

**Supplementary materials:**

*Frequency of missing data in MLTD :*

Table 1 shows the frequency of missing data across dayparts and time (only morning 1, afternoon 1 and evening 1). The distribution of missing data for the other dayparts is comparable and not displayed here.

Part of Day	time	Percentage missing		
		Activity only	Mood only	Both
	1	1	4	1
	2	0	9	1
	3	1	14	2
	4	0	14	7
	5	1	13	6
	6	1	13	3
Morning 1	7	4	5	6
	8	2	2	7
	9	2	8	10
	10	5	5	5
	11	5	8	4
	12	4	8	6
	1	3	11	12
	2	7	12	13
	3	2	11	8
	4	3	5	4
	5	4	7	4
	6	2	9	8
Afternoon 1	7	7	12	8
	8	2	10	10
	9	2	10	9
	10	4	11	11
	11	8	12	10
	12	5	9	13
	1	3	8	8
	2	3	11	4
	3	4	5	7
	4	3	4	5
	5	1	1	5
	6	3	4	6
Evening 1	7	3	5	11
	8	4	7	6
	9	3	4	9

	10	4	8	11
	11	1	8	10
	12	0	1	7

Table 1 of supplementary material. Percentage missing of Activity and Mood

From Table 1, it can be seen that the variable Mood and activity contain 5-25% and 1-18% missing observations across time (and dayparts), respectively. The variable LSW, however, does not contain any missing observations.

*Classification of missing data :*

1. *Missing completely at random (MCAR):* the missing data mechanism depends neither on observed nor on unobserved values. In the case study, this may occur when the probability of a missing observation for an arbitrary person does not depend on his/her mood nor on whether he/she is (not) engaged in some activities or the fact that he/she is (not) living in a large-scale ward. On top of these, the missing observation does not depend on any other variable (observed or unobserved) not considered in the analysis. Consequently, we can assume that the chance of recording (observing) the mood's score for all participants (i.e., available and unavailable participants) is the same, which makes the missing data mechanism MCAR, and the complete cases can be considered as a random sample of the target population.
2. *Missing at random (MAR):* the missing data mechanism depends only on the observed values (and not on the unobserved values). If, for example, it is more probable that the mood's score of a participant is missing when he/she is not participating in an activity, then the activity is an observed predictor of the probability of observing the mood's score. MAR will occur if conditional on activity, the missing data mechanism is MCAR. It should be noted that information of the activity (by means of the variable Activity) should be included in the analysis as a predictor of Mood.

3. *Missing not at random (MNAR)*: the missing data mechanism depends on the unobserved values (and perhaps also on observed values). Suppose participants with positive mood's scores are more likely to have missing observations on the mood's score. Consequently, there will be an under-representation of the estimated number of participants with an observed positive mood than there really are.

*Details of simulation study :*

We considered a longitudinal design with three time points. First, correlated binary variables  $X_1$ ,  $X_2$  and  $X_3$  were generated<sup>1</sup> with equal marginal probabilities (i.e.,  $P(X_1 = 1) = P(X_2 = 1) = P(X_3 = 1) = 0.5$ ) and equal correlations (i.e.,  $\text{cor}(X_1, X_2) = \text{cor}(X_1, X_3) = \text{cor}(X_2, X_3) = .5$ ). The outcome  $Y$  was then generated according to the random intercept model

$$Y_{it} = 2 + 0.5X_{it} + u_i + \varepsilon_{it},$$

for subject  $i = 1, \dots, 115$  and time point  $t = 1, 2, 3$ . The random intercept  $u_i$  follows a normal distribution with mean zero and variance 0.12 and the residual  $\varepsilon_{it}$  follows a multivariate (3-variate) normal distribution with mean zero, variance one and a compound symmetric correlation structure (with correlation equal to 0.6). This covariance structure implies that the outcome variables  $Y_1$ ,  $Y_2$  and  $Y_3$  are correlated.

Four different settings representing different types of missing data mechanism and different patterns of missing data were considered in the simulation study. Initially, we assumed the outcome  $Y$  and independent variable  $X$  were fully observed at the first time point (i.e.,  $Y_1$  and  $X_1$  were complete) while these variables at the other time points (i.e., time points 2 and 3) can have missing values.

- a. *Scenario 1*: Missing observations in both outcome and independent variables.

The outcomes  $Y_2$  or  $Y_3$  (or both) were missing, each with a constant probability 0.3. Similarly, the independent variables  $X_2$  or  $X_3$  (or both) were missing with the same constant probability. This represents the MCAR mechanism (i.e., the probability of observing either variable was the same for all cases/participants). In total, 50% of the

---

<sup>1</sup> The R package 'bindata' was used to generate correlated binary data.

cases was incomplete, and the outcome and independent variables were never jointly missing.

b. *Scenario 2: Missing observations in the outcome.*

The outcomes  $Y_2$  or  $Y_3$  (or both) were missing.  $Y_2$  was missing if  $Y_1 \leq \bar{Y}_1$ ;  $Y_3$  was missing if  $Y_2 \leq \bar{Y}_2$ . The missing data mechanism is MAR because the probability of missingness for  $Y_2$  depends only on observed values of  $Y_1$ . Likewise, the probability of missingness for  $Y_3$  depends only on observed values of  $Y_2$ . Approximately 50 % of the outcome variables was incomplete.

c. *Scenario 3: Missing observations in the independent variable.*

The independent variables  $X_2$  or  $X_3$  (or both) were missing.  $X_2$  was missing if  $Y_2$  was smaller than or equal to its first quartile; Likewise,  $X_3$  was missing if  $Y_3$  was smaller than or equal to its first quartile. This is a comparable MAR mechanism as in b, and approximately 40% of independent variables was incomplete.

d. *Scenario 4: Missing observations in both outcome and independent variables.*

Missing values on  $Y_2$  or  $Y_3$  were created as in scenario 2, or missing values on  $X_2$  or  $X_3$  were created as in scenario 3 (but not both). This is also a comparable MAR mechanism as in b or c where both outcome and independent variables are incomplete, and approximately 50% of cases were incomplete.