Bias

At small sample sizes, FE Bayes unrepresentative 1 produced the largest absolute bias in many scenarios, however, ME Bayes unrepresentative 1 produced the largest absolute bias in scenario 2.2. The bias produced by all methods approached zero as the sample size increased for all scenarios (Figure SM 8).

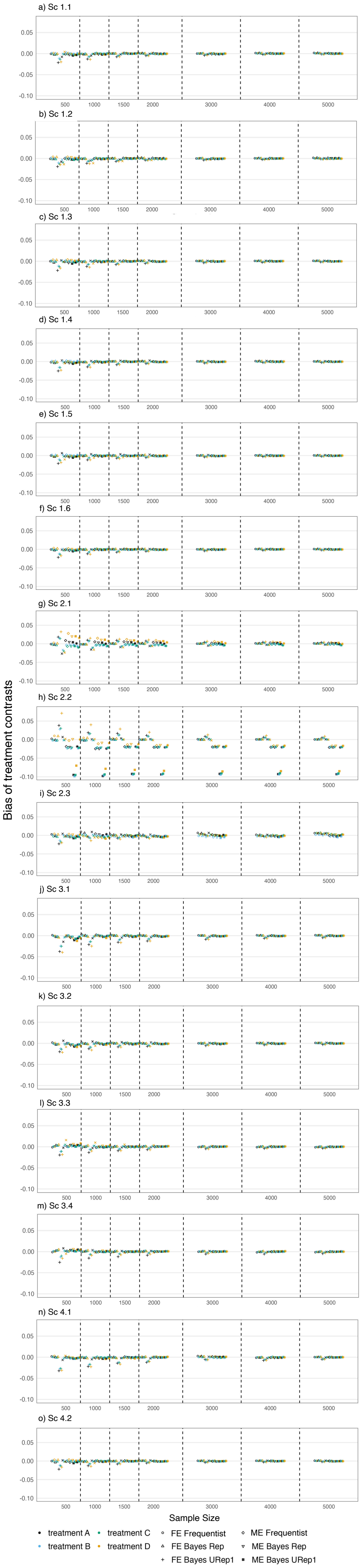

Fig. SM 8 Bias in each treatment (A-D) coefficient for all scenarios, when using a fixed and mixed effects model using frequentist and Bayesian approaches for sample sizes ranging between N = 500-5,000.

Coverage probability,

The coverage probability was around 80% for all FE methods for scenarios 1.1-2.2 & 3.1-4.2 and the ME frequentist method for scenarios 1.1-1.6 & 3.1-4.2 (Figure SM 9). It was often between 85%-90% for the ME Bayes methods for scenarios 1.1-1.6 & 3.1-4.2, where the treatments most common in each of the randomisation lists (treatments B and C) had the larger coverage probability. For scenarios 2.1-2.2, the ME frequentist and Bayesian methods gave a coverage probability approaching 100%. In scenario 2.3 the coverage probability was small for all methods (which decreased as the sample size increased).

a) Sc 1.1

b) Sc 1.2

c) Sc 1.3

d) Sc 1.4

e) Sc 1.5

f) Sc 1.6

g) Sc 2.1

h) Sc 2.2

i) Sc 2.3

j) Sc 3.1

k) Sc 3.2

l) Sc 3.3

m) Sc 3.4

n) Sc 4.1

o) Sc 4.2

Coverage probability

1.00
0.75
0.50
0.25
0.00

500 1000 1500 2000 3000 4000 5000

Sample Size

treatment A
treatment B
treatment C
treatment D
FE Frequentist
FE Bayes Rep
FE Bayes URep1
FE Bayes URep2
ME Frequentist
ME Bayes Rep
ME Bayes URep1
ME Bayes URep2

Fig. SM 9 Coverage probability of each treatment (A-D) coefficient 80% confidence/credible intervals for all scenarios, when using a fixed and mixed effects model using frequentist and Bayesian approaches for sample sizes ranging between N = 500-5,000.

Mean squared error

As expected, the MSE decreased as sample size increased, for all scenarios and methods (Figure SM 10). The four ME methods tended to have smaller MSE than the four FE methods, although the difference between the methods decreased as sample size increased. The treatments most common in each of the randomisation lists (treatments B and C) had the smaller MSE.

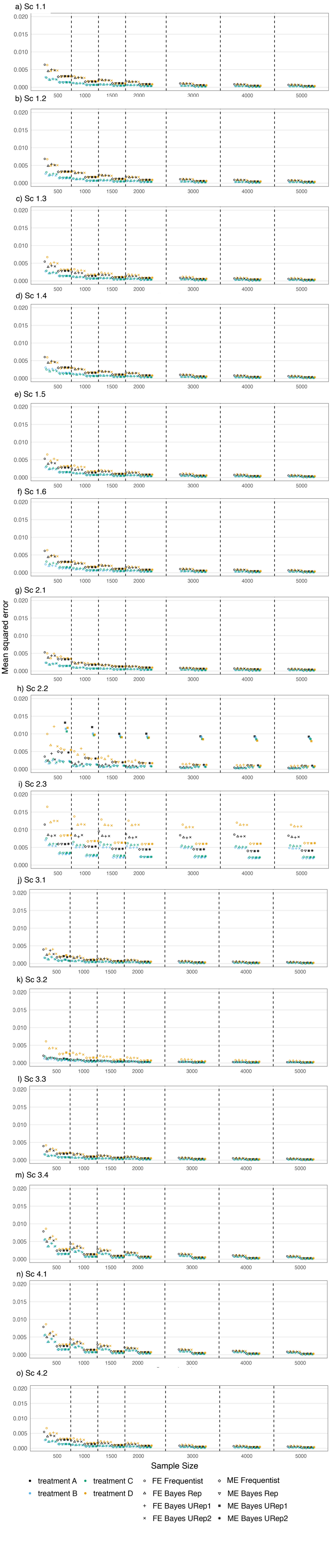

Fig. SM 10 Mean squared error in each treatment (A-D) coefficient for all scenarios, when using a fixed and mixed effects model using frequentist and Bayesian approaches for sample sizes ranging between N = 500-5,000.

Further to the analysis included in the main text and here, we investigated the FE method utilising non-informative and weakly informative priors, which produced similar results to the FE frequentist method, and hence, we omit these results from these supplementary materials.

ABBREVIATIONS

FE	Fixed effects
IORC	Improvement over random choice
ME	Mixed effects
MSE	Mean squared error
P_{best}	Best treatment probability
$P_{near\text{-}best}$	Near best treatment probability
P_{IS}	Probability of interval separation
P_{IIS}	Probability of incorrect interval separation
PRACTical	Personalised randomised controlled trial
RMR	Reduction in mortality ratio
RCT	Randomised controlled trial
SoC	Standard of care

REFERENCES

18 Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

14 Package ‘rstanarm’. https://cran.r-project.org/web/packages/rstanarm/rstanarm.pdf (accessed 14 Jul 2023).

7 Lee KM, Turner RM, Thwaites GE, *et al.* The Personalised Randomized Controlled Trial: Evaluation of a new trial design. *Statistics in Medicine* 2023;**42**:1156–70. doi:10.1002/sim.9663