Consequences of missing days on daily diary studies: A simulation study

Mustafa Yildiz[1], Maria Yefimova[2], Christopher D. Maxwell[3], Tami Sullivan[4], Carolyn E. Pickering[1]

[1]University of Alabama at Birmingham School of Nursing, [2]University of California, San Francisco, Department of Physiological Nursing, [3]Michigan State University, School of Criminal Justice, [4]Yale University, School of Medicine
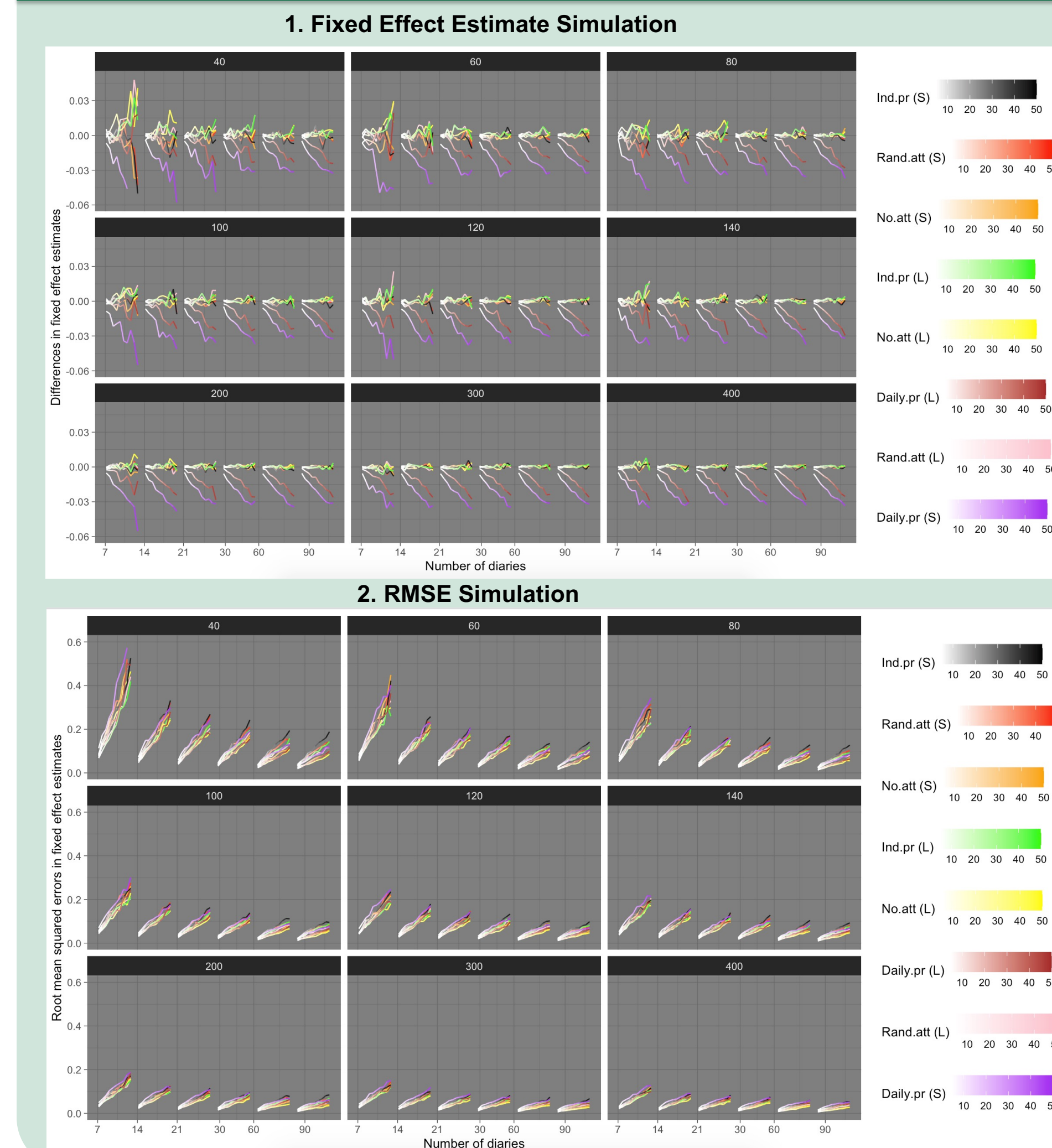
Motivation

Micro-longitudinal data collection, such as daily diaries are increasingly used in social sciences research.

Impact of missing data, both on the item-level and day-level, is unknown.

Simulation of various assumptions and missing data conditions can inform future study designs.

Aim

To investigate the impact of missing data on **fixed effects, standard errors, and power** in a multi-level model of a daily diary study using simulation.

Results

1. Fixed Effect Estimate Simulation

2. RMSE Simulation

3. Bias of Standard Errors Simulation

4. Power Simulation

Main Findings

✓ **Bias of fixed effects** was small across all simulation conditions except when missing diaries were predicted by day level measures

✓ **RMSE for fixed effects** increase as number of diaries and sample size decrease and proportion of missing diaries increases.

✓ **Standard errors** for the diaries with missing days are consistently overestimated under all conditions.

✓ The strongest predictor of **power** was effect size.

LEGEND: Ind.pr= person level aggregated predictors, Daily.pr= Day level predictors, Rand.att= Random attrition, No.att=No attrition, (S)=Small effect size, (L)=Large effect size
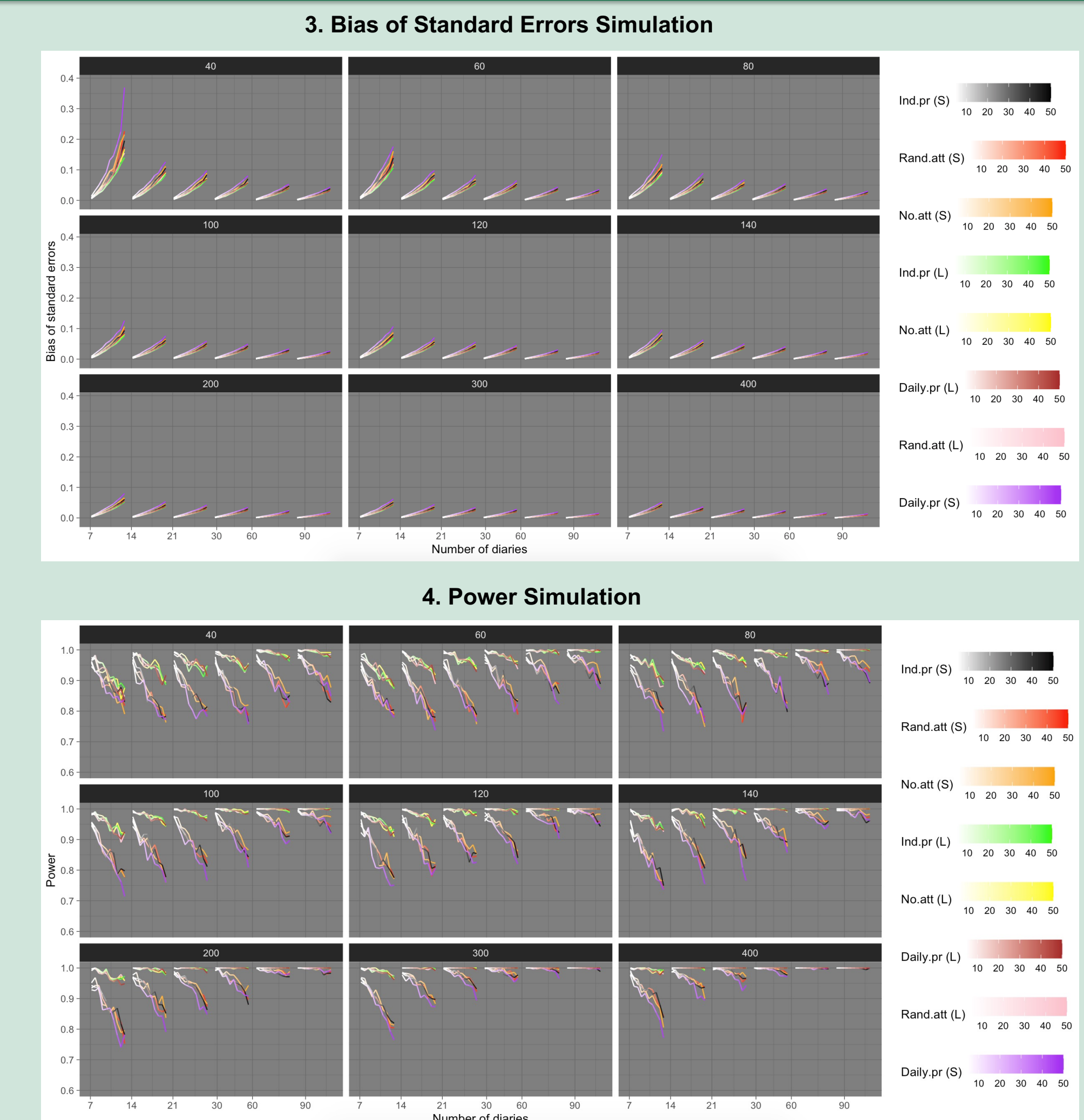
Methods

Simulation conditions: sample size (30 to 400), number of days (7 to 90), predictors of missing days (no attrition, random attrition, person level predictors, day level predictors), effect size (small and large), and proportion of missing days (5% to 50%).

Predictors of missing days condition: (i) random days missing without attrition, (ii) random days missing with attrition, (iii) attrition is predicted by day level aggregate scores, and (iv) attrition is predicted by day level predictors.

Outcome variables: bias and RMSE for fixed effects, power, and differences in standard errors.

Simulation from a multivariate normal distribution with a predetermined correlation matrix for a binary outcome given a set of fixed effect parameters for the small effect size condition (0.35, -0.5, 0.37, 0.79, 0.24, -0.70) and large effect size condition (1.42, 0.06, 1.45, 2.2, 1.27, 0.50). Fifty replications were run for each of 1080 conditions.

Conclusion and Implications

• High levels of missing diaries can be tolerated by increasing sample size and/or increasing number of diaries.

• Studies with small effect sizes are more likely to find a null value compared to the studies with large effect sizes when there was a true effect.

• Differences in standard errors were greatly reduced by increasing sample size and number of diaries.

• Both number of diaries and sample size contribute to the quality of estimation. Under estimation of fixed effects may occur if the day level variables predict missing days.

UAB NURSING
The UAB School of Nursing