a) Sc 1.1 Identified 1 best treatment Identified 2 better treatments Identified 3 better treatments 0.05 0.04 0.03 0.02 0.01 0.00 4000 Sample Size b) Sc 2.1 Identified 1 best treatment Identified 2 better treatments Identified 3 better treatments 0.05 0.04 0.03 0.02 0.01 0.00 1000 1500 2000 5000 3000 Sample Size c) Sc 3.1 Identified 1 best treatment Identified 2 better treatments Identified 3 better treatments 0.05 ≗ 0.03 0.02 0.00 4000 500 1000 1500 2000 5000 5000 500 1000 1500 3000 5000 Sample Size d) Sc 3.3 Identified 1 best treatment Identified 2 better treatments Identified 3 better treatments 0.05 0.04 0.03 0.02 0.01 0.00 1000 1500 2000 5000 1000 1500 2000 3000 4000 1000 5000 500 Sample Size e) Sc 4.1 Identified 1 best treatment Identified 3 better treatments 0.04 0.03 0.02 0.01 0.00 1500 2000 3000 4000 5000 3000 4000 5000 4000 5000 Sample Size Fixed ···· Mixed ME Bayes URep1

Fig. SM 4 Probability of incorrect interval separation for five null scenarios, when using a fixed and mixed effects model using frequentist and Bayesian approaches for sample sizes ranging between N = 500-5,000.

ME Bayes URep2

FE Bayes URep2