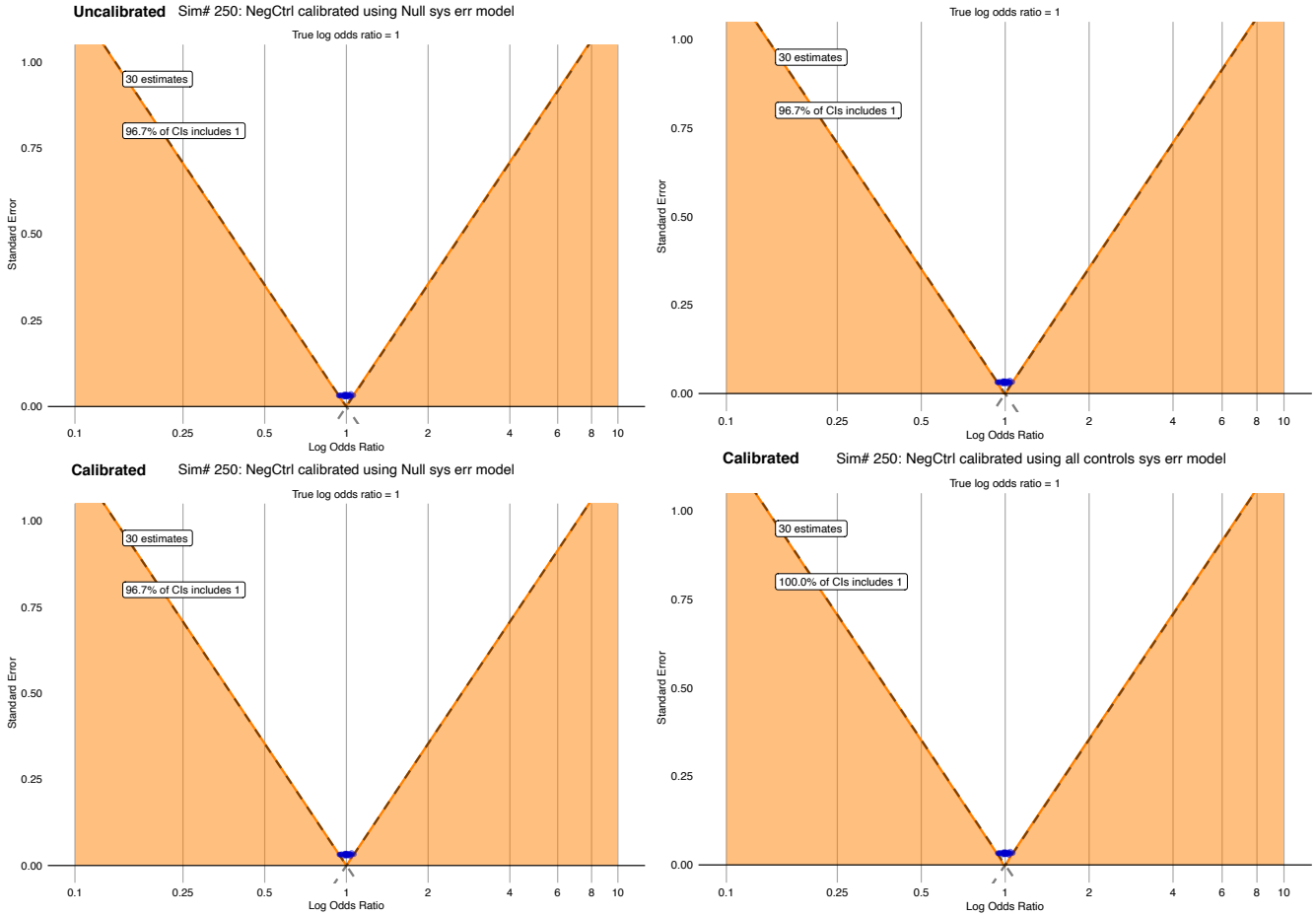








3.5.3 Calibration of Negative Controls (30)



3.5.4 Calibration of Positive Controls (30)