***Journal of Motor Learning and Development*, 2021, 9, 109–128**
https://doi.org/10.1123/jmld.2020-0018

Human Kinetics
REVIEW

Missing Data Reporting and Analysis in Motor Learning and Development: A Systematic Review of Past and Present Practices

Priya Patel, Seungmin Lee, Nicholas D. Myers, and Mei-Hua Lee
Michigan State University

Missing data incidents are common in experimental studies of motor learning and development. Inadequate handling of missing data may lead to serious problems, such as addition of bias, reduction in power, and so on. Thus, this study aimed to conduct a systematic review of the past (2007) and present (2017) practices used for reporting and analyzing missing data in motor learning and development. For this purpose, the authors reviewed 309 articles from five journals focusing on motor learning and development studies and published in 2007 and 2017. The authors carefully reviewed each article using a six-stage review process to assess the reporting and analyzing practices. Reporting of missing data along with reasons for their presence was consistently high across time, which slightly increased in 2017. Researchers predominantly used older methods (mainly deletion) for analysis, which only showed a small increase in the use of newer methods in 2017. While reporting practices were exemplary, missing data analysis calls for serious attention. Improvements in missing data handling may have the merit to address some of the major issues, such as underpowered studies, in motor learning and development.

***Keywords*:** bias, maximum likelihood, multiple imputation, power

Missing data are a problem commonly encountered by researchers in quantitative studies (Peugh & Enders, 2004; Pigott, 2001). Often referred to as a nuisance (Schafer & Graham, 2002), missing data may be encountered due to technical errors, participant withdrawals, data not matching with the performance criteria, or sometimes due to reasons unknown. Such data loss can be classified into different types, such as unit nonresponse (entire data fails), item nonresponse (data is partially available), or wave nonresponse (in longitudinal studies when data are missing for one or more sessions; Graham, Cumsille, & Elek-Fisk, 2003; Schafer & Graham, 2002). Experimental studies involving infants may encounter

The authors are with the Department of Kinesiology, Michigan State University, East Lansing, MI, USA. Patel (patelp49@msu.edu) is corresponding author.

an additional data loss in which part of the data is not available for analysis when they are crying or frustrated. Data during this time are not considered representative of typical infant behavior (Thelen, 1981), hence generally excluded from the analysis. Irrespective of the type, missing data are generally inevitable in at least some experimental studies, which makes it important to carefully account for data loss since every data collection costs time, energy, and resources of everyone involved (Pigott, 2001).

Accurate missing data reporting is critical from several aspects. The American Psychological Association (APA) Task Force on Statistical Inference released a report stating, "Before presenting results, report complications, protocol violations, and other unanticipated events in data collection. These include missing data, attrition, and nonresponse" (Wilkinson & Task Force on Statistical Inference, 1999, p. 597). An important aspect of missing data reporting is to account for the possible reasons behind data loss. These reasons may help in understanding the potential problems related to experimental design in a study and increase its replicability in the future. For example, presence of too many outliers or withdrawals in a study may indicate that the task is either too difficult or is not interesting enough for the participants. Sometimes, it may also indicate task inappropriateness for particular groups of people. Similarly, too many dropouts/missing waves in a longitudinal study may indicate that the experimental design is not feasible for participants. Identifying and reporting such reasons may increase the overall replicability of an experiment and provide directions to the primary investigator(s) and others in that field for future research studies. Therefore, it is critical to report the presence of missing data and provide possible reasons for their presence in research articles.

In addition to accurate reporting, appropriate handling of missing data during data analysis is equally important. Inappropriate handling of missing data may lead to serious issues in data analysis, thereby reducing the validity of results (Jeličić, Phelps, & Lerner, 2009). Problems, such as reduction in power (Bennett, 2001; Enders, 2013), bias in parameter estimates (Jones, 1996), bias in standard errors (SEs) and test statistics (Glasser, 1964), inefficient use of the data (Afifi & Elashoff, 1966), reduction in precision (Wood, White, & Thompson, 2004), may occur when missing data are ignored during data analysis. Take for instance a hypothetical study aimed at assessing the learning rate of a new motor skill in children. For this, researchers tested 20 children to reach the intended sample size of 10 complete data sets. Data were missing for 10 participants, as some could not finish the study while others decided to withdraw. Here the low completion rate of children may question the appropriateness of the experimental task in measuring the learning rate. Alternatively, one may question if the sample consisting of complete data sets alone is a fair representation of the population on which inferences are made. In this situation, conducting data analysis using complete data sets alone may reduce power and introduce bias in the results (Enders, 2013). However, one may address these concerns by appropriately handling missing data during the analysis.

An important step toward appropriate missing data handling is to make reasonable assumptions about the nature of missingness (Peugh & Enders, 2004; Wood et al., 2004). There are three assumptions based on possible relationships between the data and the distribution of missingness (Little & Rubin, 1987; Rubin, 1976). Missing completely at random (MCAR) assumption holds when the reason for missingness is not related to any of the measured variable(s) in the study as well as the values of missing data. Missing at random (MAR) assumption holds when it may be related to one or more of the measured variables in the study but is not related to the values of missing data. Missing not at random (MNAR) assumption holds when the reason for missingness is related to the values of the missing data. In our previous hypothetical study example, MCAR would hold if data were missing due to technical errors, MAR would hold if children who withdrew or who could not finish the study had any common demographics (socioeconomic status, ethnicity, etc.), and MNAR would hold if these children had poor motor learning ability. The MCAR assumption is testable (Little, 1988) but is probably the most unrealistic form of missingness in the real world (Bennett, 2001; Enders, 2013; Jeličić et al., 2009). The MAR is not testable (Allison, 2003; Enders, 2013; Schafer & Graham, 2002). Finally, data analysis with MNAR assumption is often difficult as the true values of missingness are unknown (Enders, 2013; Jeličić et al., 2009). Owing to the complexity of MNAR-based methods,[1] we will focus only on methods based on MCAR and MAR in this review.

Missing data assumptions are important when selecting a method for data analysis, as their violation may produce detrimental effects on overall results (Enders, 2010; Little & Rubin, 1987). Methods that are either robust to violations of assumptions or those that perform well under multiple assumptions are usually preferred. The "older" methods such as listwise, pairwise deletion typically function only under MCAR (Peugh & Enders, 2004). Moreover, listwise deletion can lead to substantial power reduction (Enders, 2013). The APA described these methods as among the worst methods available for practical applications (Wilkinson & Task Force on Statistical Inference, 1999). On the contrary, "newer" methods, such as maximum likelihood estimation and most multiple imputations (developed under MAR assumption), produce unbiased and efficient parameter estimates under both—MAR and MCAR (Peugh & Enders, 2004). Moreover, these methods retain all the data. Therefore, the use of newer methods is encouraged in methodological literature due to their effectiveness (Enders, 2001a, 2001b, 2003; Enders & Bandalos, 2001; Gold & Bentler, 2000; Graham & Schafer, 1999; Muthén, Kaplan, & Hollis, 1987). However, an inclination toward the use of older methods over newer methods of analysis sometimes is observed in applied research fields (Jeličić et al., 2009).

While missing data may be ubiquitous in most empirical fields, the challenges in reporting and conducting data analysis in their presence may be at least somewhat unique to each of the fields. Peugh and Enders (2004) conducted a systematic review of missing data practices in education and applied psychology. They found an increase in explicit reporting of missing data from 33.75% (1999) to 74.24% (2003) owing to the APA report (Wilkinson & Task Force on Statistical Inference, 1999). However, there was no substantial change in the use of methods for analysis as most studies continued to use deletion methods. Similar systematic reviews were conducted in applied psychology and education (Roth, 1994), developmental science (Jeličić et al., 2009), political science (King, Honaker, Joseph, & Scheve, 2001), epidemiology (Eekhout, de Boer, Twisk, de Vet, & Heymans, 2012), and medical (Burton & Altman, 2004; Wood et al., 2004) to assess field-specific trends in missing data practices. While the overall trends from these studies were similar to each other, the methods used for missing data analysis and complications encountered were more field-specific. For example, deletion methods were more prevalent in developmental science (Jeličić et al., 2009) but last observation carried forward method was used commonly in medical studies (Wood et al., 2004). While such missing data analyzing trends are known for some empirical fields, to our knowledge, no such information is available for motor learning and development.

Since most studies in motor learning and development are experimental and involve human participants, occurrences of missing data incidents are quite likely. Given the importance of missing data in research studies (Wilkinson & Task Force on Statistical Inference, 1999) and the present scenario wherein a *Nature* survey found 90% of researchers agreeing to have a "reproducibility crisis" (Baker, 2016), a check on missing data reporting and analyzing practices may not be uncalled for. Threats such as under- or over-power, *P* hacking, publication bias to reproducible science addressed in different fields (Camerer et al., 2018; Munafò et al., 2017; Open Science Collaboration, 2015) have also surfaced in motor learning and development (Lohse, Buchanan, & Miller, 2016; Ranganathan, Tomlinson, Lokesh, Lin, & Patel, 2020). It is found that studies without sufficient sample size often find difficulty in detecting real underlying effects due to sampling variability (Button et al., 2013; Lohse et al., 2016). While missing data directly affects the sample size, using older methods, such as case deletion or mean substitution, in studies with a small sample size may exacerbate the problems of power reduction, inefficient use of data, and bias in parameter estimates. Although missing data reporting and analysis are not a silver bullet for all the threats to reproducible science, it may have the potential to be an important adjunct to the proposed methods (Munafò et al., 2017) for increasing replicability and transparency of research studies.

Given that appropriate reporting and handling of missing data has several advantages for the field of motor learning and development, we focused on past and present practices in the field. We conducted a systematic review of publications from two separate years (2007 and 2017) for practices used to report and analyze missing data in motor learning and development. We focused on three specific research questions:

(a) How often are studies reporting missing data in motor learning and development?
(b) Which methods did the studies predominantly use for handling missing data?
(c) Is there a change in missing data reporting and analyzing practices from 2007 to 2017?

Method

We designed our method based on previous systematic missing data reviews (Jeličić et al., 2009; Peugh & Enders, 2004; Wood et al., 2004) conducted in other related fields instead of using structured guidelines provided by PRISMA for conducting reviews and meta-analysis. Our main rationale behind this was to keep our study comparable to other missing data reviews. We conducted the systematic review of missing data reporting and analyzing practices by dividing the process into six main stages. The primary coder (first author) reviewed all the articles under inclusion for missing data reporting and handling practices. In addition, the secondary coder (second author) independently reviewed 25% of the total number of included studies for interrater agreement and reliability testing (Adolph et al., 2012; Franchak & Adolph, 2012; Wood et al., 2004). We explain below the six main stages of missing data review and schematically present them in Figure 1.

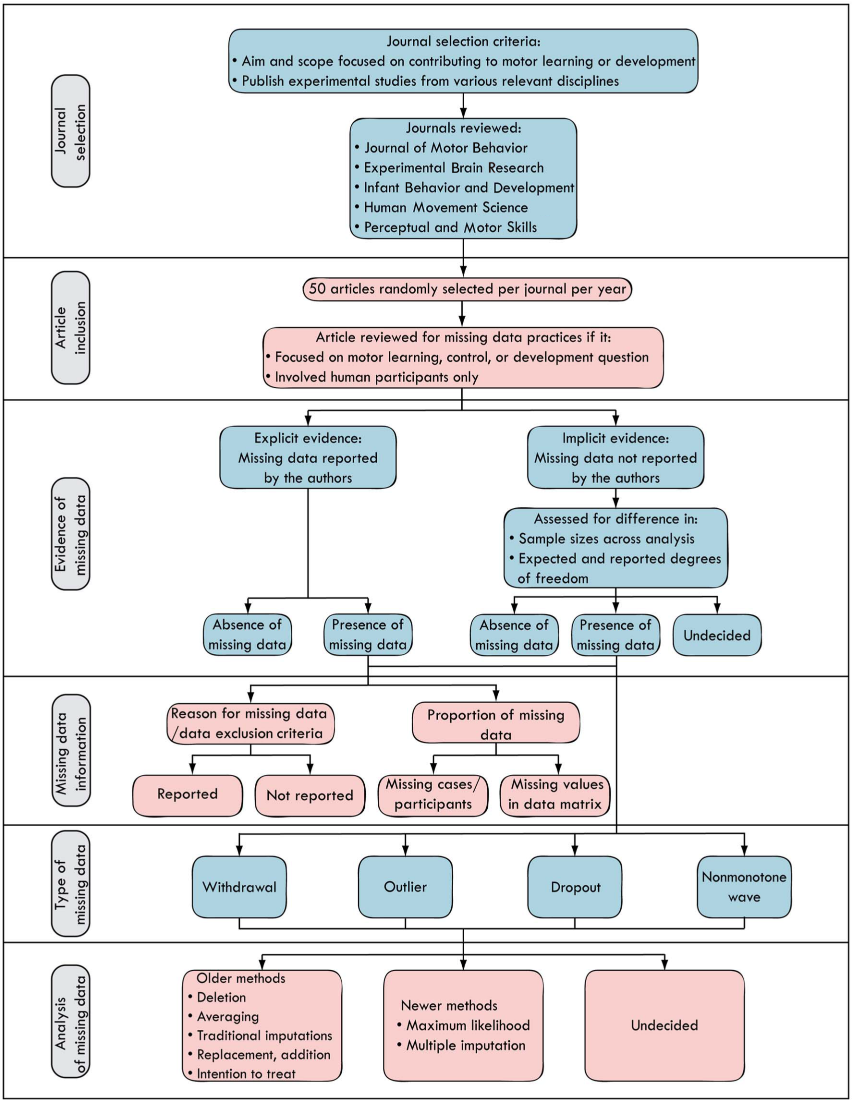

Figure 1 — Schematic presentation of a systematic review of missing data reporting and analysis conducted using six stages.

Journal Selection

Based on previous review studies (Jeličić et al., 2009; Peugh & Enders, 2004), we set two criteria for journal selection: (a) its aim and scope should be focused on contributing to the understanding of motor learning or development and (b) it should publish experimental studies in various relevant disciplines. Based on these criteria, we selected five journals as being representative of typical studies: (a) *Journal of Motor Behavior*, (b) *Experimental Brain Research*, (c) *Infant Behavior and Development*, (d) *Human Movement Science*, and (e) *Perceptual and Motor Skills*. To account for chronological trends in missing data practices, we reviewed the journals for the years 2007 and 2017.

Article Inclusion

The total number of articles published in each journal varied from 45 to 600, so we set an operational frequency of 50 articles per year per journal for conducting the review (Peugh & Enders, 2004). At first, we assigned an identification number to all the articles published in a journal for a given year based on their publication order. Next, we generated a random sequence of 50 whole numbers. We then selected all the articles with identification numbers corresponding to the numbers in that random sequence. For example, if a journal had more than 50 articles published, then we generated a random sequence of 50 numbers and included the articles with corresponding identification numbers for review. If a journal had less than 50 articles published in a year, then we included all of them for review. Next, we included each of the selected articles for missing data review if it met with the following two criteria: (a) focus on motor learning, control, or development question and (b) experimental study involving humans. Furthermore, we considered articles with multiple studies (Study 1, Study 2, etc.) as separate units and reviewed them individually.

Evidence of Missing Data

We examined every article that met the inclusion criteria for evidence of missing data. We categorized evidence of missing data into two types: (a) explicit and (b) implicit. Explicit evidence referred to when the authors directly reported the presence or absence of missing data. Implicit evidence referred to when the authors did not directly report any information on either the presence or absence of missing data. For implicit evidence, we examined the articles for: (a) differences in sample sizes across analyses and (b) differences in expected and reported degrees of freedom (*df*). We considered implicit evidence of “presence of missing data” (also known as blanket removal of missing data; Peugh & Enders, 2004) if we found a discrepancy in either of the previously mentioned two conditions. We considered implicit evidence of “absence of missing data” if we found no discrepancy in the sample sizes or *df* in the data analyses. However, the analysis of different variables that does not require the entire data may also lead to a discrepancy in the *df*. In such cases, it is difficult to conclude if the discrepancy was a result of blanket removal of cases or not; therefore, we coded the presence of missing data as “undecided” for such studies.

Missing Data Information

We reviewed all the studies that explicitly reported the presence of missing data for two additional details: (a) reason for missing data and (b) proportion of missing data.

While data can be unavailable for analysis due to different reasons, we considered two main reasons here: (a) data are missing or recorded incorrectly or (b) experimenter excludes the data based on the performance criteria. In either case, reporting the reason for missing data or the data exclusion criteria may help the readers to understand the results and increase the overall replicability of the study. Thus, we reviewed if the studies reporting the presence of missing data also reported the reason behind missing data or criteria for data exclusion.

Next, we reviewed the amount of missing data/data loss in these studies as the proportion of (a) cases/participants lost/excluded, and/or (b) missing values in the data matrix (e.g., number of trials excluded, etc.). Since there is no established range or cutoff for defining the proportion of missing data as acceptable, we set up an operational range based on some heuristics provided in Schafer (1999) and Bennett (2001). According to this operational range, we counted the number of studies that had: (a) <5% missing data (may be considered inconsequential), (b) 5–10% missing data, and (c) >10% missing data (statistical analysis is likely to be biased). While it is important to report the proportion of missing data, studies also emphasize that the missingness assumptions for handling the missing data are more critical than the amount of missing data in deciding the method of analysis and their impact on research results (Enders, 2013; Tabachnick & Fidell, 2012).

Type of Missing Data

The missing data type is important to know in motor learning and development as certain types are under the control of researchers while some are not. For example, data loss due to infant crying is not in the control of the researcher. However, there is no fixed classification or nomenclature available in motor learning and development to describe different types of missing data. Thus, we generated an operational nomenclature[2] for commonly observed missing data types in motor learning and development. We based this nomenclature on whether missing data were researcher- or participant-driven and included two specific types of data loss seen in longitudinal studies. Different types of missing data were defined as follows:

(a) Outlier (researcher driven): When entire/part of the data for a participant was removed, as it did not meet the performance criteria set by the experimenter or there was a technical fault.

(b) Withdrawal (participant driven): When entire/part of the data for a participant was not available, as they decided to stop due to fatigue, boredom, non-cooperation, and so on. In infant studies, data where an infant is crying or agitated are often removed.

(c) Dropout (longitudinal study-specific): When a participant misses a testing session and never returns for subsequent testing sessions (Ibrahim & Molenberghs, 2009).

(d) Nonmonotone wave (longitudinal study-specific): When a participant misses a session but returns for subsequent testing sessions (Ibrahim & Molenberghs, 2009).

Analysis of Missing Data

We next examined the method used for handling missing data analysis. Based on the type of method used, we classified missing data handling technique as older or newer (Peugh & Enders, 2004; Schafer & Graham, 2002). Older methods included deletion, averaging, traditional imputations (e.g., mean substitution, regression imputation, single imputation, etc.), replacement and/or addition of participants and/or trials, intention to treat, and so on. Newer methods included maximum likelihood estimation and multiple imputations. Whenever the method of missing data analysis was unclear, the primary coder discussed with the third author to reach a conclusion. If the issue did not resolve by discussion, then we approached the corresponding authors of respective studies to determine the type of analysis used. We approached corresponding authors of nine such studies, of which six responded. We arranged a short phone interview with them to discuss the nature of missing data analysis used in their study and coded the study according to the information they provided. In the three studies where the corresponding authors did not respond, the method of analysis remained unclear, and we coded them as undecided.

Results

We conducted descriptive analyses based on previous systematic missing data reviews in other related fields (Jeličić et al., 2009; Peugh & Enders, 2004; Wood et al., 2004) to keep our review findings comparable with these studies. We present results schematically in Figure 2.

Interrater Reliability

Interrater reliability between the two coders for coding different aspects of missing data review was generally high as indicated by percentage agreement between them and Cohen's Kappa statistic. We got a 95.18% agreement with the Kappa statistic of .90 on the evidence of missing data (explicit vs. implicit). We got 87.95% agreement with the Kappa statistic of .81 for coding the presence of missing data. Furthermore, we got 96.38% agreement and Kappa statistic of .93 on the reason of presence of missing data/criteria for data exclusion. We had 91.56% agreement and Kappa statistic of .80 on the type of missing data. Finally, we got 90.48% agreement and Kappa statistic of .84 on the type of missing data handling technique. The primary and secondary coder resolved all the initial disagreements by discussion.

Descriptive Measures

We selected a random sample of 243 articles in 2007 and 250 articles in 2017 from all the five journals for this review. Out of these randomly selected articles,

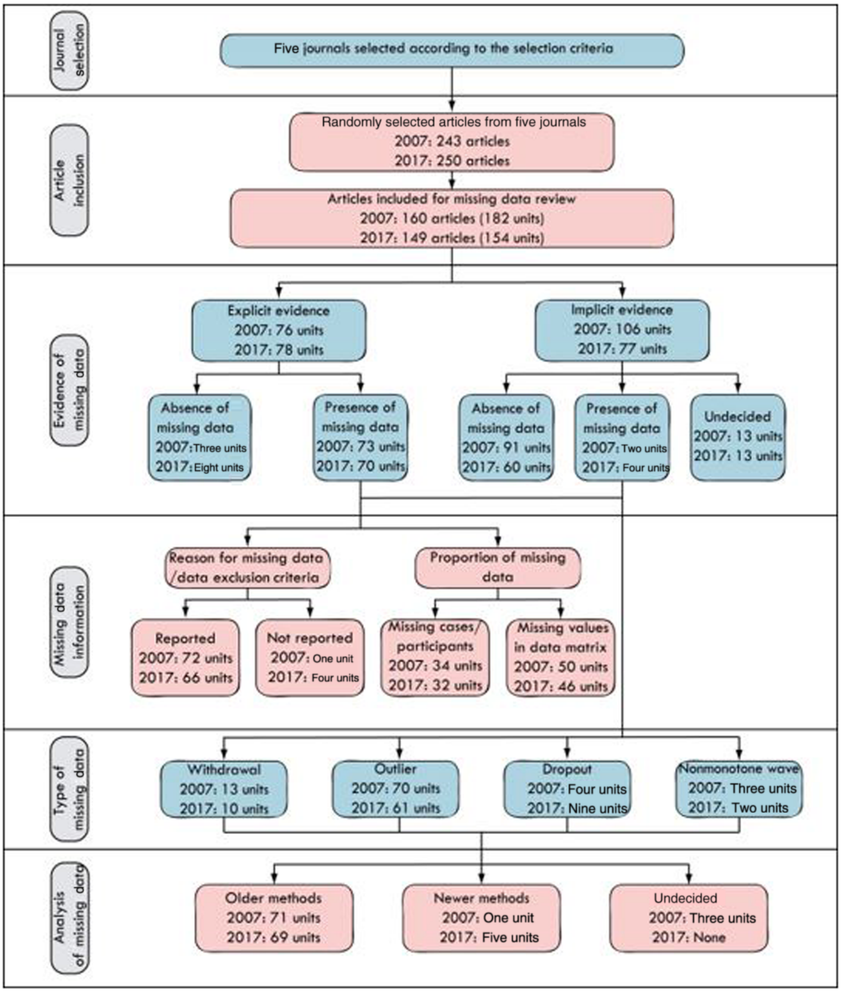

Figure 2 — Schematic of the missing data trends as the number of studies encountered at each stage of the review process. As we encountered articles with multiple studies (e.g., Study 1, Study 2, etc.), we considered each study as one unit and reviewed it independently. We conducted all statistical analyses using the total number of units instead of the total number of articles.

160 (65.84%) in 2007 and 149 (59.60%) in 2017 met the article inclusion criteria, and we further reviewed them for missing data practices. All the five journals contributed evenly toward the total number of articles reviewed for missing data practices (see Table 1); thereby indicating that these journals were review relevant. As several articles consisted of more than one study (Study 1, Study 2, etc.), we individually reviewed 182 units from 160 articles in 2007 and 154 units from 149 articles in 2017. We conducted all statistical analyses using the total number of

Table 1 Frequency Distribution of Articles That Met With Review Criteria Across Five Journals

Journal name	Number of articles meeting review criteria 2007	2017
Experimental Brain Research	34 (21.25%)	30 (20.13%)
Human Movement Science	31 (19.38%)	33 (22.15%)
Infant Behavior and Development	27 (16.88%)	16 (10.74%)
Journal of Motor Behavior	41 (25.63%)	37 (24.83%)
Perceptual and Motor Skills	27 (16.88%)	33 (22.15%)

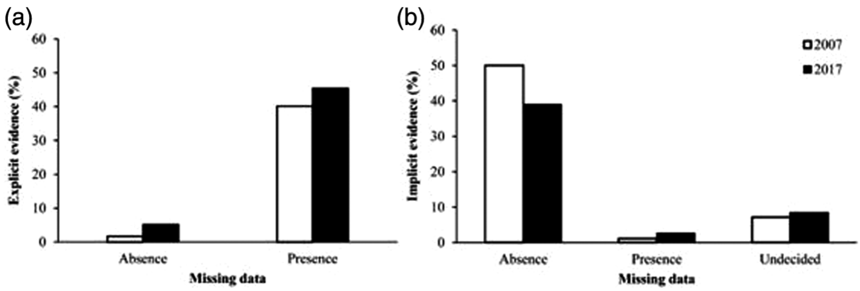

Figure 3 — (a) Explicit and (b) implicit evidence of missing data across 2007 and 2017. Explicit evidence referred to when the authors directly reported the presence or absence of missing data in their study. Implicit evidence referred to when the authors did not report any information on missing data, but the authors of this review found missing data presence or absence. Whenever present, studies meticulously reported missing data across time. The absence of missing data reporting showed an increase in 2017.

units (182 in 2007; 154 in 2017) instead of the total number of articles as we reviewed each unit independently.

Evidence of Missing Data

Explicit reporting about missing data presence or absence increased from 41.75% in 2007 to 50.65% in 2017 (see Figure 3a). To be specific, reporting of missing data presence increased from 40.11% (73 out of 182 units) in 2007 to 45.45% (70 out of 154 units) in 2017. Similarly, reporting of missing data absence increased from 1.65% (three out of 182 units) in 2007 to 5.19% (eight out of 154 units) in 2017.

Implicit presence of missing data was found in 1.10% studies (two out of 182 units) in 2007 and 2.60% studies (four out of 154 units) in 2017 (see Figure 3b). Missing data presence was undecided in 7.14% studies (13 out of 182 units) in 2007 and 8.44% studies (13 out of 154 units) in 2017. Overall, 50% studies (91 out of 182 units) in 2007 and 38.96% studies (60 out of 154 units) in 2017 showed no detectable missing data.

Missing Data Information

Studies consistently reported the reasons for missing data presence/criteria for data exclusion across both the years—98.63% (72 out of 73 units) in 2007 and 94.29% (66 out of 70 units) in 2017. The proportion of missing data (cases/participants or missing values) was consistent across time in terms of both magnitude and reporting rate. The proportion of missing/excluded cases/participants ranged from 2.55% to 60% in 2007 and from 1.08% to 60.87% in 2017. The proportion of missing values in the data matrix ranged from 0.97% to 54% in 2007 and from 0.25% to 76% in 2017. The number of studies within each category of missing data proportions (<5%, 5–10%, and >10%) did not show dramatic change over time (see Table 2). In terms of reporting practices, studies were more prone to reporting the proportion of cases/participants missing/excluded (100% in 2007; 93.75% in 2017) than reporting the proportion of missing values in the data matrix (42% in 2007; 39.13% in 2017).

Type of Missing Data

We found the occurrence of outliers more common (93.33% in 2007; 82.43% in 2017) than that of withdrawal (17.33% in 2007; 13.51% in 2017). Dropout rates increased over time (5.33% in 2007; 12.16% in 2017) while nonmonotone rates decreased over time (4% in 2007; 2.70% in 2017). While most studies encountered only one type of missing data, about 20% studies (15 out of 75 units) in 2007 and 13.51% studies (10 out of 74 units) in 2017 had more than one type of missing data incidences.

Although these trends remained similar across time, we cannot generalize this finding. Motor development-related journals were relatively less in this review pool, so commonly found missing data in this field, such as dropout, nonmonotone, or withdrawal were encountered less than outlier that is commonly found in motor learning studies. As a result, we cannot generalize the frequency of various missing data types across the journals.

Analysis of Missing Data

Studies predominantly used older methods of missing data analysis across both the years—94.67% in 2007 and 93.24% in 2017 (see Figure 4a). Out of 75 units with

Table 2 Frequency Distribution of Articles With Missing Data Proportion in Each Category of Operational Range

	Number of articles with missing participant/cases		Number of articles with missing values in data matrix	
Missing data proportion	**2007**	**2017**	**2007**	**2017**
<5%	6 (17.65%)	3 (9.38%)	10 (20%)	6 (13.04%)
5–10%	7 (20.59%)	9 (28.13%)	5 (10%)	4 (8.70%)
>10%	21 (61.76%)	18 (56.25%)	6 (12%)	8 (17.39%)
Amount not reported	0 (0%)	2 (6.25%)	29 (58%)	28 (60.86%)

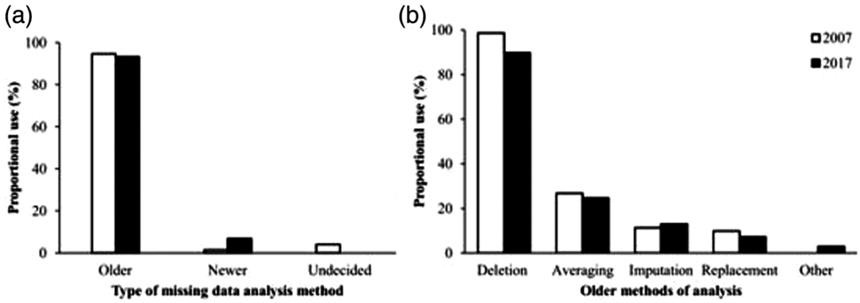

Figure 4 — Proportional use of (a) different types of missing data analysis methods and (b) different older methods of analysis across 2007 and 2017. Studies mainly used older methods as primary methods of analysis throughout 2007 and 2017. Among older methods, studies predominantly used deletion methods across time, sometimes in combination with other older methods.

detectable missing data in 2007, 71 studies used older methods, one study used newer methods, and it was undecided for the rest of the studies. Out of 74 units with detectable missing data in 2017, 69 studies used older methods, and five studies used newer methods.

Among the older methods, deletion methods were most commonly used (98.59% in 2007; 89.85% in 2017), followed by averaging (26.76% in 2007; 24.64% in 2017; see Figure 4b). A small percentage of studies used traditional imputation methods, such as mean substitution, regression, or single imputation (11.27% in 2007; 13.04% in 2017); replacement of participants or trials, and addition of new trials (9.85% in 2007; 7.25% in 2017); and intention to treat (2.90% in 2017). While 45.07% studies (32 out of 71 units) in 2007 used a combination of two or more types of older methods, this reduced to 34.78% studies (24 out of 69 units) in 2017.

Discussion

We conducted a systematic review of the past (2007) and present (2017) practices used to report and analyze missing data in motor learning and development. We found the incidence of missing data not just common (41.21% in 2007; 48.05% in 2017) but quite high in proportion (see Table 2). This, to our knowledge, is the first estimate of missing data in motor learning and development. The two key findings from this study indicate that researchers (a) meticulously reported missing data whenever present and (b) primarily used older methods to analyze data with missing cases. Moreover, we also found some secondary findings, such as consistent reporting of reason/criteria for data loss, rigorous reporting of proportion of missing participant/cases but not so robust reporting of proportion of missing values in data matrix, salient increase in reporting of missing data absence, and use of newer methods for data analysis.

Presence of Missing Data Reported Meticulously Across Time

Whenever missing data were present, their reporting rate was high across time, this only slightly increased in 2017. Researchers consistently reported the reasons for data loss with the proportion of missing data when participants were missing/excluded. However, studies were not so keen on reporting the proportion of missing values in the data matrix. The overall meticulous reporting practice indicates that researchers in motor learning and development understand the importance of missing data incidents. We can attribute this practice to the APA report (Wilkinson & Task Force on Statistical Inference, 1999) which in general increased the awareness among researchers regarding reporting of missing data events (Peugh & Enders, 2004). On the other hand, we saw a subtle (yet interesting!) increase in reporting of missing data absence in 2017 (~5%). This seems to be an indirect effect of the growing urge among the research community for increasing reproducible science (Baker, 2016; Camerer et al., 2018; Munafò et al., 2017; Open Science Collaboration, 2015). Lately, researchers are encouraged to use open science framework, registered reports, reporting checklists, data sharing, and so on. (Munafò et al., 2017) to improve replicability and transparency of the studies. Reporting the absence of missing data may only be an indirect effect of adapting to these practices; however, it has the virtue to increase the overall study transparency.

Reporting the missing data absence is particularly helpful in studies involving intricate data analysis. In studies with multiple parameter estimates requiring only part of the data or data in different forms or complex within-subjects' analysis, *df* may be different across analysis even in absence of missing data. In such cases, unless studies explicitly report missing data absence, the reason behind different *df*s is difficult to predict. We encountered 7.14% (2007) and 8.44% (2017) such studies and labeled them "undecided" for the presence of missing data. If not for the purpose of this review, we would assume that these studies used all the available data for analysis based on trust. However, as quoted in the recent *Nature* article—"Transparency is superior to trust" (Munafò et al., 2017, p. 5), explicit reporting of missing data absence may increase the study transparency. Similarly, reporting the proportion of missing values in the data matrix (we found this reporting to be less rigorous) is equally important in increasing the transparency of studies especially with intricate data analysis. Therefore, we recommend reporting of both the absence and presence of missing data along with the proportion of data missing/excluded for different analysis, as a measure to improve study transparency.

Older Methods Predominantly Used to Analyze Missing Data

After approximately 20 years of the APA report (Wilkinson & Task Force on Statistical Inference, 1999) that condemned the use of older methods, we still found their predominant use in missing data analysis. Deletion methods were the most commonly used methods, and several studies used them in combination with other older methods. While there was a salient increase (~5% in 2017) in the use of newer methods, the overall trends are similar to education and developmental psychology (Jeličić et al., 2009; Peugh & Enders, 2004); thereby calling for serious attention. A gap is often found between the inferential methods recommended in the statistical research and techniques commonly used in applied research (Keselman et al., 1998). Amidst multiple reasons responsible for such a gap and use of older methods, we suspect the main reason is the lack of awareness about newer methods of analysis. Researchers may recognize the problems with older methods; they may even be aware of newer methods but may lack implementation knowledge. We are not discounting the increase found in newer methods' use in 2017, but it could be a corollary effect of using mixed-effects models which itself is in the nascent stage. As reported in a recent motor learning and development review (Lohse, Shen, & Kozlowski, 2020), researchers heavily relied on conventional analysis of variance-based methods over mixed-effects regression models, where the latter are better at handling missing data due to the use of MAR assumption and maximum likelihood. Irrespective of the reasons, it is first important to discuss why using older methods should be discouraged in motor learning and development.

Recently, problems due to underpowered studies are called for serious attention in motor learning, development, and other related fields (Button et al., 2013; Lohse et al., 2016; Ranganathan et al., 2020). While we found deletion methods as the primary method of analysis, using them (particularly listwise) may lead to a notable reduction of power (Enders, 2013; Enders & Bandalos, 2001; Peugh & Enders, 2004). This may be particularly deleterious for small sample size studies that are not uncommon in motor learning and development (Lohse et al., 2016; Ranganathan et al., 2020). Moreover, older methods being less robust to missingness assumptions, there is a consensus that most of these methods introduce bias in parameter estimates when data are not MCAR (Allison, 2003; Enders, 2013; Peugh & Enders, 2004; Schafer & Graham, 2002). Another commonly used older method—"averaging," may sometimes introduce bias under MCAR (Schafer & Graham, 2002). In addition, all conventional forms of imputation methods may cause underestimates of *SE*s (Allison, 2003; Little & Rubin, 1987). Due to these reasons, we recommend using newer methods over older methods, and if the latter are used, then the data should be first tested for MCAR assumption.

MCAR Assumption Tests and Newer Methods of Analysis

A formal evaluation of the MCAR assumption is an important first step in missing data analysis (Jamshidian, Jalal, & Jansen, 2014). For multivariate continuous data, the MCAR assumption is tested using Little's test (Little, 1988) which is a likelihood ratio-based single test statistic assessing the differences between means of different missing data patterns. For repeated categorical data, MCAR can be tested using a test procedure based on weighted least square (Park & Davis, 1993). Both these tests are conservative to small sample sizes and require moderately large samples. Other MCAR assumption tests include, but are not limited to, tests based on generalized estimating equations (Park & Lee, 1997) and tests based on generalized least squares rationale (Kim & Bentler, 2002). Although both these tests are relatively stable for small sample sizes, researchers should check the criterion for small sample in these tests prior to their use. In addition, software packages, such as MissMech (Jamshidian & Jalal, 2010; Jamshidian et al., 2014), allow easy MCAR assumption testing but may show variability for small sample sizes. Essentially, we recommend testing MCAR assumption with either of these tests depending on the data type before considering older methods of analysis.

In our opinion, newer methods of missing data analysis are more appropriate in motor learning and development mainly due to the retention of data. Maximum likelihood estimation[3] is a newer method used in structural equation modeling, mixed models, and so on (Jeličić et al., 2009; Peugh & Enders, 2004). Under multivariate normality, this method estimates the population parameters that maximize likelihood function based on the sample data (Enders, 2013). During this process, temporary imputations are generated for the missing cases using the observed values of other variables; however, no new data is created or imputed in the data set. On the other hand, multiple imputation[4] generates several imputed data sets for missing cases. This helps to overcome the issue of lack of variability present in most regression-based imputations (Allison, 2003). All the imputation data sets are used for the analysis of interests, and the parameter estimates are then averaged to generate a single set of results. To our knowledge, multiple imputation is currently available in SAS, S-Plus, R, and SPSS 17.0 (missing values analysis add-on module needed) and maximum likelihood estimation is available in SPSS AMOS, R, SAS, and Mplus—and likely many other statistical software packages. Although both these methods are more efficient than the older methods for missing data analysis (Peugh & Enders, 2004), we do not claim them to be gold standard.

The efficiency of maximum likelihood and multiple imputation in handling missing data depends on different factors, such as study design, distribution of missingness, and so on. For longitudinal studies, mixed-effects regression (uses maximum likelihood estimation) is better than repeated-measures analysis of variance owing to its relatively efficient handling of missing data (Garcia & Marder, 2017; Lohse et al., 2020). These studies also provide an excellent understanding of different statistical approaches used in longitudinal studies and a road map to which methods work better in certain situations. On the other hand, the use of multiple imputation in studies with randomized control trials is debatable. While alternative methods that are better or equally efficient but less daunting than multiple imputation are found for randomized control trials (Sullivan, White, Salter, Ryan, & Lee, 2018), there is evidence that multiple imputation can be better than maximum likelihood in certain cases (Rombach, Jenkinson, Gray, Murray, & Rivero-Arias, 2018). In terms of missing data distribution, both—maximum likelihood and multiple imputation require an appropriate missingness assumption. Theoretically, multiple imputation does not require MAR assumption, but most of its analyses are developed under this assumption (Schafer & Graham, 2002). Thus, if data are MNAR, then using a MAR-based multiple imputation or maximum likelihood will more likely give biased results (Peugh & Enders, 2004). One must remember before using the newer methods that they differ in terms of their generality and range of analysis (Peugh & Enders, 2004). Therefore, instead of recommending a particular method, we suggest choosing an appropriate method based on study aim(s), data type, analyses of interest, and the statistical tool available.

Limitations and Future Directions

This review has three main limitations. First, we did not have an equal representation of motor development to motor learning studies since the journals focusing on the latter were more than the former. As a result, findings are more generalizable for motor learning than development; however, we still get insights about the missing data scenarios in the latter. Second, we did not conduct the risk of bias assessment and inferential statistics, instead performed an interrater reliability test and descriptive statistics for data analysis. Since we based our study on other missing data reviews, our method and data were not fully in the scope of conducting the risk of bias and inferential statistics. However, basing our study on previous missing data reviews provided a foundation for our study and kept it comparable with them. This was particularly essential since, to our knowledge, this review is the first of its kind in motor learning and development. Third, we did not demonstrate the use of newer methods and mainly focused on conducting a systematic review. However, we consider this as a future direction to provide a demonstration of testing MCAR assumption followed by a comparison of older versus newer methods for handling missing data in experimental studies.

In summary, this study provides a detailed review of missing data reporting and analyzing practices in motor learning and development. With missing data prevalence quite common here, the reporting practices used by researchers are commendable. However, the methods used for missing data analysis call for serious attention. Finally, we propose two essential recommendations: (a) report the absence of missing data and the proportion of data loss for all the analysis to increase study transparency and (b) test MCAR assumption and use newer methods of analysis. Different resources, such as demonstration articles, workshops, seminars, coursework, journal guidelines, and so on can help to increase awareness among researchers and thereby achieve these recommendations. Finally, we believe that improvements in the overall handling of missing data may contribute toward solving some of the most important issues in motor learning and development.

Notes

1. For more information on methods of analysis based on MNAR assumption: Enders (2010, 2011); Muthén, Asparouhov, Hunter, and Leuchter (2011).

2. The authors would like to acknowledge that some types defined here might be used interchangeably by some studies. Such synonymous use is not incorrect in absence of standard classification. However, for this review, missing data were classified using this nomenclature strictly, even in the instances when it was named differently in the study.

3. For detailed information refer: Allison (2003); Enders (2013); Enders and Bandalos (2001); Peugh and Enders (2004); Schafer and Graham (2002).

4. For detailed information refer: Allison (2003); Graham and Hofer (2000); Harel and Zhou (2007); Horton and Lipsitz (2001); Schafer and Graham (2002).

Acknowledgment

This review was supported by the Summer Research Renewal Fellowship awarded through the College of Education at the Michigan State University to Priya Patel.

References

Adolph, K.E., Cole, W.G., Komati, M., Garciaguirre, J.S., Badaly, D., Lingeman, J.M., . . . Sotsky, R.B. (2012). How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science, 23*(11), 1387–1394. doi:10.1177/0956797612446346

Afifi, A.A., & Elashoff, R.M. (1966). Missing observations in multivariate statistics I. Review of the literature. *Journal of the American Statistical Association, 61*(315), 595–604. doi:10.1080/01621459.1966.10480891

Allison, P.D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545. PubMed ID: 14674868 doi:10.1037/0021-843X.112.4.545

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*(7604), 452–454. PubMed ID: 27225100 doi:10.1038/533452a

Bennett, D.A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*(5), 464–469. PubMed ID: 11688629 doi:10.1111/j.1467-842X.2001.tb00294.x

Burton, A., & Altman, D.G. (2004). Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines. *British Journal of Cancer, 91*(1), 4–8. PubMed ID: 15188004 doi:10.1038/sj.bjc.6601907

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., & Munafò, M.R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. PubMed ID: 23571845 doi:10.1038/nrn3475

Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644. PubMed ID: 31346273 doi:10.1038/s41562-018-0399-z

Eekhout, I., de Boer, R.M., Twisk, J.W., de Vet, H.C., & Heymans, M.W. (2012). Missing data: A systematic review of how they are reported and handled. *Epidemiology, 23*(5), 729–732. PubMed ID: 22584299 doi:10.1097/EDE.0b013e3182576cdb

Enders, C.K. (2001a). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*(4), 352. doi:10.1037/1082-989X.6.4.352

Enders, C.K. (2001b). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*(5), 713–740. doi:10.1177/0013164401615001

Enders, C.K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*(3), 322. PubMed ID: 14596494 doi:10.1037/1082-989X.8.3.322

Enders, C.K. (2010). *Applied missing data analysis*. New York, NY: Guilford press.

Enders, C.K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*(1), 1–16. PubMed ID: 21381816 doi:10.1037/a0022640

Enders, C.K. (2013). Analyzing structural equation models with missing data. In G.R. Hancock& R.O. Mueller, *Structural equation modeling: A second course* (2nd ed., pp. 493–519). Charlotte, NC: Information Age Publishing.

Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430–457. doi:10.1207/S15328007SEM0803_5

Franchak, J.M., & Adolph, K.E. (2012). What infants know and what they do: Perceiving possibilities for walking through openings. *Developmental Psychology, 48*(5), 1254–1261. PubMed ID: 22390664 doi:10.1037/a0027530

Garcia, T.P., & Marder, K. (2017). Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model. *Current Neurology and Neuroscience Reports, 17*(2), 14. PubMed ID: 28229396

Glasser, M. (1964). Linear regression analysis with missing observations among the independent variables. *Journal of the American Statistical Association, 59*(307), 834–844.

Gold, M.S., & Bentler, P.M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling, 7*(3), 319–355. doi:10.1207/S15328007SEM0703_1

Graham, J.W., Cumsille, P.E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. Schinka& W. Velicer (Eds.), *Handbook of Psychology: Research Methods in Psychology* (Vol. 2, pp. 87–114). New York, NY: Wiley. doi:10.1002/0471264385.wei0204

Graham, J.W., & Hofer, S.M. (2000). Multiple imputation in multivariate research. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201–218, 269–281). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Graham, J.W., & Schafer, J.L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R.H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1–29). Thousand Oaks, CA: Sage.

Harel, O., & Zhou, X.-H. (2007). Multiple imputation: Review of theory, implementation and software. *Statistics in Medicine, 26*(16), 3057–3077. PubMed ID: 17256804 doi:10.1002/sim.2787

Horton, N.J., & Lipsitz, S.R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician, 55*(3), 244–254. doi:10.1198/000313001317098266

Ibrahim, J.G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: A review. *Test, 18*(1), 1–43. PubMed ID: 21218187 doi:10.1007/s11749-009-0138-x

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika, 75*(4), 649–674. PubMed ID: 21720450 doi:10.1007/s11336-010-9175-3

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software, 56*(1), 1–31. doi:10.18637/jss.v056.i06

Jeličić, H., Phelps, E., & Lerner, R.M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology, 45*(4), 1195. PubMed ID: 19586189 doi:10.1037/a0015665

Jones, M.P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association, 91*(433), 222–230. doi:10.1080/01621459.1996.10476680

Keselman, H.J., Huberty, C.J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donahue, B., . . . Levin, J.R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*(3), 350–386. doi:10.3102/00346543068003350

Kim, K.H., & Bentler, P.M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika, 67*(4), 609–623. doi:10.1007/BF02295134

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review, 95*(1), 49–69. doi:10.1017/S0003055401000235

Little, R.J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Lohse, K., Buchanan, T., & Miller, M. (2016). Underpowered and overworked: Problems with data analysis in motor learning studies. *Journal of Motor Learning and Development, 4*(1), 37–58. doi:10.1123/jmld.2015-0010

Lohse, K., Shen, J., & Kozlowski, A.J. (2020). Modeling longitudinal outcomes: A contrast of two methods. *Journal of Motor Learning and Development, 8*(1), 1–21. doi:10.1123/jmld.2019-0007

Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie du Sert, N., . . . Ioannidis, J.P.A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 1–9. doi:10.1038/s41562-016-0021

Muthén, B., Asparouhov, T., Hunter, A.M., & Leuchter, A.F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STARD antidepressant trial. *Psychological Methods, 16*(1), 17–33. PubMed ID: 21381817 doi:10.1037/a0022634

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*(3), 431–462. doi:10.1007/BF02294365

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Park, T., & Davis, C.S. (1993). A test of the missing data mechanism for repeated categorical data. *Biometrics, 49*(2), 631–638. PubMed ID: 8369395 doi:10.2307/2532576

Park, T., & Lee, S.-Y. (1997). A test of missing completely at random for longitudinal data with missing observations. *Statistics in Medicine, 16*(16), 1859–1871. PubMed ID: 9280038

Peugh, J.L., & Enders, C.K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525–556. doi:10.3102/00346543074004525

Pigott, T.D. (2001). A review of methods for missing data. *Educational Research and Evaluation, 7*(4), 353–383. doi:10.1076/edre.7.4.353.8937

Ranganathan, R., Tomlinson, A., Lokesh, R., Lin, T.-H., & Patel, P. (2020). A tale of too many tasks: Fragmentation of tasks in motor learning and a call for model task paradigms. *Experimental Brain Research*. Advance online publication. doi:10.1007/s00221-020-05908-6

Rombach, I., Jenkinson, C., Gray, A.M., Murray, D.W., & Rivero-Arias, O. (2018). Comparison of statistical approaches for analyzing incomplete longitudinal patient-reported outcome data in randomized controlled trials. *Patient Related Outcome Measures, 9,* 197–209. PubMed ID: 29950913 doi:10.2147/PROM.S147790

Roth, P.L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537–560. doi:10.1111/j.1744-6570.1994.tb01736.x

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. doi:10.1093/biomet/63.3.581

Schafer, J.L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3–15. PubMed ID: 10347857

Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. PubMed ID: 12090408 doi:10.1037/1082-989X.7.2.147

Sullivan, T.R., White, I.R., Salter, A.B., Ryan, P., & Lee, K.J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical Methods in Medical Research, 27*(9), 2610–2626. PubMed ID: 28034175

Tabachnick, B.G., & Fidell, L.S. (2012). *Using multivariate statistics* (6th ed.). Needham Heights, MA: Allyn & Bacon.

Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology, 17*(3), 237–257. doi:10.1037/0012-1649.17.3.237

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594–604. doi:10.1037/0003-066X.54.8.594

Wood, A.M., White, I.R., & Thompson, S.G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials, 1*(4), 368–376. PubMed ID: 16279275 doi:10.1191/1740774504cn032oa