

Missing Step Count Data? Step Away From the Expectation–Maximization Algorithm

Mia S. Tackney,¹ Daniel Stahl,² Elizabeth Williamson,¹ and James Carpenter^{1,3}

¹Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom; ²Department of Biostatistics & Health Informatics, King's College London, London, United Kingdom; ³MRC Clinical Trials Unit at University College London, London, United Kingdom

In studies that compare physical activity between groups of individuals, it is common for physical activity to be quantified by step count, which is measured by accelerometers or other wearable devices. Missing step count data often arise in these settings and can lead to bias or imprecision in the estimated effect if handled inappropriately. Replacing each missing value in accelerometer data with a single value using the Expectation–Maximization (EM) algorithm has been advocated in the literature, but it can lead to underestimation of variances and could seriously compromise study conclusions. We compare the performance in terms of bias and variance of two missing data methods, the EM algorithm and Multiple Imputation (MI), through a simulation study where data are generated from a parametric model to reflect characteristics of a trial on physical activity. We also conduct a reanalysis of the 2019 MOVE-IT trial. The EM algorithm leads to an underestimate of the variance of effects of interest, in both the simulation study and the reanalysis of the MOVE-IT trial. MI should be the preferred approach to handling missing data in accelerometer, which provides valid point and variance estimates.

Keywords: accelerometer, wearable devices, multiple imputation, missing data

Wearable devices, such as pedometers and accelerometers, are becoming a popular tool in clinical and epidemiological studies for measuring participants' physical activity (Bravata et al., 2007). For example, accelerometers have evaluated the impact of interventions aiming to increase exercise in a number of clinical trials (Harris et al., 2015, 2017, 2018; Ismail et al., 2019; Murray et al., 2006). These devices measure acceleration in three dimensions in very fine intervals of time, called epochs, which are then aggregated to obtain step counts on an hourly, daily, or weekly level. Compared to self-report approaches, measurements from these devices do not suffer from recall and desirability bias and there is reduced participant burden (Ae Lee & Gill, 2018). However, missing step count data is a common issue in this setting. Participants may not wear the device as per protocol, and there may be entire days or parts of days where no step counts are recorded. There may also be technical issues such as the battery running out, or water damage to the device, and leading to loss of information. If the analysis does not account for the missing data in an appropriate way, the resulting estimates may be biased or imprecise.

Accelerometer data raise a number of broader missing data issues (Tackney et al., 2021), but we focus here on comparing the

Expectation–Maximization (EM) algorithm and Multiple Imputation (MI) as methods for handling missing data. Analysis of data with missing values requires assumptions about the way in which data become missing—the missingness mechanism. These mechanisms were categorized into three broad classes by Rubin (1976), which we describe in the accelerometer context:

- The *missing completely at random* (MCAR) assumption states that the probability that a step count is missing does not depend on the observed or unobserved data; for example, if a number of accelerometers become faulty by chance and stop recording data, the missingness mechanism is MCAR.
- The *missing at random* (MAR) assumption states that the probability that a step count is missing depends on the observed data, but not on the unobserved data; for example, if younger people are more likely to forget to wear the accelerometer, but their activity levels on days where they forget the device is similar to the activity levels of younger people on days where they wear the device, the missingness mechanism for step counts is MAR given age group.
- The *missing not at random* assumption states that the probability that a step count is missing depends on the unobserved data; this would occur, for example, if people decide not to wear the accelerometer on days where they are less active.


Here, we consider settings where daily step counts are collected, some of which are missing. The primary analysis model has step count as the outcome, and aims to compare step counts between groups. Typically, baseline step counts are accounted for in the model. We assume that the missing data mechanism is MAR. We note that in practice it is not possible to verify that the MAR assumption is met using the observed data; however, it is a natural assumption to conduct the primary analysis under. Sensitivity analysis is recommended to assess robustness of the analysis to violations of the MAR assumption; this is beyond the scope of this article. Our focus is on the statistical properties of the EM

© 2022 The Authors. Published by Human Kinetics, Inc. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0, which permits unrestricted noncommercial and commercial use, distribution, and reproduction in any medium, provided the original work is properly cited, the new use includes a link to the license, and any changes are indicated. See <http://creativecommons.org/licenses/by/4.0>. This license does not cover any third-party material that may appear with permission in the article.

Stahl  <https://orcid.org/0000-0001-7987-6619>

Williamson  <https://orcid.org/0000-0001-6905-876X>

Carpenter  <https://orcid.org/0000-0003-1629-0607>

Tackney (Mia.Tackney@lshtm.ac.uk) is corresponding author,  <https://orcid.org/0000-0003-3868-0550>

algorithm and MI for handling the missing data, in particular, bias and precision of the estimates.

There are various ways of dealing with missing data. First, maximum likelihood methods can handle missing outcome data for linear regression or mixed models (Snijders & Bosker, 2011), which give unbiased effect estimates and valid estimates of variances under the MAR assumption. However, maximum likelihood cannot readily handle missing values in both the outcome and covariates (Carpenter & Smuk, 2021), which is likely to occur in the accelerometer setting as baseline step counts are often incorporated as a covariate in the primary analysis model. This would lead to exclusion of participants with missing covariates, which leads to loss of information and potentially a reduction in statistical power. Thus, in the accelerometer setting, there are two common approaches to handling missing data: single imputation using the EM algorithm, and MI (Ae Lee & Gill, 2018; Borghese et al., 2019; Xu et al., 2018). The literature on the design and analysis of clinical trials caution against the use of single imputation, as it can lead to underestimation of standard errors (SEs) (Dziura et al., 2013; Jakobsen et al., 2017). This has also been demonstrated in simulation studies using observational data (Avtar et al., 2019). In accelerometer studies, however, there has been some misunderstanding in the recommended approach to handling missing data. Using a simulation study, Catellier et al. (2005) compared the EM algorithm and MI in handling intermittent missing data such as missing intervals within days, or missing days within a week. They found that the estimates of mean step counts are similar in terms of bias and precision. Though they acknowledge that the EM algorithm can lead to underestimation of the variance estimates in general, the results from their simulation showing similar performances between the EM algorithm and MI have been used to justify the EM approach to imputation in other accelerometer studies. In this study, we aim to illustrate the EM and MI approaches to handling missing data in the accelerometer setting and demonstrate their statistical properties. We carefully elucidate their performances in terms of the bias, variance, and confidence intervals (CIs) of the treatment effect in a simulation study of a simple trial set up. We then conduct a reanalysis of the 2019 MOVE-IT trial to compare the two approaches to imputation in a more complex setting, and discuss the implications.

EM Algorithm and MI

The EM algorithm is an approach to finding maximum likelihood estimates in the presence of missing data under the MAR assumption (Schafer, 1997). In the context of accelerometer outcomes, the algorithm can provide point predictions for average daily step counts, conditional on participant characteristics such as sex, age, and treatment arm. The missing daily step counts can then be imputed (replaced) by these point predictions from the EM algorithm. This results in a “complete” data set which can then be used for the primary analysis. In this analysis, all values in the “complete” data set are treated equally, regardless of whether the step count was actually observed or imputed using the EM algorithm. This may not be appropriate, because the prediction of the missing values is more uncertain than the observed values—but this information is not used by the primary analysis model, which gives predictions for missing values the same status as observed values.

Multiple imputation is an alternative approach to handling missing data under the MAR assumption, which considers the uncertainty due to the missing values. Given an imputation model,

which in the accelerometer setting can be a joint model for average daily step counts, conditional on characteristics such as sex, age, and treatment arm, MI creates M imputed data sets by replacing each missing value by M different plausible values generated from the imputation model. In each of the M imputed data sets, the imputed value is different, reflecting the uncertainty around the missing value. The imputed data sets are analyzed separately and the results of the M analyses are combined in a pooling step. The point estimates from the M data sets are averaged to get the pooled effect estimate, and the pooled estimate of the SEs incorporate the variability *within* and *between* the M imputations (Rubin, 1976). Thus, MI gives missing observed values a different status to observed values, and the uncertainty around the missing values is taken into account.

The two approaches are illustrated in Figure 1. Technical details of each procedure are provided in the Appendix.

Software

Both the EM algorithm and MI can be implemented in a wide range of statistical software by readily available packages and options.

- SPSS: Both single imputation using the EM algorithm and MI can be conducted (IBM Corp., 2020a, 2020b).
- R: The package `norm` carries out the EM algorithm (Novo & Schafer, 2013). The package `JOMO` implements MI (Quartagno & Carpenter, 2020), and can be run with the interface `mitml`

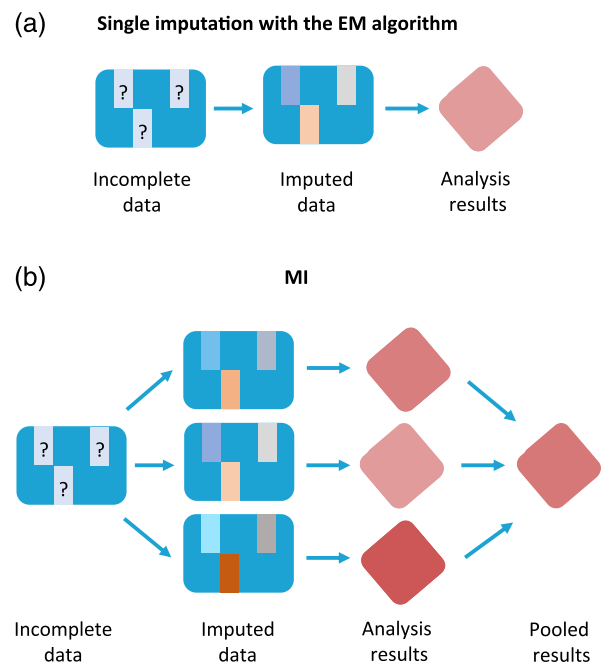


Figure 1 — An illustration of (a) single imputation using the EM algorithm, where missing values are imputed once and the resulting data set is analyzed, and (b) MI, where missing values are imputed M times to create M imputed data sets, which are each analyzed separately, and the results are pooled. MI = multiple imputation; EM = expectation-maximization. This figure is adapted from “Missing Data and Bias in Physics Education Research: A Case for Using Multiple Imputation,” by J. Nissen, R. Donatello, & B. Van Dusen, 2019, *Physical Review Physics Education Research*, 15(2), p. 20106. Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license.

(Grund et al., 2019), which provides tools for visualizing and analyzing multiple imputed data sets. A tutorial for JOMO and mitml is provided by Quartagno et al. (2019). MI can also be implemented in R using mice (van Buuren & Groothuis-Oudshoorn, 2011), which has an associated online vignette by Vink and van Buuren (n.d.)

- Stata: The command `mi impute mvn` can be used to conduct both single imputation with the EM algorithm as well as MI. Furthermore, MI can be performed using the command `mi impute chained` (Statacorp, 2021).
- SAS software: The procedures PROC MI and PROC MIANALYSE implement single imputation using the EM algorithm and MI (SAS Institute Inc., 2021). A tutorial is provided by Yuan (2011).

Simulation

We compare the performance of the EM algorithm and MI for handling missing data by simulating a simple randomized trial setting. We focus on the bias, variance, and CIs of the estimates of the treatment effect obtained under the two methods. In this simulation, we assume that participants provide an accelerometer step count at baseline. They are then randomized to either the treatment or control arm, and then provide a step count after 1 year. The step counts at baseline are fully observed, but there are step counts at Year 1 which are MCAR. While this setup is simplistic, it will provide insight into the statistical properties of the two methods in more general MAR mechanisms.

We generate step count data for this simulation through a parametric model, with parameter settings chosen to reflect characteristics of a trial on physical activity. We denote by $y_{i,0}$ the step count for the i th patient at baseline. Assuming that there is just one observation per person, which is centered around a mean of 7,000 steps, we generate $y_{i,0}$ as:

$$y_{i,0} = 7,000 + \epsilon_i,$$

where ϵ_i is normally distributed with mean 0 and standard deviation 1,700. We note that, while daily step counts are large enough to be treated as continuous data, they are typically right-skewed, so a log-transformation may be necessary for linear regression. We assume for simplicity in this simulation that step counts are normally distributed. We assume that patients are randomized to one of two arms (treatment and control). The variable arm_i is an indicator variable for whether patient i received the treatment. We denote by $y_{i,1}$ the step count for patient i postintervention, one year after baseline. The step counts at Year 1 are drawn from a normal distribution, conditional on arm and step count at baseline:

$$y_{i,1} = 1,000 + 300arm_i + 0.7y_{i,0} + v_i,$$

where v_i is normally distributed with mean 0 and standard deviation 2,000, and ϵ_i and v_i have a correlation of 0.6. The true treatment effect is 300, which is an effect that could realistically be observed in a trial for physical activity. We wish to test the null hypothesis that there is no effect of treatment, against the alternative hypothesis that there is an effect of treatment, with a Type I error of 5%. We generate $n = 500$ patients in each simulation. In this simple setting, we assume that there are no missing step counts at baseline, but there are missing step counts at Year 1. We explore scenarios with the proportion of missing data at Year 1 ranging from 0.1 to 0.9. Missing values are imputed using the EM

algorithm and MI. The EM algorithm is implemented using the R package norm (Novo & Schafer, 2013). MI is conducted via joint modeling using the R packages JOMO (Quartagno & Carpenter, 2020) and mitml (Grund et al., 2019). We use 30 imputations. A small number of imputations, typically greater than five, is sufficient for most applications (Carpenter & Kenward, 2013), but a larger number of imputations is needed for stable estimates of the standard error when the proportion of missing data is large (von Hippel, 2018). For each scenario, we simulate 2,000 data sets to ensure that we estimate the empirical SEs with a Monte Carlo SE of less than 2%. Full details of the implementation are provided in the Appendix. We then analyze the imputed data sets using a linear regression model with step count at Year 1 as the outcome, and step count at baseline and treatment as covariates. From this linear regression model, we estimate the treatment effect and its variance.

We evaluate the two methods by considering the mean, variance, and 95% CI of the treatment effect. The mean of the treatment effect across the 2,000 simulations has expected value of 300 if the treatment effect estimate is unbiased. Furthermore, we expect the theoretical variance of the treatment effect to be similar to the empirical variance (the sample variance of the treatment effect across simulations). If the theoretical variance is underestimated by an approach, the corresponding CIs will be too narrow; conversely, if the theoretical variance is overestimated, the CIs will be too wide. Thus, we assess the performance of each approach by considering the following measures across the 2,000 replications:

- Mean of the estimated treatment effect, which has an expected value of 300.
- Means of the theoretical variance and the empirical variance, which we expect to have similar values.
- Coverage, the proportion of 95% CIs which contain the true treatment effect (300), which we expect to be 0.95. The proportion of CIs that are smaller than the true effect, and the proportion of CIs that are larger than the true effect, should be 0.025.

In Figure 2, we see in the top panel that the estimates of the treatment effect are centered around the true value of 300 for both the MI and EM approaches; this is expected as the missing data mechanism is MCAR. We also observe that the variability of the estimates of the means increases as the proportion of missing data increases; more missing data lead to more uncertainty in the estimated treatment effect. In the middle panel, we observe that the means of the theoretical variances are very different for the two methods; while the means of the variances for MI increase as the proportion of missing data increases, the means of the variances for the EM algorithm remain constant. When we compare this to the plot of the means of the empirical variances in the bottom panel, we observe that, for MI, the theoretical variances are a reasonable estimate of the empirical variances, but for the EM algorithm, the theoretical variances are underestimating the empirical variances. The underestimate of the variances by the EM algorithm becomes increasingly large as the proportion of missing data increases.

In Figure 3, we see in the top panel that the proportion of CIs that contain the true treatment effect decreases as the proportion of missing data increases for the EM algorithm, while for MI, it appears to remain fairly constant. We also observe that the proportion of CIs that are smaller than the true value of the treatment effect (middle panel) and the proportion that are larger than it (bottom panel) increase as the proportion of missing data increases for the EM algorithm, but stays constant for MI.

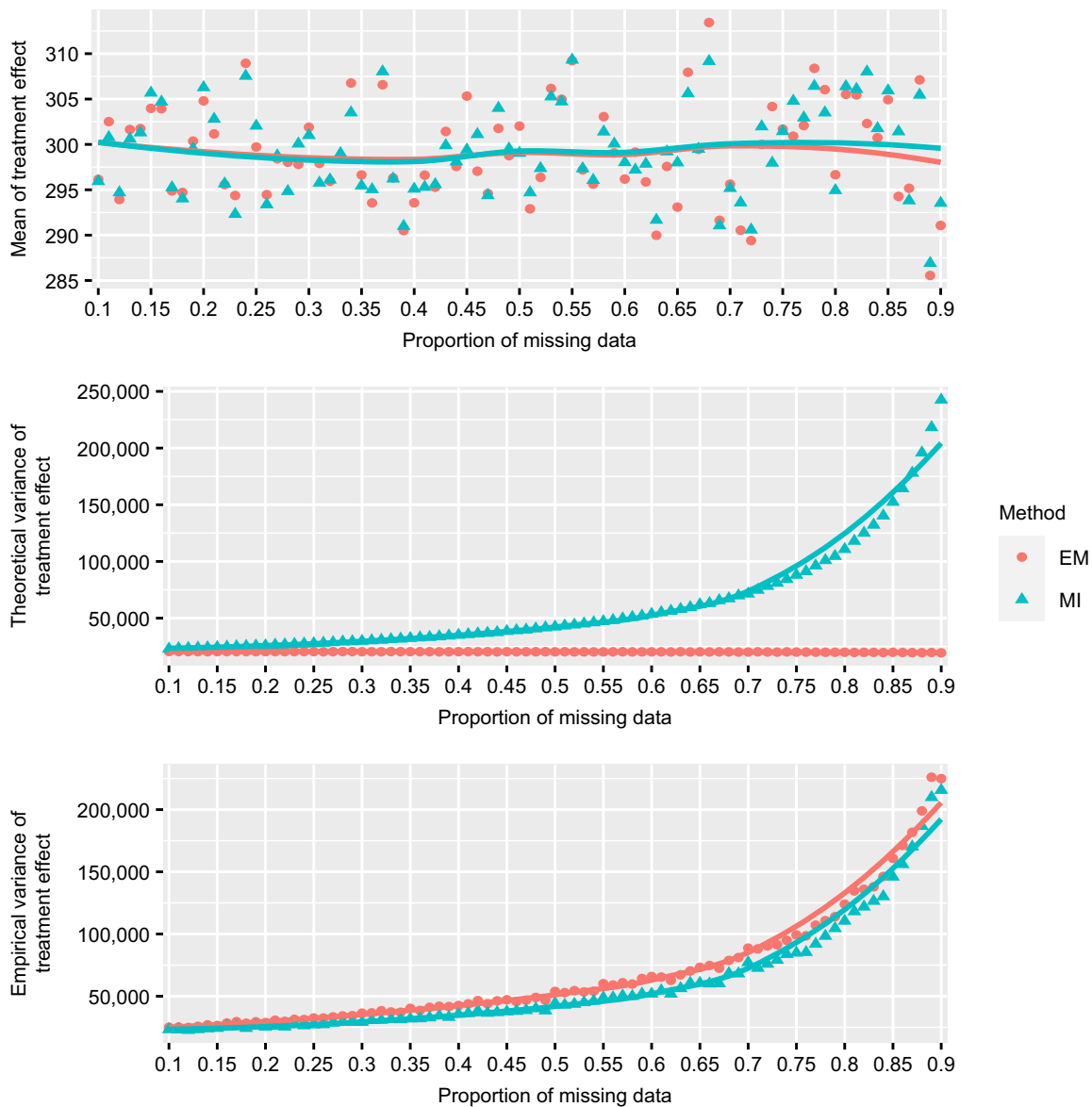


Figure 2 — Plots displaying change in the mean of treatment effect (top), the theoretical variance of treatment effect, (middle) and the empirical variance of the treatment effect (bottom). The blue triangles indicate estimates obtained by MI and the red circles indicate the estimates obtained by the EM algorithm; smoothed lines are added for each method. CI=confidence interval; MI= multiple imputation; EM= expectation–maximization.

Overall, the simulation demonstrates that the EM algorithm underestimates the variance of the treatment effect, where the extent of underestimation increases as the proportion of missing data becomes larger. This leads to CIs that include the true treatment effect less than 95% of the time and provide a false sense of precision around the treatment effect estimate. This implies that Type I error is inflated.

The result is illustrative of the implications in more complicated settings. For example, the same variance underestimation will occur if the missingness mechanism of the Year 1 step counts is MAR. The variance underestimation will also occur if additional variables, such as baseline step count, are MAR. If the missing mechanism is MNAR, both approaches would lead to bias in addition to variance underestimation for the EM algorithm. Furthermore, if one is interested in modeling summaries of step counts, such as weekly averages, where there are intermittent days with

missing data, the underestimation of the variance using the EM algorithm is also of concern.

Next, we illustrate the application of the EM algorithm and MI to the analysis of the MOVE-IT trial. Using real data, we explore a more complex setting where there are three treatment groups, and three time periods at which step counts are measured for each individual. We assume a MAR missingness mechanism and the primary analysis has weekly averaged step counts as the outcome.

Application to the MOVE-IT Trial

We compare the EM and MI approaches to imputation in the analysis of the 2019 MOVE-IT trial (Bayley et al., 2015; Ismail et al., 2019, 2020). The MOVE-IT trial investigated the effects of motivational interviewing and motivational group therapy in

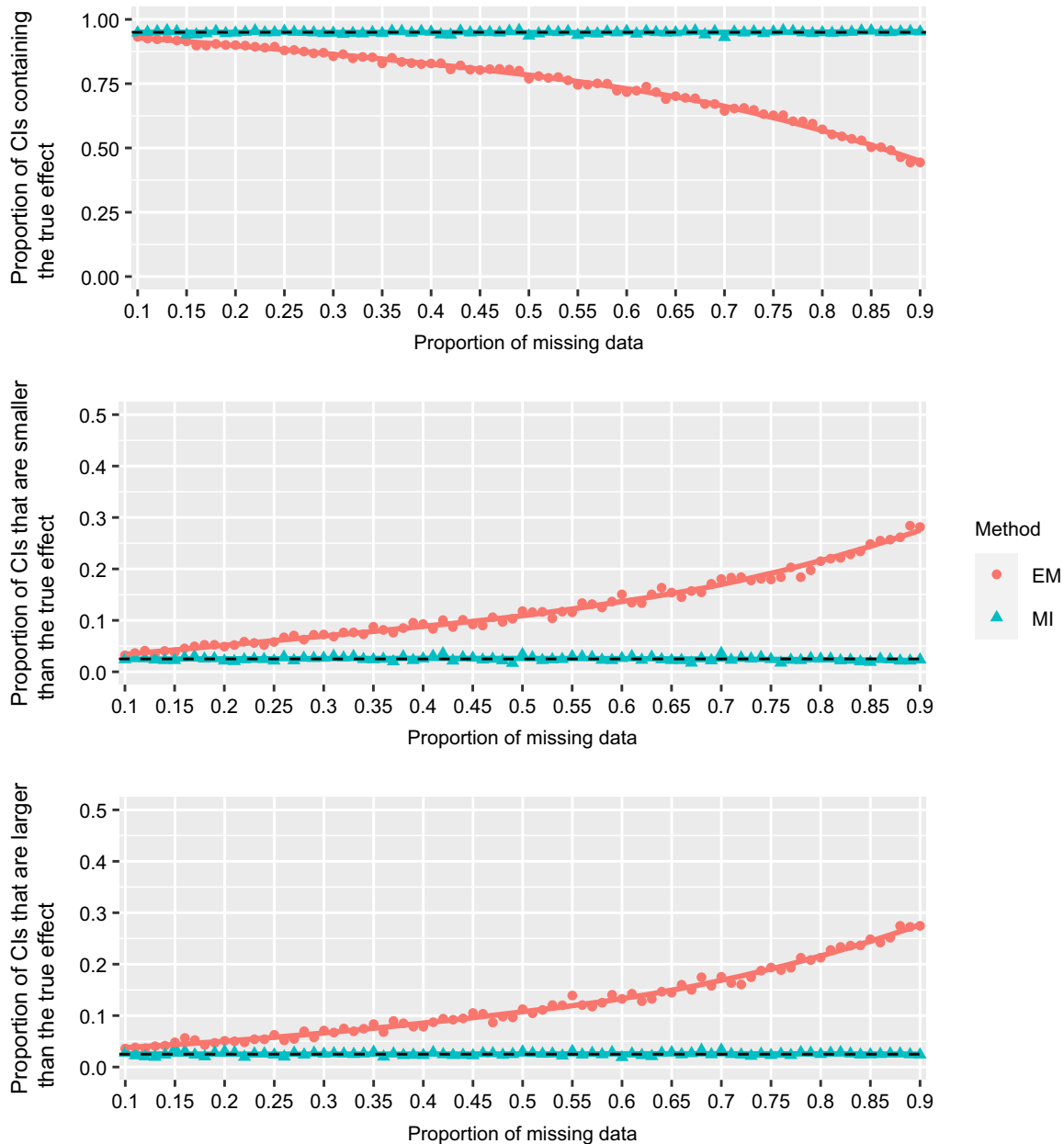


Figure 3 — Plots displaying the proportion of CIs containing the true effect (top), the proportion of CIs that lie below the true effect (middle), and the proportion of CIs that lie above the true effect (bottom) as the proportion of missing data increases. The blue triangles indicate proportions obtained by MI and the red circles indicate the proportions obtained by the EM algorithm; smoothed lines are added for each method. The expected proportion is displayed with a dashed line (0.95 for the top panel and 0.025 for the middle and bottom panels). CI=confidence interval; MI=multiple imputation; EM=expectation-maximization.

reducing weight and increasing physical activity for patients who are at high risk of cardiovascular disease (QRISK2 of 20% or higher; [National Institute for Health and Care Excellence, 2015](#)). The trial randomized patients between three arms: individual motivational interviewing (Arm 1), motivational group therapy (Arm 2), or usual care (Arm 3). Motivational interviewing and motivational group therapy consisted of 10 sessions over the course of a year. The participants recorded their daily physical activity with an ActiGraph GT3X accelerometer (ActiGraph) for a period of seven consecutive days on three occasions: baseline, Year 1, and Year 2. The trial provided insufficient evidence to recommend

either intervention for reducing weight or increasing physical activity.

The outcome of interest is the average step count across a 7-day period ([Ismail et al., 2019](#)). Our analysis model is a mixed model where we have the average step daily count as the outcome. The covariates are year (Year 1 or Year 2), arm, arm-year interaction, baseline average step count, the interaction between baseline average step count and year, gender, and age, and we have an unstructured covariance matrix. We wish to estimate the difference in average step count between the individual therapy and usual care, and the difference in average step count between the

group therapy and usual care, at Year 1 and Year 2. Full details of the analysis and imputation models are provided in the [Appendix](#).

We wish to impute missing days where participants provide at least one observed day during the trial; this means that there is some information from the participant from which information on the missing days can be recovered. Out of 1,742 patients who were randomized to a treatment, 25 did not provide any data on any of the three measurement periods, so they will be excluded from this analysis. Some participants wore the device for longer than 7 days in a measurement period. Data from Days 1 to 7 are used for the analysis, unless participants provided insufficient data on the first day, in which case data from Days 2 to 8 are used instead. If participants wore the device for less than 540 minutes in a day, this observation is considered missing ([Ismail et al., 2020](#)). Table 1 shows the percentage of the 1,717 participants who have missing data on each day at each year.

We impute the daily step count under the assumption that each of the 21 step counts (for the 7 days at baseline, Year 1, and Year 2) are jointly normally distributed, and dependent on gender, age, and treatment arm, and further assuming that the data are MAR. We impute the missing values separately within each arm with the EM algorithm and also using MI ($M = 30$ imputations). Both methods are implemented using the R package `norm`.

Figure 4 displays the 95% CIs for the difference in average step count for each intervention compared with usual care for Year 1 and Year 2. While the point estimates for the differences between individual therapy and usual care are larger than that between group therapy and usual care within each year, and the point estimates for Year 1 are larger than those for Year 2, neither intervention is effective at the 5% significance level. Importantly, the CIs provided by the EM algorithm are smaller than those obtained by MI, consistent with the results of the simulation study. Detailed results are provided in the [Appendix](#), which illustrate that the SEs of all effects are lower when using the EM algorithm compared with MI. These differences between the two methods are nontrivial; in our study, we found that the length of the 95% CIs for the differences in average step count are between 11.7% and 13.7% lower when the EM algorithm is used instead of MI. Such differences could potentially lead to different conclusions in other studies.

Discussion

While the theoretical advantages of using MI over single imputation are well known, how this plays out in practice is less clear, especially when the relatively sophisticated EM algorithm is used for single imputation. Therefore, despite the fact that guidance on

Table 1 Percentage of the Participants Who Have Missing Step Counts (Defined as Less Than 540 min of Wear Time per Day) on Each Day at Each Year

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Baseline	18.6	25.4	29.2	31.4	29.9	33.4	28.4
Year 1	36.6	28.3	40.2	39.4	42.3	46.7	44.0
Year 2	35.0	37.5	40.0	38.2	40.5	43.7	42.3

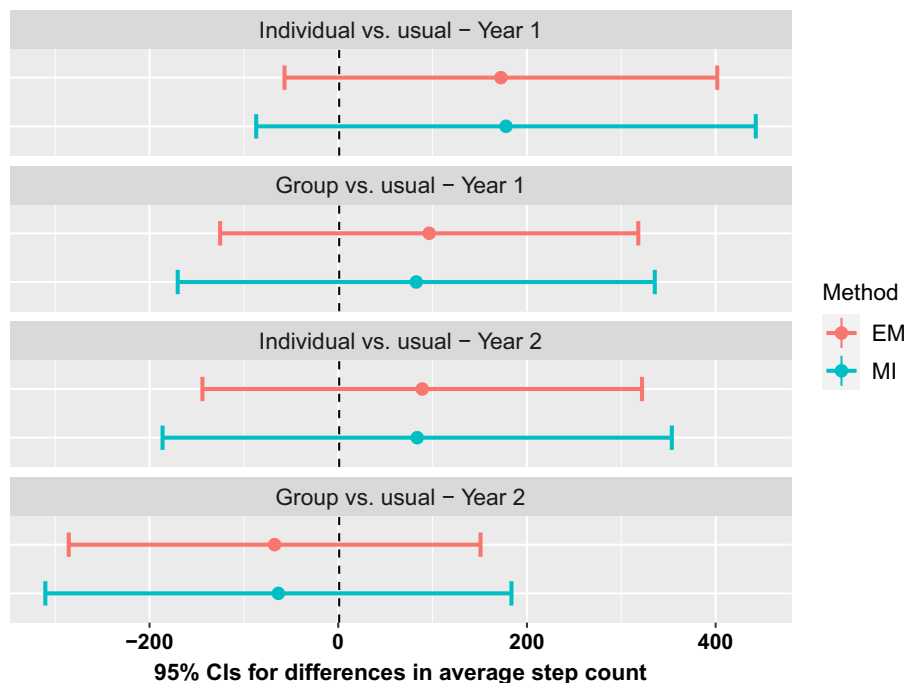


Figure 4 — Forest plot showing 95% confidence intervals for the difference in average step count per week for individual therapy versus usual care and group therapy versus usual care for Years 1 and 2. Missing values have been imputed using the EM algorithm (shown in red, at the top of each panel) and with MI (shown in blue, at the bottom of each panel). CI = confidence interval; MI = multiple imputation; EM = expectation–maximization.

handling missing data for clinical trials (e.g., by Dziura et al., 2013, and Jakobsen et al., 2017) caution against the use of single imputation, it is important to critically compare the two methods in a practically relevant context derived from a real clinical trial with accelerometer outcomes.

In this paper, we therefore evaluated two approaches to handling missing data in accelerometer outcomes; single imputation of missing values using the EM algorithm (advocated by Catellier et al., 2005), and MI (Carpenter & Kenward, 2013; Rubin, 1976). Specifically, we compared the two approaches in a simulation study of a simple trial setting where the outcome is a daily step count and the data are MCAR. The results showed that the EM algorithm leads to a practically important underestimation of the variance of the treatment effect, and also reduced coverage probability; the extent of these two issues increases with increased proportion of missingness in the data.

We also compared the two approaches in the analysis of the MOVE-IT trial. In this more complex setting, the outcome is the average of seven consecutive days of step counts. Our analysis assumes that the data are MAR. Again, we found that the SEs of all effects are lower when using the EM algorithm compared with using MI; in consequence, using the EM algorithm can lead to an increase in Type I error. Similar results were found in an observational study of accelerometer outcomes (Avtar et al., 2019).

In applications, valid imputation of missing accelerometer outcome data requires careful consideration of a number of further issues. First, defining missingness for accelerometer outcomes is a complex task with no consensus (Lee & Gill, 2018). Second, analysis by MI typically benefits from the inclusion of carefully selected auxiliary variables which must be good predictors of the missing accelerometer values. If they also predict the chance of those values being observed, they may correct for any bias (Carpenter & Kenward, 2013, p. 64). Inclusion of auxiliary variables can improve plausibility of the MAR assumption. Third, analyses typically assume that the data are MAR; sensitivity analyses to explore the impact of deviation from this assumption on the results should be conducted (Carpenter & Smuk, 2021; Cro et al., 2020). For a practically grounded discussion of these issues, we refer readers to a framework for handling missing accelerometer data (Tackney et al., 2021).

In summary, our results, together with theoretical considerations, show that it's time to step away from the EM algorithm for missing step count data.

Acknowledgments

Tackney and Stahl were supported by Health Data Research UK, which is funded by the U.K. Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), and British Heart Foundation and Wellcome. Stahl received financial support by the National Institute for Health Research Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust and King's College London. Carpenter is supported by the Medical Research Council, grant numbers MC UU 12023/21 and MC UU 12023/29. Williamson is supported by Medical Research Council project grants MR/S01442X/1 and MR/R013489/1. The MOVE-IT trial was funded by the National Institute for Health Research Health Technology Assessment program (Project: 10/62/03). The views expressed are those of the

author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. The MOVE-IT trial was reviewed and approved by the Dulwich Ethics Committee (reference: 12/LO/0917). Written informed consent was gained from all participants prior to undergoing screening in order to validate their eligibility to participate.

References

- Ae Lee, J., & Gill, J. (2018). Missing value imputation for physical activity data measured by accelerometer. *Statistical Methods in Medical Research*, 27(2), 490–506. <https://doi.org/10.1177/0962280216633248>
- Avtar, S.S., Khuneswari, G.P., Abdullah, A.A., McColl, J.H., Wright, C., & Team, G.M.S. (2019). Comparison between EM algorithm and multiple imputation on predicting children's weight at school entry. *Journal of Physics: Conference Series*, 1366(1), Article 012124. <https://doi.org/10.1088/1742-6596/1366/1/012124>
- Bayley, A., de Zoysa, N., Cook, D.G., Whincup, P.H., Stahl, D., Twist, K., ... Ismail, K. (2015). Comparing the effectiveness of an enhanced MOtiVational intErviewing InTervention (MOVE IT) with usual care for reducing cardiovascular risk in high risk subjects: Study protocol for a randomised controlled trial. *Trials*, 16, 112. <https://doi.org/10.1186/s13063-015-0593-5>
- Borghese, M.M., Borgundvaag, E., McIsaac, M.A., & Janssen, I. (2019). Imputing accelerometer nonwear time in children influences estimates of sedentary time and its associations with cardiometabolic risk 11 medical and health sciences 1117 public health and health services. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 1–12. <https://doi.org/10.1186/s12966-019-0770-0>
- Bravata, D.M., Smith-Spangler, C., Sundaram, V., Gienger, A.L., Lin, N., Lewis, R., Stave, C.D., Olkin, I., & Sirard, J.R. (2007). Using pedometers to increase physical activity. A systematic review. *JAMA*, 298(19), 2296–2304. <https://doi.org/10.1001/jama.298.19.2296>
- Carpenter J., & Kenward, M. (2013). *Multiple imputation and its application*. John Wiley & Sons. <https://doi.org/10.1002/9781119942283.ch5>
- Carpenter, J.R., & Smuk, M. (2021). Missing data: A statistical framework for practice. *Biometrical Journal*, 63(5), 915–947. <https://doi.org/10.1002/bimj.202000196>
- Catellier, D.J., Hannan, P.J., Murray, D.M., Addy, C.L., Conway, T.L., Yang, S., & Rice, J.C. (2005). Imputation of missing data when measuring physical activity by accelerometry. *Medicine & Science in Sports & Exercise*, 37(Suppl. 11), 555–562. <https://doi.org/10.1249/01.mss.0000185651.59486.4e>
- Cro, S., Morris, T.P., Kenward, M.G., & Carpenter, J.R. (2020). Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Statistics in Medicine*, 39(21), 2815–2842. <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8569>
- Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- Dziura, J.D., Post, L.A., Zhao, Q., Fu, Z., & Peduzzi, P. (2013). Strategies for dealing with missing data in clinical trials: From design to analysis. *Yale Journal of Biology and Medicine*, 86(3), 343–358.
- Grund, S., Robitzsch, A., & Luedtke, O. (2019). Mitml: Tools for multiple imputation in multilevel modeling. <https://cran.r-project.org/package=mitml>
- Harris, T., Kerry, S.M., Limb, E.S., Furness, C., Wahlich, C., Victor, C.R., ... Cook, D.G. (2018). Physical activity levels in adults and older adults 3–4 years after pedometer-based walking interventions: Long-term follow-up of participants from two randomised controlled trials

- in UK primary care. *PLoS Medicine*, 15(3), Article e1002526. <https://doi.org/https://dx.doi.org/10.1371/journal.pmed.1002526>
- Harris, T., Kerry, S.M., Limb, E.S., Victor, C.R., Iliffe, S., Ussher, M., ... Cook, D.G. (2017). Effect of a primary care walking intervention with and without nurse support on physical activity levels in 45- to 75-year-olds: The Pedometer and Consultation Evaluation (PACE-UP) cluster randomised clinical trial. *PLoS Medicine*, 14(1), 1–19. <https://doi.org/10.1371/journal.pmed.1002210>
- Harris, T., Kerry, S.M., Victor, C.R., Ekelund, U., Woodcock, A., Iliffe, S., ... Cook, D.G. (2015). A primary care nurse-delivered walking intervention in older adults: PACE (Pedometer Accelerometer Consultation Evaluation)—lift cluster randomised controlled trial. *PLoS Medicine*, 12(2), 1–23. <https://doi.org/10.1371/journal.pmed.1001783>
- IBM Corp. (2020a). Estimating statistics and imputing missing values. <https://www.ibm.com/docs/en/spss-statistics/27.0.0?topic=analysis-estimating-statistics-imputing-missing-values>
- IBM Corp. (2020b). Multiple imputation. Retrieved December 9, 2021 from <https://www.ibm.com/docs/en/spss-statistics/27.0.0?topic=edition-multiple-imputation>
- Ismail, K., Bayley, A., Twist, K., Stewart, K., Ridge, K., Britneff, E., ... Stahl, D. (2020). Reducing weight and increasing physical activity in people at high risk of cardiovascular disease: A randomised controlled trial comparing the effectiveness of enhanced motivational interviewing intervention with usual care. *Heart*, 106(6), 447–454. <https://heart.bmj.com/content/106/6/447>
- Ismail, K., Stahl, D., Bayley, A., Twist, K., Stewart, K., Ridge, K., ... Winkley, K. (2019). Enhanced motivational interviewing for reducing weight and increasing physical activity in adults with high cardiovascular risk: The MOVE IT three-arm RCT. *Health Technology Assessment*, 23(69), 1–144. <https://doi.org/10.3310/hta23690>
- Jakobsen, J.C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Medical Research Methodology*, 17(1), 1–10. <https://doi.org/10.1186/s12874-017-0442-1>
- Lee, J.A., & Gill, J. (2018). Missing value imputation for physical activity data measured by accelerometer. *Statistical Methods in Medical Research*, 27(2), 490–506. <https://doi.org/10.1177/0962280216633248>
- Little R.J.A., & Rubin, D.B. (1987). *New Developments in Autism*. Wiley. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119013563>
- Murray, D.M., Stevens, J., Hannan, P.J., Catellier, D.J., Schmitz, K.H., Dowda, M., ... Yang, S. (2006). School-level intraclass correlation for physical activity in sixth grade girls. *Medicine & Science in Sports & Exercise*, 38(5), 926–936. <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=16672847>
- National Institute for Health and Care Excellence. (2015). *Cardiovascular risk assessment and lipid modification*.
- Nissen, J., Donatello, R., & Van Dusen, B. (2019). Missing data and bias in physics education research: A case for using multiple imputation. *Physical Review Physics Education Research*, 15(2), Article 020106. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020106>
- Novo A.A., & Schafer, J.L. (2013). Norm: Analysis of multivariate normal datasets with missing values. <https://cran.r-project.org/package=norm>
- Quartagno, M., & Carpenter J. (2020). Jomo: A package for multilevel joint modelling multiple imputation. <https://cran.r-project.org/package=jomo>
- Quartagno, M., Grund, S., & Carpenter, J. (2019). Jomo: A flexible package for two-level joint modelling multiple imputation. *R Journal*, 11(2), 205–228. <https://doi.org/10.32614/rj-2019-028>
- Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.2307/2335739>
- SAS Institute Inc. (2021). SAS/stat user's guide the mi procedure 2021.1.1. https://documentation.sas.com/doc/en/pgmsascdc/v_012/statug/statug_mi_overview.htm?homeOnFail
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall.
- Snijders, T.A.B., & Bosker, R.J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. <https://books.google.co.uk/books?id=N1BQvcomDdQC>
- Statacorp. (2021). Stata multiple-imputation reference manual release 17. Retrieved December 9, 2021 from <https://www.stata.com/manuals/mi.pdf>
- Tackney, M.S., Cook, D.G., Stahl, D., Ismail, K., Williamson, E., & Carpenter, J. (2021). A framework for handling missing accelerometer outcome data in trials. *Trials*, 22(1), 1–18. <https://doi.org/10.1186/s13063-021-05284-8>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Vink, G., & van Buuren, S. (n.d.). miceVignettes. Retrieved December 9, 2021 from <https://www.gerkovink.com/miceVignettes/>
- von Hippel, P.T. (2018). How many imputations do you need. *Sociological Methods & Research*, 49(3), 1–17. <https://doi.org/10.1177/0049124117747303>
- Xu, X., Tupy, S., Robertson, S., Miller, A.L., Correll, D., Tivis, R., & Nigg, C.R. (2018). Successful adherence and retention to daily monitoring of physical activity: Lessons learned. *PLoS One*, 13(9), 1–14. <https://doi.org/10.1371/journal.pone.0199838>
- Yuan, Y. (2011). Multiple imputation for missing data: Concepts and new development (version 9.0). <https://support.sas.com/rnd/app/stat/papers/multipleimputation.pdf>

Appendix

Details of the EM Algorithm and MI

We describe the technical details of the EM algorithm and MI approach to handling missing data in the accelerometer setting. Suppose that there are n participants in a trial, and participants provide data on p variables. Some of these variables are step counts which may have missing values. The $n \times p$ matrix Y denotes the matrix of data on all participants, which can be partitioned into an observed component Y^{obs} , and a missing component Y^{mis} . We assume that the missing accelerometer data are MAR, meaning that the missingness depends only on the observed step counts and/or the covariates. Given the observed step counts and covariates, the missing values do not depend on the unobserved pattern of step counts. We can then model the joint model for the data Y as a multivariate distribution, $Y \sim \text{MVN}(\theta, \Sigma)$, where θ is the vector of means for the j variables, and Σ is the variance–covariance matrix (Schafer, 1997, Ch. 2).

The EM algorithm is an iterative procedure for finding the maximum likelihood of θ in the presence of missing data. The maximum likelihood estimate is then used to replace each missing value with a single plausible value. More specifically, given an initial guess for θ , the EM algorithm iterates between an expectation step and a maximization step (Dempster et al., 1977; Little & Rubin, 1987; Schafer, 1997, Ch. 3). In the expectation step, the expectation of the log-likelihood function for θ with respect to the conditional distribution of Y^{mis} given Y^{obs} and the current value of θ is obtained.

In the maximization step, the parameter values that maximize this expectation are computed. These two steps are iterated until a convergence criterion is met; typically, this criterion is when successive estimates of θ vary by less than 0.0001. The resulting maximum likelihood estimate is used to draw plausible values for the missing data. Importantly, each missing value is imputed a single time. The imputed data set is then analyzed as if it were the observed data; the uncertainty due to the missing values is disregarded.

Multiple imputation creates M completed data sets, called imputed data sets, by replacing missing values by plausible values. Each imputed data set is analyzed separately and then combined in a pooling step to take into account the uncertainty of the imputed values. We describe the joint modeling approach here, though the alternative method of full conditional specification, also known as imputation using chained equations, can be taken (Carpenter & Kenward, 2012, p. 85). Given an initial guess for the parameter values θ , the algorithm iterates between two steps; in the first step, missing values are imputed by draws from the predictive distribution of Y^{mis} conditional on Y^{obs} and the current value of θ . In the second step, θ is updated by drawing from the distribution of θ given the observed data and current imputed values for Y^{mis} . Repeating these two steps leads to a stochastic sequence of values for θ and Y^{mis} . After a sufficiently large number of iterations, the stochastic sequence reaches a stationary distribution. Imputations of Y^{mis} are drawn from the stationary distribution M times to obtain M sets of complete data. The analysis model is fitted to each of the M imputed data sets, and the results are combined using Rubin’s rules (Rubin, 1976).

Details of Imputation in Simulation Study

The EM algorithm is implemented using the R package norm (Novo and Schafer, 2013). Default starting values are used, which set the mean on the transformed scale to be a vector of zeros, and

the covariance matrix on the transformed scale to be the identity matrix. The convergence criterion is such that the algorithm stops when the maximum relative differences in the estimated means, variances, and covariances between two iterations differ by no more than 0.0001.

The MI is conducted via joint modeling using the R packages JOMO (Quartagno & Carpenter, 2020) and mitml (Grund et al., 2019). The number of burn-in iterations is set to 1,000, the number of iterations between successive imputations is set to 1,000, and we use 30 imputations.

Details of MOVE-IT Analysis Model

Suppose that $y_{i,j,k}$ is the step count for patient i at year j on day k , where $i \in \{1, \dots, n\}$, $j \in \{0, 1, 2\}$ to represent baseline, Year 1, and Year 2, respectively, and $k \in \{1, 2, \dots, 7\}$. We define $\bar{y}_{i,j,\cdot}$ as the average step count across seven consecutive days for patient i at year j :

$$\bar{y}_{i,j,\cdot} = \frac{1}{7} \sum_{k=1}^7 y_{i,j,k},$$

and we denote by arm_i the arm that participant i was assigned to, which takes values 1, 2, or 3, corresponding to individual therapy, group therapy, and usual care, respectively. We regress the average step count at Year 1 and Year 2 on a baseline–year interaction; a treatment–year interaction; and fixed effects for sex and age, and so can be written, for $j \in \{1, 2\}$:

$$\begin{aligned} \bar{y}_{i,j,\cdot} = & \beta_0 + \beta_1 \text{year}2_{ij} + \beta_2 I(\text{arm}_i = 1) + \beta_3 I(\text{arm}_i = 2) \\ & + \beta_4 \text{year}2_{ij} \times I(\text{arm}_i = 1) + \beta_5 \text{year}2_{ij} \times I(\text{arm}_i = 2) \\ & + \beta_6 \bar{y}_{i,0,\cdot} + \beta_7 \text{year}2_{ij} \times \bar{y}_{i,0,\cdot} + \beta_8 \text{female}_i + \beta_9 \text{age}_i + e_{ij}, \end{aligned} \quad (1)$$

where covariates in the model include:

- year2, the dummy variable for whether the observation is from Year 2.
- female, the dummy variable for whether the participant is female.
- age, the age of the participant at baseline in years (centered).

To take account of the effect of the intervention on the variance, the primary analysis also specifies an unstructured correlation matrix for the residuals by arm:

$$(e_{i,1}, e_{i,2})^T \sim N(0, \Sigma_{\text{arm}_i}),$$

where the variance–covariance matrix is given by

$$\Sigma_{\text{arm}_i} = \begin{bmatrix} \sigma_{1,\text{arm}_i}^2 & \sigma_{12,\text{arm}_i} \\ \sigma_{21,\text{arm}_i} & \sigma_{2,\text{arm}_i}^2 \end{bmatrix},$$

where $\sigma_{j,\text{arm}_i}^2$ denotes the variance of residuals for year j for patients assigned to arm_i , and $\sigma_{12,\text{arm}_i} = \sigma_{21,\text{arm}_i}$ denotes the covariance between residuals at Year 1 and Year 2 for patients assigned to arm_i .

Details of Imputation Model

We assume that the step counts for each day for each year are jointly normally distributed, and dependent on gender, age, and arm:

$$\begin{aligned}
 y_{i,0,1} &= \alpha_{01} + \gamma_{01}\text{female}_i + \delta_{01}\text{age}_i + \kappa_{01}\text{arm}_i + e_{01,i}, \\
 &\vdots \\
 y_{i,0,7} &= \alpha_{07} + \gamma_{07}\text{female}_i + \delta_{07}\text{age}_i + \kappa_{07}\text{arm}_i + e_{07,i}, \\
 y_{i,1,1} &= \alpha_{11} + \gamma_{11}\text{female}_i + \delta_{11}\text{age}_i + \kappa_{11}\text{arm}_i + e_{11,i}, \\
 &\vdots \\
 y_{i,1,7} &= \alpha_{17} + \gamma_{17}\text{female}_i + \delta_{17}\text{age}_i + \kappa_{17}\text{arm}_i + e_{17,i}, \\
 y_{i,2,1} &= \alpha_{21} + \gamma_{21}\text{female}_i + \delta_{21}\text{age}_i + \kappa_{21}\text{arm}_i + e_{21,i}, \\
 &\vdots \\
 y_{i,2,7} &= \alpha_{27} + \gamma_{27}\text{female}_i + \delta_{27}\text{age}_i + \kappa_{27}\text{arm}_i + e_{27,i},
 \end{aligned}$$

where, $(e_{01,i} \dots e_{07,i} \ e_{11,i} \dots e_{17,i} \ e_{21,i} \dots e_{27,i}) \sim N(0, \Sigma)$.

Detailed Results of MOVE-IT Analysis

Table A1 provides point estimates and standard errors for the coefficients in the primary analysis model (Equation 1) using MI versus using the EM algorithm. We observe that the standard errors

are lower when the EM algorithm is used, for all effects. Table A2 provides estimates of the variances of the residuals under both approaches; these estimates are similar under the two approaches, as expected.

Table A2 Residual Variances of the Analysis of the MOVE-IT Trial Where Missing Values Have Been Imputed Using MI and by the EM Algorithm

	MI	EM algorithm
arm 1	$\begin{bmatrix} 2,021.4 & 0.487 \\ 0.487 & 1,961.3 \end{bmatrix}$	$\begin{bmatrix} 2,061.8 & 0.503 \\ 0.503 & 1,969.6 \end{bmatrix}$
arm 2	$\begin{bmatrix} 2,195.1 & 0.565 \\ 0.565 & 1,970.4 \end{bmatrix}$	$\begin{bmatrix} 2,243.0 & 0.559 \\ 0.559 & 1,992.4 \end{bmatrix}$
arm 3	$\begin{bmatrix} 1,737.3 & 0.573 \\ 0.573 & 1,868.6 \end{bmatrix}$	$\begin{bmatrix} 1,674.2 & 0.562 \\ 0.562 & 1,844.5 \end{bmatrix}$

Note. MI = multiple imputation; EM = expectation-maximization.

Table A1 Fixed Effect Estimates, SEs, and p Values of the Analysis of the MOVE-IT Trial Where Missing Values Have Been Imputed Using MI and by the EM Algorithm

	MI			EM algorithm		
	Estimate	SE	p	Estimate	SE	p
arm1	177.8	135.0	.19	172.4	117.0	.14
arm2	82.69	128.9	.52	96.42	113.1	.39
year2	-11.93	160.4	.94	-70.47	130.6	.59
arm1 year2	-94.13	134.8	.49	-83.36	114.6	.47
arm2 year2	-146.1	122.0	.23	-163.9	105.5	.12
$\bar{y}_{i,0..}$	0.788	0.0198	.00	0.781	0.0171	.00
year2 $\bar{y}_{i,0..}$	-0.0434	0.0199	.03	-0.0334	0.0160	.04
female	-429.4	139.1	.00	-361.2	0.49	-83.36
age	-30.33	12.15	.01	-45.53	0.23	-163.9
intercept	1,447.0	163.3	.00	1,519.2	0.00	.781

Note. MI = multiple imputation; EM = expectation-maximization.