

Supplement

2023-03-28

Additional model results

Per our pre-registration, we ran a set of 6 models crossing 3 outcome measures (subjective replication score, whether the replication result was within the prediction interval of the original, and p-original on the hypothesis that both came from the same distribution) with 2 sets of predictors (with or without statistical predictors). These 6 models required 3 tiers of data: the subjective replication score without statistical predictors applies to all the data; the p_original and prediction interval models apply to the subset of data with numeric outcomes that can be compared; and the statistical predictor models need the smaller subset of data with p-values and original SMD in particular.

Due to low sample sizes and large numbers of predictors, even with regularizing priors, the coefficient estimates generally have a lot of uncertainty.

Sensitivity Analysis

As a check on whether our results were sensitive to the inclusion of pairs that were marginal in some way, we repeated the 6 models including only studies that were not marginal.

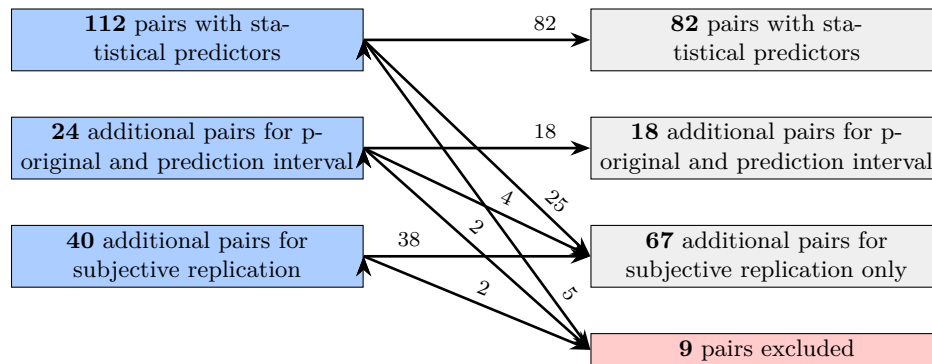


Figure 1: Diagram of what studies were downgraded or excluded for the sensitivity analysis.

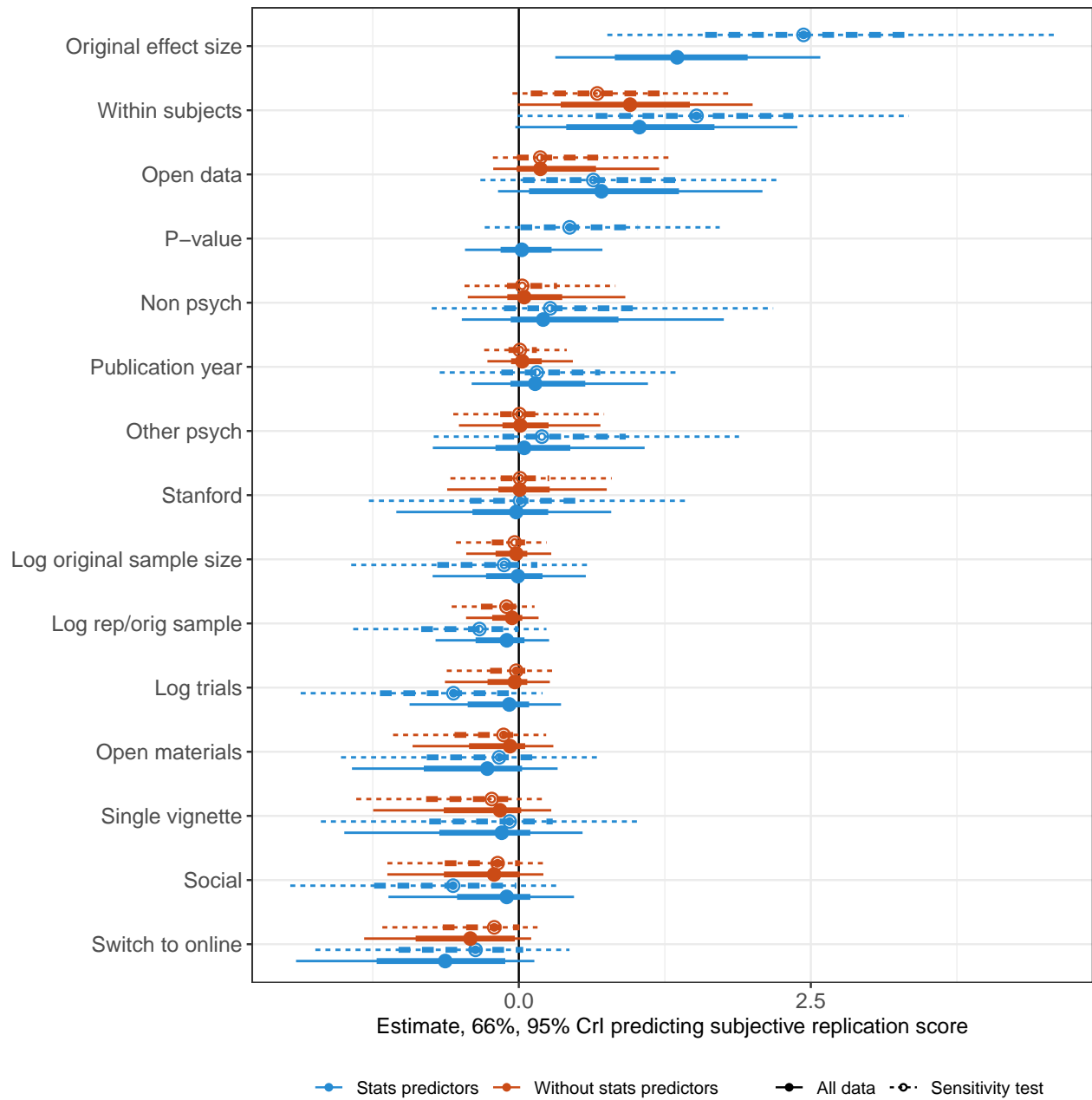


Figure 2: Coefficient estimates and uncertainty from ordinal models predicting subjective replication scores. Solid lines correspond to models run on as much of the data as possible; dashed lines are on the subset of the data for sensitivity analysis. Red is run on all relevant data with experimental predictors only; blue is on relevant data where there are statistical predictors.

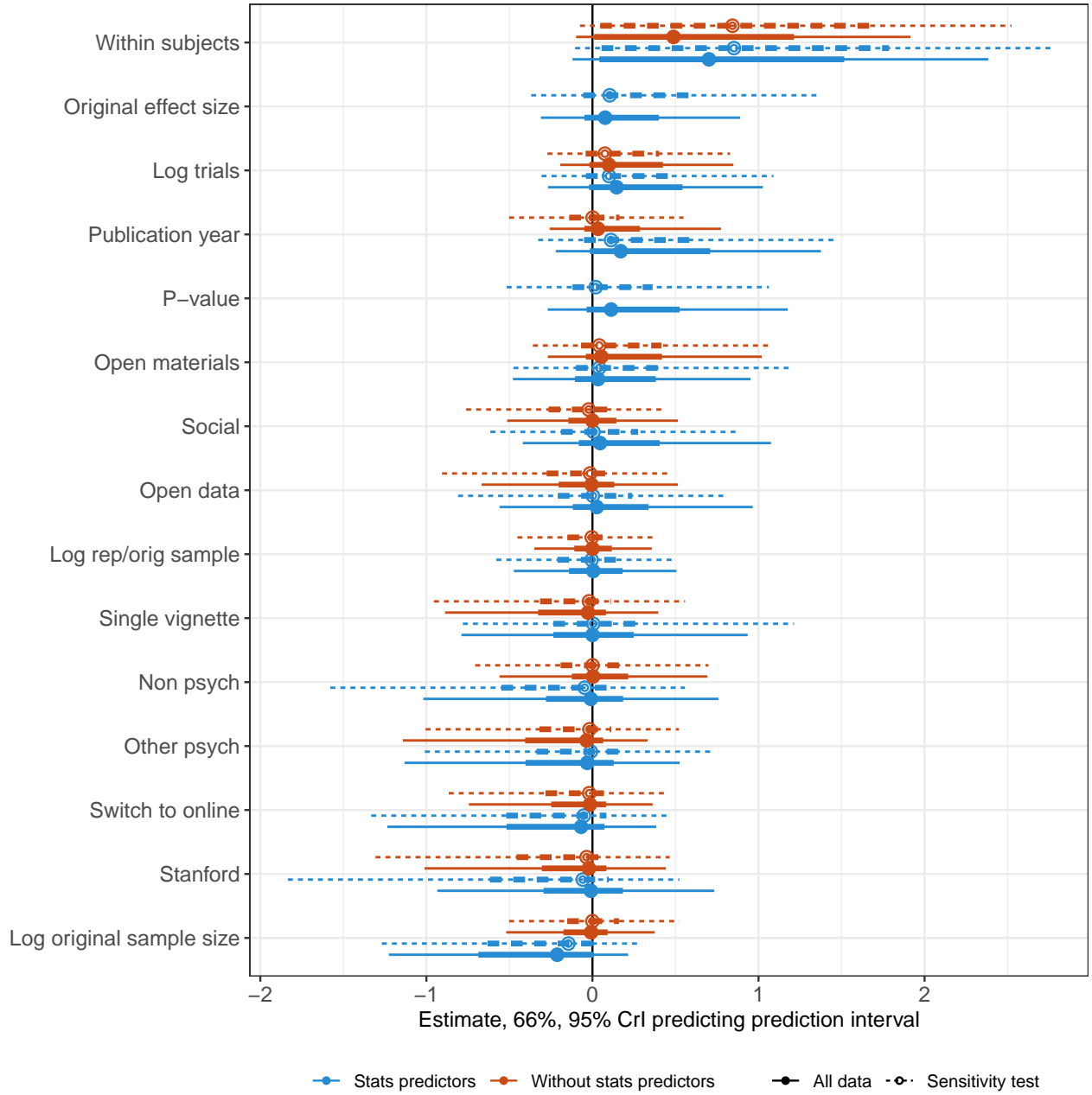


Figure 3: Coefficient estimates and uncertainty from logistic models predicting prediction intervals. Solid lines correspond to models run on as much of the data as possible; dashed lines are on the subset of the data for sensitivity analysis. Red is run on all relevant data with experimental predictors only; blue is on relevant data where there are statistical predictors.

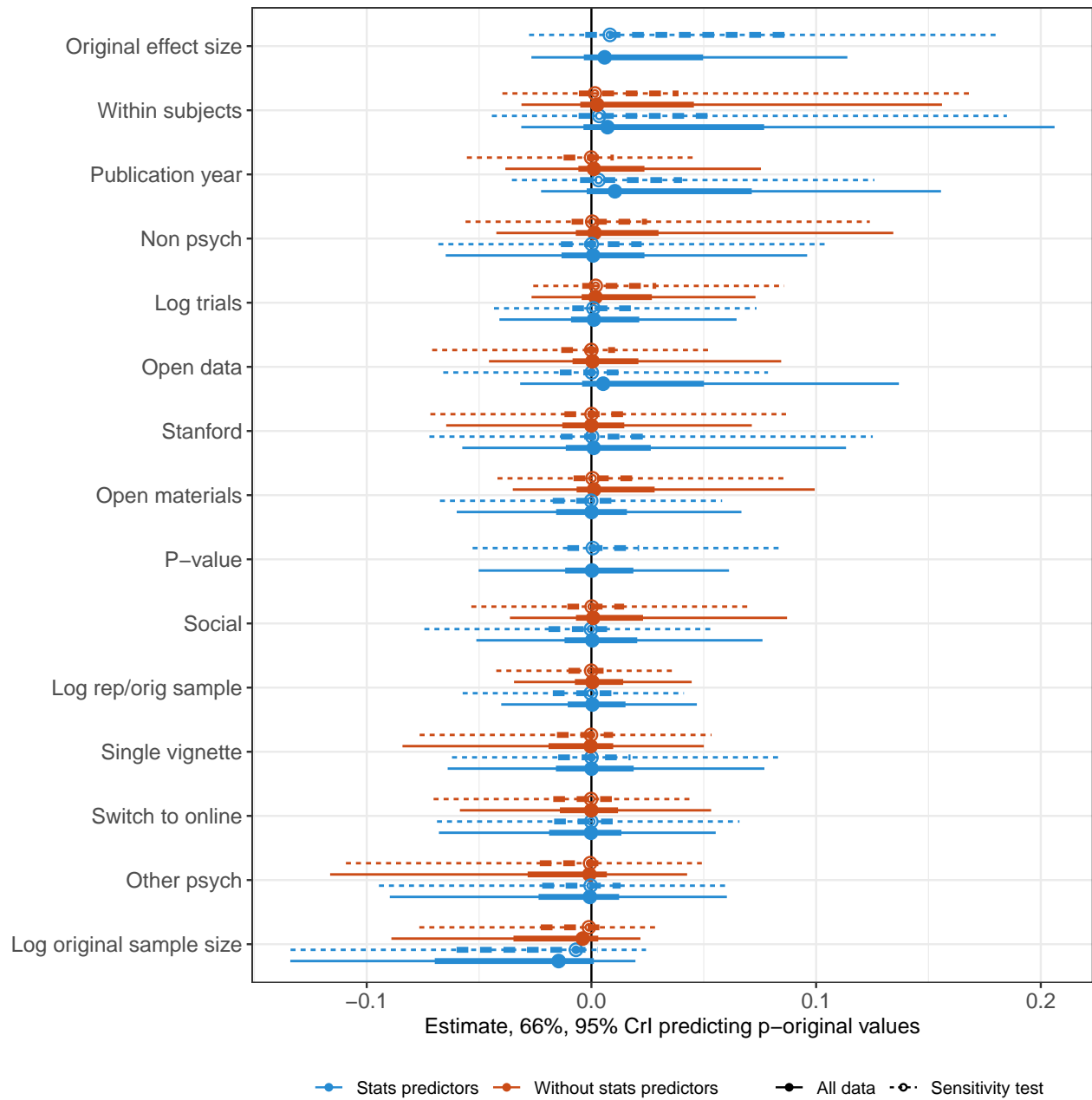


Figure 4: Coefficient estimates and uncertainty from linear models predicting p-original values. Solid lines correspond to models run on as much of the data as possible; dashed lines are on the subset of the data for sensitivity analysis. Red is run on all relevant data with experimental predictors only; blue is on relevant data where there are statistical predictors.