

Electronic Supplementary Materials

We here present the scatterplots of one replication in the most severe missing data condition of each simulation study. The bias in person parameter estimates was rather similar across the whole range of person ability level. This is also true for the other simulation conditions not depicted here.

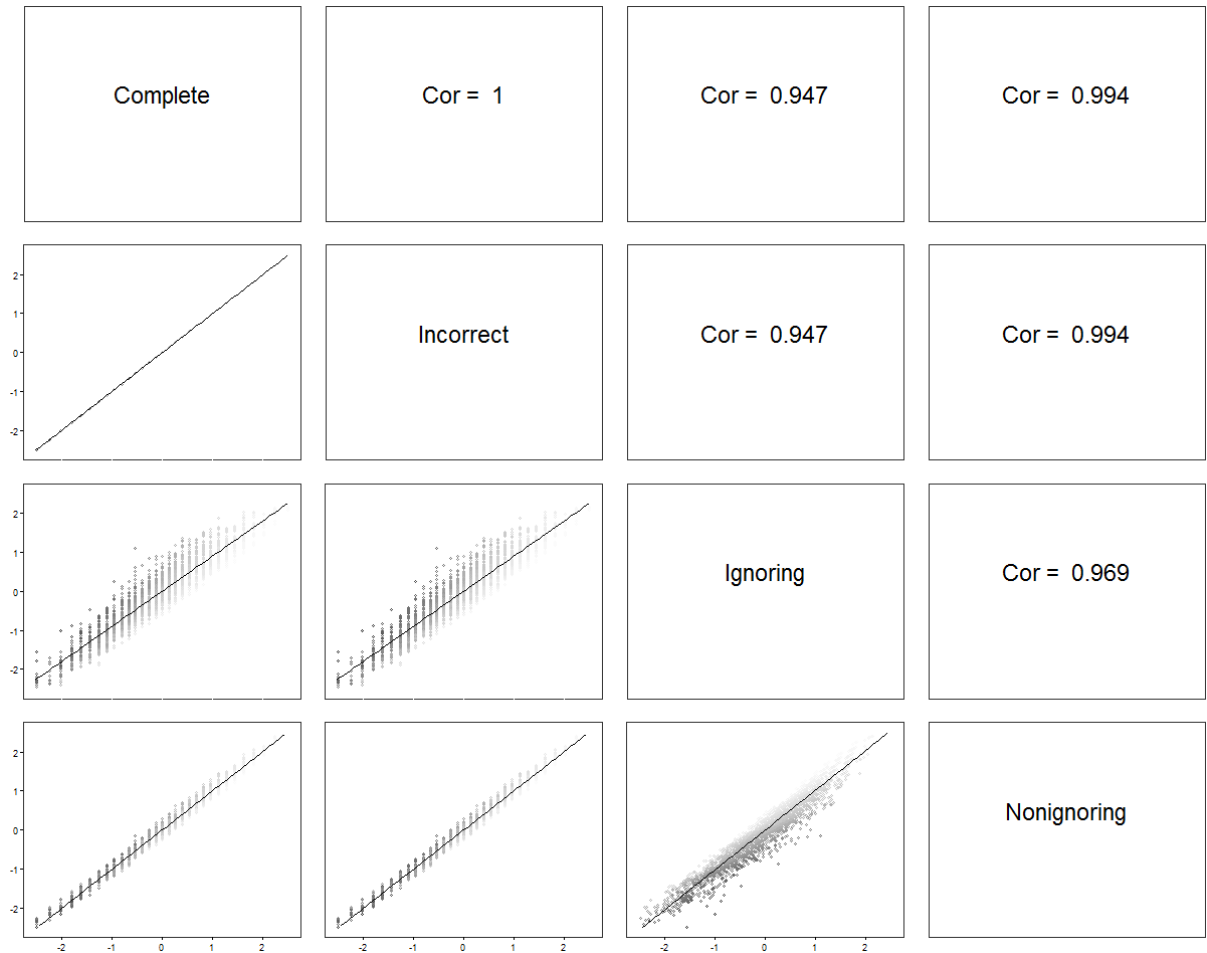


Figure S1. EAP-estimates of person ability in the strong MNAR condition with 30% missing values in simulation 1. Darker points represent more missing values. Note that for this graph the average bias was set to zero.

Figure S1 depicts the individual person parameter estimates for one replication in the most extreme nonignorable condition of 30% missing responses under strong MNAR. The figure shows the scatterplot and the correlation of the individual person parameter estimates for the complete

data analyses (as the reference) and each of the three analysis approaches on the incomplete data. As expected, in this data generation condition, incorrect scoring resulted in the same estimates as in the complete data analysis. When ignoring missing values, the ability of persons with many missing values were overestimated. Interestingly the approach for nonignorable missing values showed a similar effect, however, with a much smaller range of bias for individual persons. Thus, the rank order of persons was rather well preserved in the approach for nonignorable missing values.

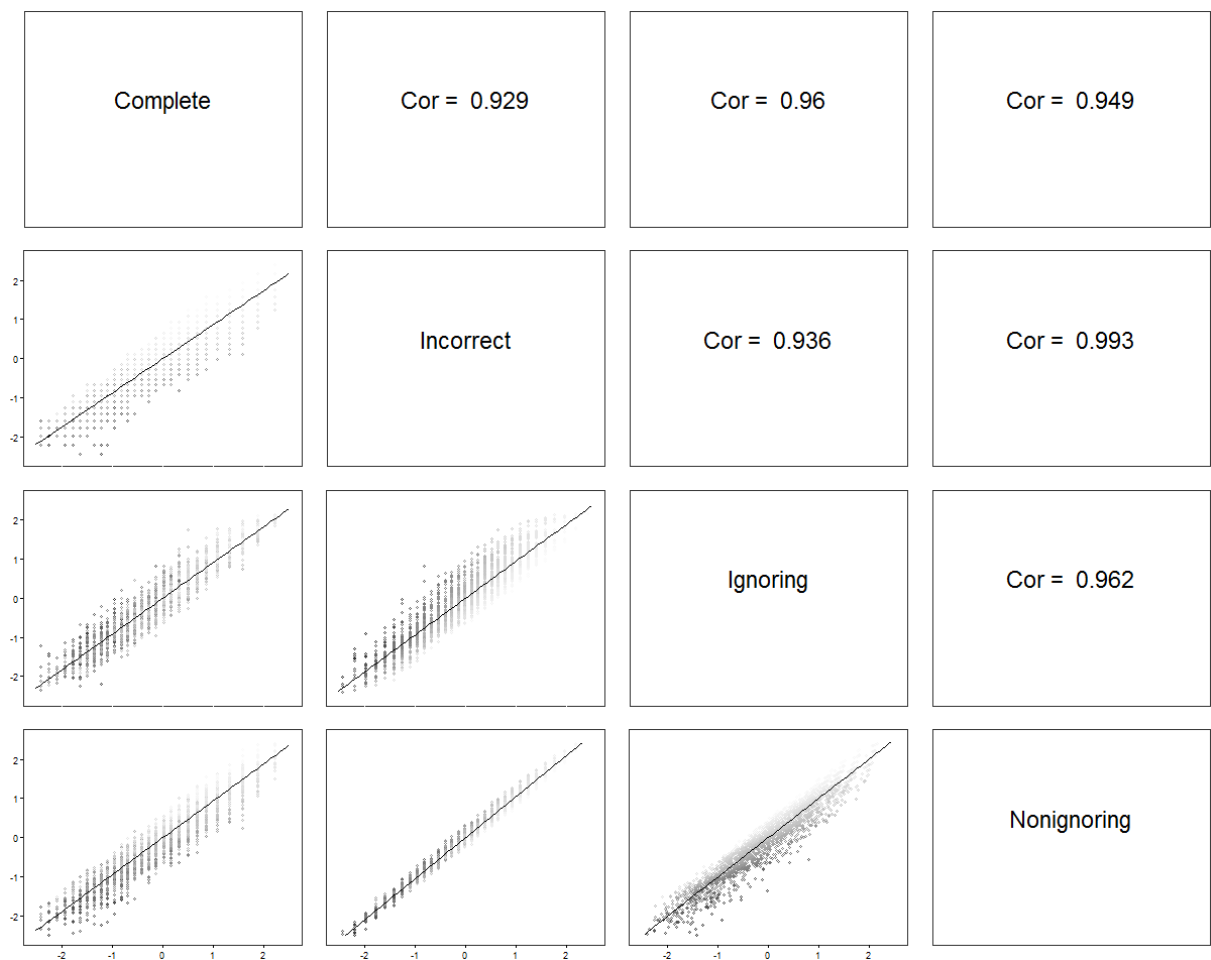


Figure S2. EAP-estimates of person ability using different analyses methods in simulation

2. Darker points represent persons with more missing values. Note that for this graph the average bias was set to zero.

Individual person parameter estimates for one replication in the complete case as well as for all three analyses in the incomplete case are depicted in Figure S2. The relative differences in person parameter estimates between the three approaches was about the same as in simulation 1. Again, incorrect scoring and nonignoring resulted in a rather similar rank order of persons, the only difference is a general shift in average ability estimate (see Table 4). The relationship of the person parameter estimates of the three approaches to the complete data case differs between simulation 1 and 2.

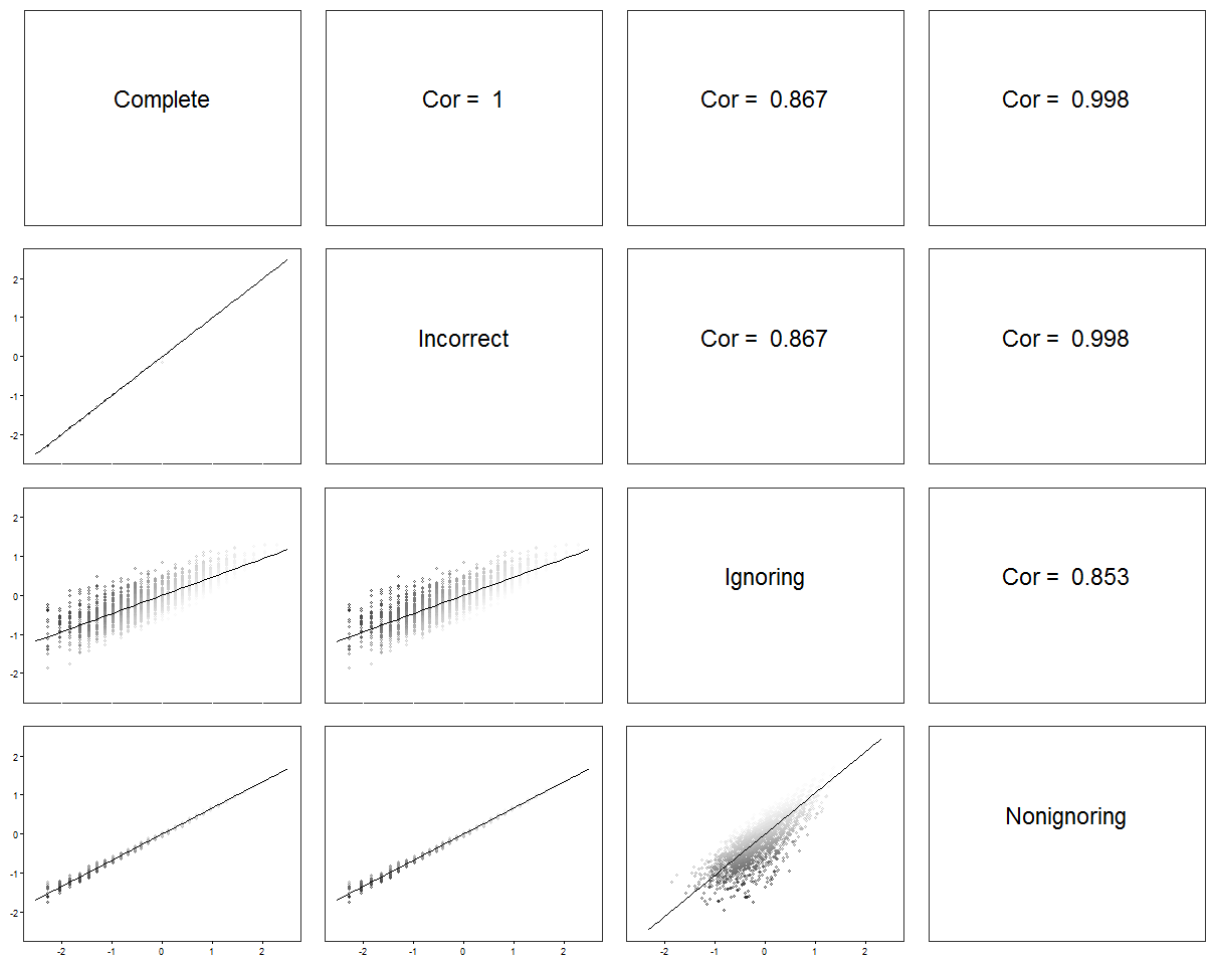


Figure S3. EAP-estimates of person ability in the condition with 30% missing values and strong impact of the item response in simulation 3. Darker points represent more missing values. Note that for this graph the average bias was set to zero.

Individual person parameter estimates for one replication in the most extreme condition of 30% missing responses and strong impact of item responses on the missing probability in simulation 3 are depicted in Figure S3. The pattern of differences across the three approaches is similar to the pattern in simulation 1. Again, incorrect scoring and nonignoring result in a rather similar rank order. The differences in rank order are rather small and mainly occur for persons with many missing values.