Supplment Four: S4 Multiple Imputation Analytic Strategy, Results, and Discussion

**Analytic Strategy**

Modern missing data analyses of multiple imputation and full information maximum likelihood were utilised. Both techniques require less stringent assumptions than complete case analysis, as they are able to retrieve unbiased results in the presence of more complex missing data patterns (Graham, 2009). Complete case analysis can be unbiased, but only when missingness is in its most ‘simple form’, completely at random (MCAR). For example, if a computer virus deletes records in a completely random fashion. However, multiple imputation and full information maximum likelihood are unbiased even when there is systematic bias underlying missing data, provided that variables in the dataset are (i) completely able to predict the mechanisms of missingness, or (ii) the variable with missing data itself (Pedersen et al., 2017; Enders, 2017, Graham, 2009). This is formally known as missing at random (MAR), which is a requirement for multiple imputation and full information maximum likelihood to be unbiased (Enders, 2017; Graham et al., 2009).

Here, multiple imputation was implemented in SPSS v 25. Following current recommendations (Pedersen et al., 2017; Azur, Stuart, Frangakis, Leaf, 2011) that a larger number of datasets yield more precise parameter estimation, 20 datasets were generated. Moreover, as inhibitory control tasks showed sizeable skew and kurtosis with noticeable departure from a normal distribution (Appendix 4 Table 1), predictive mean matching was utilised. PMM fills missing values with values from observed cases, selected at random from a pool of cases which have values close to regression based predicted values. In this way, PMM is a semi-parametric procedure and is expected to relax assumptions of normality required for typical entirely regression based forms of multiple imputation. Also consistent with recommendations (Pedersen et al., 2017; Azur et al., 2011) was the addition of auxiliary variables not in the main analyses. This was done by including variables that were conceptually already considered to have an expectable association with task performance, so as to increase the likelihood of meeting MAR assumptions. This included 5 variables considered as possible covariates in the main analyses; gender, maternal age, child’s actual and post conception age, and maternal exposure. Also, given correlations between aspects of executive functioning, including working memory (e.g., Miyake et al 2000), all 6 task variables were considered simultaneously to further allow the meeting of MAR assumptions (i.e. NSE and SE variants of relational memory, cognitive flexibility and inhibitory control).

Full information maximum likelihood was implemented in Mplus v 7.4 by *removing* the syntax “listwise = on” from all models in the main analyses. This resulted in relational memory and inhibitory control models estimated with MAR assumption, as these continuous variables are analysed by maximum likelihood. The cognitive flexibility task of DCCS yields a ‘pass’ or ‘fail’ binary variable, and are analysed by weighted least squares means and variance. As this model does not utilise maximum likelihood, it is not as robust in handling missing data, only utilising information from predictors of maternal sensitivity and SES but not SE and NSE versions of the task, to inform missing data (known as MARX ; Asparouhov & Muthen, 2010)

**Results**

Below we present a side by side comparison of the descriptives for complete and multiply imputed data (Supplementary Table 2) as well as the beta’s and path models (Supplementary Table 3) for analyses based upon the complete cases and imputed data sets.

Appendix 4 Table 1. Descriptives for Complete Cases and Multiply Imputed Data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Complete Case Analysis | | | | Multiply Imputed Analysis | | | | Difference (%) | |
| Study Variable | M | SD | N | Skew/  Kurtosis | Pooled M | Pooled SD | N | Range Skew/  Kurtosis | M | SD |
| Sensitivity | 0.22 | 0.43 | 401 | -0.36  -1.08 | 0.221 | 0.43 | 401 | -0.36 to -0.36  -1.08 to -1.08 | 0.00% | 0.00% |
| SES | 3.36 | 0.91 | 394 | -0.23  -0.87 | 3.36 | 0.91 | 401 | -0.25 to -0.22  -0.90 to -0.86 | 0.00% | 0.01% |
| Maternal Age | 30.72 | 5.29 | 401 | 0.08  -0.53 | 30.72 | 5.29 | 401 | 0.08 to 0.08  -0.53 to -0.53 | 0.00% | 0.00% |
| Age at test | 1257.76 | 28.25 | 327 | 1.34  4.55 | 12567.01 | 27.98 | 401 | 1.15 to 1.43  4.00 to 5.01 | 0.06% | -0.94% |
| Postconceptual Age at Test (Age plus Gestational Age) | 1529.26 | 28.23 | 327 | 1.27  4.67 | 1528.44 | 27.88 | 401 | 1.15 to 1.44  4.14 to 5.26 | 0.05% | -1.22% |
| Maternal Exposure | 203/ 54.6% |  | 372 |  | 217.6/ 54.3% |  | 401 |  | -0. 3% |  |
| Gender (%Male) | 219  /54.6% |  | 401 |  | 219  /54.6% |  | 401 |  | 0.00% |  |
| Relational Memory SE | 2.68 | 1.00 | 239 | 0.10  0.00 | 2.19 | 1.40 | 401 | -0.32 to 0.68  -1.29 to -0.31 | 18.4% | 40.13% |
| Relational Memory NSE | 2.93 | 1.02 | 243 | -0.03  -0.14 | 2.61 | 1.42 | 401 | -0.64 to 0.44  -1.35 to 0.11 | 11.12% | 39.85% |
| DCCS SE (Pass) | 58  /24.3% |  | 239 |  | 104.2  /26.0% |  | 401 |  | 1.71% |  |
| DCCS NSE (Pass) | 62  /26.7% |  | 232 |  | 105.4  /26.3% |  | 401 |  | -0.45% |  |
| Inhibitory  Control SE | 7.86 | 1.29 | 273 | -2.05  6.12 | 7.69 | 1.42 | 401 | -2.30 to -1.62  2.89 to 7.19 | 2.12% | 10.27% |
| Inhibitory Control NSE | 7.90 | 1.30 | 277 | -2.23  7.34 | 7.78 | 1.40 | 401 | -2.31 to -1.76  4.7 to 7.55 | 1.57% | 7.28% |

Table Note: Appendix Four Table 1 presents descriptive statistics across all 20 multiply imputed datasets. Comparison between complete case and multiply imputed data indicates relative similar means, which generally differ by less than 5%. However, Relational Binding SE and NSE task means differed by 18.4% and 11.12% respectively.

Generally, pooled standard deviations were similar, with the range of standard deviations in multiply imputed data and the standard deviation of complete case data differing by less than 5%. However, Relational Binding and Inhibition tasks differed by 39.9% to 40.1% and 7.28 to 10.27%.

These values do not appear to indicate an error in the multiple imputation procedure, as these are below a suggested ‘rule of thumb’ of differences in the mean of a large magnitude of 2 standard deviations of the variable, and of differences of a ratio of 2.0 times the variance or larger (Azur, Stuart, Frangakis, Leaf, 2011).

Supplement 4 Table 2. Standardised Beta’s for Path Models

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Complete Case  (n = 236, 203, 255) | | | Multiply imputed data  (n = 401) | | | Full or limited maximum likelihooda  (n = 243, 258, 286) | | |
|  | SE | NSE | Wald-test | SE | NSE | Wald-test | SE | NSE | Wald-test |
| *Relational Memory* | | | | | | | | | |
| Maternal Sensitivity | -0.16 (0.06)  p = .009 | 0.09  (0.07)  p = .200 | 7.67,1,  p = .006 | -0.04 (0.08)  p = .590 | 0.10 (0.09)  p = .251 | 1.20, 1,  p = . 274 | -0.16 (0.06)  p = .009 | 0.09 (0.07)  p = .185 | 7.86,1,  p = .005 |
| Socio-economic status | 0.22  (0.07)  p = .001 | -0.03 (0.06)  p = .588 | NA | 0.08 (0.10)  p = .394 | -0.01 (0.10)  p = .9120 | NA | 0.22 (0.07)  p=  .001 | -0.02 (0.06)  p = .748 | NA |
| *Cognitive Flexibility* | | | | | | |  |  |  |
| Maternal Sensitivity | 0.04  (0.10)  p =.679 | 0.16  (0.10)  p = .114 | 1.16,1  p = .282 | 0.04 (0.09)  p =.636 | 0.10 (0.09)  p = .249 | 0.33, 1  p = .564 | 0.07 (0.10)  p = .483 | 0.09 (0.09)  p = .315 | 0.05, 1,  p = .832 |
| Socio-economic status | 0.33  (0.09)  p = .000 | 0.25  (0.09)  p = .007 | NA | 0.30 (0.09)  p = .001 | 0.24 (0.09)  p = .009 | NA | 0.34 (0.08)  p = .000 | 0.26 (0.09)  P = .004 | NA |
| *Inhibitory Control* | | | | | | |  |  |  |
| Maternal Sensitivity | 0.04  (0.07)  p =.583 | 0.14  (0.07)  p = .039 | 2.10, 1  p = .147 | 0.09 (0.07)  p =.190 | 0.14 (0.06)  p =.023 | 0.56, 1  p = .454 | 0.05 (0.07)  p = .422 | 0.13 (0.06)  p = .035 | 1.69,1  p = .193 |
| Socio-economic status | -0.02 (0.07)  p = .724 | 0.00  (0.06)  p = .950 | NA | 0.01 (0.07)  p =.876 | 0.03 (0.06)  p = .608 | NA | 0.01 (0.07)  p = .898 | 0.02 (0.06)  p = .682 | NA |

Table Note: a Full information maximum likelihood used for continuous outcomes of Relational Memory and Inhibitory Control ; Limited/Reduced maximum likelihood used for categorical outcome of Cognitive Flexibility

As shown in Appendix 4 Table 2, multiply imputed data indicates that maternal sensitivity did not predict either NSE or SE forms of relational memory. Otherwise, p-values of standardised beta estimates of maternal sensitivity were consistent with complete case analyses (i.e. non-significant remained non-significant). As is also apparent, there is a smaller beta and larger standard error in the beta estimate of the multiply imputed data (i.e., 0.05, *SE* = 0.10), than in the complete case data (i.e. -0.16, *SE =* 0.06). This, in turn reduces the p value. That is, the p-value for the beta is based on the z-statistic, calculated as the beta estimate divided by the standard error (Muthen & Muthen, 2017). In this way, a smaller beta and larger standard error makes the z-statistic smaller, thus reducing the corresponding p-value.

Still, full/limited maximum likelihood results indicated that maternal sensitivity had a greater association with NSE than SE versions of relational memory. This, along with the rest of the results, were consistent with complete case analysis (i.e. significant remained significant and non-significant remained non-significant).

**Discussion**

Complete case analysis was observed to be consistent with one modern missing data technique, full/limited maximum likelihood. That is, p-values which were significant in complete case models remained significant when analysed with full/limited maximum likelihood. However, such agreement was not found with multiple imputation, where significant findings with relational memory were no longer significant with multiple imputation. This is likely due to the inclusion of 5 auxiliary variables in multiple imputation, which allows for a more inclusive missing data model, such that more information was utilised to estimate the resultant multiple imputation models than the full/limited maximum likelihood models.

Recall that, based on prior literature, not only were key variables of sensitivity and NSE and SE outcomes included, but also, 5 conceptually supported covariates were specified as auxiliary variables in this multiple imputation model, to meet MAR assumptions. Nevertheless, we cannot be certain that all these variables, and only these variables, should have been included. The contribution of variables to meeting MAR assumptions appears small. No auxiliary variable associated with NSE and SE outcomes (Table 1, main manuscript), whilst associations with missing cases though significant and numerous also appear small (Appendix 3).

Additionally we struggled to determine whether and how to include data from the DCCS in our multiply imputed data sets. DCCS missing values may occur either because the child did not take part in the task, or because the child did not pass pre-switch. Passing pre-switch is a necessary condition for a post-switch score because the DCCS is meant to reflect one specific aspect of cognitive ability, and not more general attention. Because the necessary cognitive resources for pre- and post-switch are very similar, except, the ability to switch, only when children first pass pre-switch is it possible to understand how they are performing on switching, per se.

While this makes good theoretical sense, the omission of data based on pre-switch performance leads to two issues. First, the act of imputing missing values on the DCCS may increase measurement error; children who had failed ‘pre-testing’ on the DCCS now receive an imputed score on the DCCS. Still, a further sensitivity analysis in keeping with Goh et al, 2020, which filtered out all children who failed pre-test, indicated equivalent results (e.g., no differences in p< 0.05 versus p> 0.05).

A second consideration for DCCS data was its potentially problematic role in generating other multiply imputed data sets. That is, the very fact that a child received a post-test score (whether or not he/she passed or failed the DCCS) could suggest that these children exhibited more attention during the DCCS than did most children who failed pre-switch. Because attention is also important to other forms of cognition, it may not be appropriate to only impute other cognitive variables for children with available DCCS scores. Doing so may enhance the resultant imputed sample means. On the other hand, it may also be inappropriate to impute other cognitive variables after imputing missing DCCS scores from children who did not participate in the task, because we would expect that had these children participated in the task, a certain proportion would also have failed pre-switch. Given these constraints, we ran additional multiply imputed Relational Memory and Inhibitory Control models, without including the DCCS as a predictor of other variables. Still, despite changes in the exact statistics and significance levels, there were no differences in whether p-values crossed conventional lines of statistical significance (e.g., p>0.05). Taken together, it does not appear that reported multiply imputed results are biased by including the DCCS. (Additional analyses with these different approaches to DCCS data are available upon request.)

Multiple imputation is typically expected to reduce standard errors (Little, Jorgensen, Lang, Moore, 2014; Harel et al., 2017), assuming that MAR is met. Although it would have been possible to ‘re-run’ the multiple imputation model with a new set of auxiliary variables, until standard errors became smaller, we were concerned that such an approach would ultimately be unreliable. Such an empirically driven approach would, by definition, be atheoretical, and increase Type I error. In other words, there is an increased risk that observed statistical significance and large contributions are simply a result of random chance in this sample, and not replicate.

Still, enlargement of standard error is not necessarily a sign of a ‘faulty’ multiple imputation. It may simply be that there is more variation in task performance amongst 401 multiply imputed then 203 to 255 complete case children. Moreover, the measures in this study all have some degree of measurement error, which may increase or decrease standard errors (Ree & Carretta, 2006) with a larger sample. Without knowledge of the ‘true’ data generating mechanism of the missing data, it remains an open question as to how well MAR assumptions were met, and if multiply imputed results are more or less superior to complete case analyses which are only unbiased when missing data is completely at random (Graham et al., 2009). We anticipate that this question will be better answered once future studies which examine causal mechanisms responsible for missing data in the context of sensitivity and these three tasks are available.

Please address questions regarding this supplement to the second author, Dr. Goh Kok Yew Shaun @ [shaun.goh@nie.edu.sg](mailto:shaun.goh@nie.edu.sg).

Supplemental References

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, *2010*, 1-10.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work?. *International journal of methods in psychiatric research*, *20*(1), 40-49.

Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour research and therapy*, *98*, 4-18.

Goh, S.K.Y., Yang, H., Tsotsi, S. *et al.* (2020). Mitigation of a Prospective Association Between Early Language Delay at Toddlerhood and ADHD Among Bilingual Preschoolers: Evidence from the GUSTO Cohort. *J Abnorm Child Psychol* **48,** 511–523. https://doi.org/10.1007/s10802-019-00607-5

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, *60*, 549-576.

Harel, O., Mitchell, E. M., Perkins, N. J., Cole, S. R., Tchetgen Tchetgen, E. J., Sun, B., & Schisterman, E. F. (2018). Multiple imputation for incomplete data in epidemiologic studies. *American journal of epidemiology*, *187*(3), 576-584.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of pediatric psychology*, *39*(2), 151-162.

Miyake, A., et al. (2000). "The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis." Cogn Psychol 41(1): 49-100.

Muthén, L.K. and Muthén, B.O. (1998-2017).  Mplus User’s Guide.  Eighth Edition.

Los Angeles, CA: Muthén & Muthén

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, *9*, 157.

Ree, M. J., & Carretta, T. R. (2006). The role of measurement error in familiar statistics. *Organizational Research Methods*, *9*(1), 99-112.