Contemporary Clinical Trials Communications 20 (2020) 100684

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

Intention-to-treat analysis when only a baseline value is available

Jos WR. Twisk [a,*], Judith JM. Rijnhart [a], Trynke Hoekstra [b], Noah A. Schuster [a], Marieke M. ter Wee [a], Martijn W. Heymans [a]

[a] Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, UMC Amsterdam, the Netherlands
[b] Department of Health Science, Amsterdam Public Health Research Institute, Faculty of Science, VU University, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Intention to treat analysis
Randomised controlled trial
Mixed model analysis
Single imputation
Multiple imputation
Selective imputation

ABSTRACT

Objectives: How to perform an intention to treat (ITT) analysis when a patient has a baseline value but no follow-up measurements is problematic. The purpose of this study was to compare different methods that deal with this problem, i.e. no imputation (standard and alternative mixed model analysis), single imputation (i.e. baseline value carried forward), and multiple imputation (selective and non-selective).

Study design and setting: We used a simulation study with different scenarios regarding 1) the association between missingness and the baseline value, 2) whether the patients did or did not receive the treatment, and 3) the percentage of missing data, and two real life data sets.

Results: Bias and coverage were comparable between the two mixed model analyses and multiple imputation in most situations including the real life data examples. Only in the situation when the patients in the treatment group were simulated not to have received the treatment, selective imputation using this information outperformed all other methods.

Conclusions: In most situations a standard mixed model analysis without imputation is appropriate as ITT analysis. However, when patients with missing follow-up data allocated to the treatment group did not received treatment, it is advised to use selective imputation, using this information, although the results should be interpreted with caution.

1. Introduction

The standard method to estimate treatment effects in a randomised controlled trial (RCT) is an intention-to-treat analysis. In an intention-to-treat analysis, all patients randomised into the treatment condition should be analysed as being treated, regardless of receiving the complete treatment, only partly or nothing at all.

The general principle of intention-to-treat analysis is widely recognised. However, in most methods dealing with the analyses of RCT data, the follow-up measurement(s) are used as the outcome, whereas the baseline value is used as a covariate. When using these methods, a problem arises when a baseline measurement is available for a particular patient, while all follow-up measurements are missing. The intention-to-treat principle states that these patients should be analysed according to their assigned condition. Yet, in an analysis adjusted for the baseline value, the data of these patients cannot be included in the analysis. There is a lot of discussion going on about how to deal with these patients. Some epidemiologists argue that data of these patients should not be taken into account in the analysis, as no data is available on the follow-up measurements after treatment initiation. Others argue that not including data of these patients in the analysis drives against the principle of intention-to-treat and leads to bias in the effect estimates. In other words, the definition of intention-to-treat and how to deal with this principle in statistical analyses of RCT data remains unclear [1–9].

Several suggestions are provided in the literature on analysing RCT data to deal with the above mentioned problem. The most classical solution is to impute the follow-up measurement(s) with the baseline value carried forward [5]. Although highly criticized, this method is still widely used [10]. Multiple imputation using more complicated imputation methods (such as predictive mean matching) are suggested as a better alternative [11,12]. If there is more than one follow-up measurement available, a mixed model analysis can be performed to estimate treatment effects. In mixed model analysis, data of patients with a baseline measurement but missing follow-up measurements are mostly ignored (i.e., are not part of the analysis). This is based on the idea that the use of a mixed model analysis (adjusting for the baseline value) is

* Corresponding author. Department of Epidemiology and Biostatistics, UMC Amsterdam, location VUMC, PO Box 7057, 1007 MB, Amsterdam, the Netherlands.
E-mail address: jwr.twisk@amsterdamumc.nl (J.WR. Twisk).

https://doi.org/10.1016/j.conctc.2020.100684
Received 26 June 2020; Received in revised form 29 October 2020; Accepted 22 November 2020
Available online 26 November 2020

enough to deal with the missing data. However, although it is true that a mixed model analysis is suitable for the analysis of longitudinal data with outcome missingness [13,14], the cases with only a baseline value are not included in the analysis, because no follow-up outcome measurements are available. Because of this phenomenon, an alternative mixed model analysis has been proposed in which the baseline value is part of the longitudinal outcome and the model is estimated without the inclusion of the treatment variable, but with including time and the interaction between treatment and time. Due to the fact that the treatment variable is not part of the model, the intercept of such an analysis reflects the combined baseline value for both the treatment and control group. In this alternative mixed model analysis, the regression coefficient for the interaction between treatment and time reflects the treatment effect [13,15,16]. If the follow-up outcome measurements for a particular patient are only partly missing, there is no problem in applying a mixed model analysis (adjusting for the baseline value) to deal with the missing data [17].

Because there remains heterogeneity in applied methods to deal with the problem of missing data on all follow-up measurements while the baseline value is available, the purpose of this study is to compare the performance of different methods to deal with this problem. The methods that are evaluated in this study are all frequently used in practice.

2. Methods

2.1. Missing data scenarios

Two scenarios for missing data were evaluated: 1) missing completely at random, and 2) missing at random, in which missingness was associated with the baseline value of the outcome. Both scenarios were evaluated in two situations: 1) a situation where there is no information about whether or not the patients with missing data in the treatment group actually received the treatment, and 2) a situation in which the subjects with missing data in the treatment group did not receive the treatment. For all scenarios, 5% missing, 10% missing, 20% missing and 40% missing were evaluated.

2.2. Simulations

We simulated longitudinal RCT datasets with a normally distributed outcome variable. The simulation set up included two follow-up measurements on 150 patients. A dichotomous treatment variable was created that randomly and evenly assigned the simulated patients to either the control group or the treatment group. The parameters used for the simulation set up were derived from one of the datasets used in the real-life data examples. Both the random intercept variance and the residual variance were about 0.5 (i.e. 0.7 squared), so the intraclass correlation coefficient between the two repeated measurements is equal to 0.5. The baseline value was generated from the value at the first follow-up measurement with a regression coefficient of 0.6. The model which was used for the simulation included an intercept of 1.5 and a regression coefficient for treatment of 1. For each condition 500 samples were generated. The simulations were performed with STATA and Box 1 and box 2 show the syntax used for the simulations.

To evaluate the performance of the different methods, we examined bias, and coverage probability of regression coefficients. Bias was determined by comparing the difference between the true treatment effect and the treatment effects estimated with the different methods. The percentage of times that the confidence interval of the estimated treatment effect included the true treatment effect was used to assess the coverage probability. Since a 95% confidence interval was used, the ideal coverage should have a score of 95% [18].

2.3. Statistical methods

We compared the following methods with each other: 1) no imputation, 2) single imputation, using the baseline value carried forward,

Box 1
Syntax used to simulate RCT datasets with complete data

```
Clear
program sim1
drop _all
set obs 150 // 150 patients
gen pat_id=_n // patient id
generate nu0 = sqrt(0.8)*rnormal() // random intercept per patient
gen treatment_c= runiform() // generate the treatment variable
gen treatment = treatment_c // copy of the treatment variable
replace treatment=0 if treatment_c <0.5 // generate the control condition
replace treatment=1 if treatment_c >0.5 // generate the treatment condition
expand 2 // 2 follow-up measurements per patient
bysort pat_id: gen index = _n
generate time = index - 1 // generate a time variable of 0 and 1
gen residual = sqrt(0.7)* rnormal() // residual variance
gen true_outcome =  1*treatment + 1.5 // generate the true outcome
gen outcome = true_outcome + nu0 + residual // generate the outcome
drop index residual nu0 true_outcome // drop variables not needed
reshape wide outcome, i(pat_id) j(time) // reshape file to wide format
gen baseline_outcome = 0.6*outcome0 + invnorm(uniform()) // generate the baseline
value based on the first follow-up measurement
reshape long outcome, i(pat_id) j(time) // reshape to long format
mixed outcome treatment baseline_outcome|| pat_id: // mixed model analysis
end
simulate _b _se, seed(12345) reps(500): sim1
```

Box 2
Syntax used to simulate RCT datasets with MAR data

```
clear
program sim2
drop _all
set obs 150 // 150 patients
gen pat_id=_n // patient id
generate nu0 = sqrt(0.8)*rnormal() // random intercept per patient
gen treatment_c= runiform() // generate the treatment variable
gen treatment = treatment_c // copy of the treatment variable
replace treatment=0 if treatment_c <0.5 // generate the control condition
replace treatment=1 if treatment_c >0.5 // generate the treatment condition
expand 2 // 2 follow-up measurements per patient
bysort pat_id: gen index = _n
generate time = index - 1 // generate a time variable of 0 and 1
gen residual = sqrt(0.7)* rnormal() // residual variance
gen true_outcome =  1*treatment + 1.5 // generate the true outcome
gen outcome = true_outcome + nu0 + residual // generate the outcome
drop index residual nu0 true_outcome // drop variables not needed
reshape wide outcome, i(pat_id) j(time) // reshape file to wide format
gen baseline_outcome = 0.6*outcome0 + invnorm(uniform()) // generate the baseline
value based on the first follow-up measurement

gen p0 = invlogit(-3.6 +.5*baseline_outcome) if treatment==0
replace outcome0 = . if uniform() < p0 & treatment==0 // generating missing
values for the control condition related to the baseline
gen p1 = invlogit(-3.6 +.5*baseline_outcome) if treatment==1
replace outcome0 = . if uniform() < p1 & treatment==1// generating missing values
for the treatment condition related to the baseline
replace outcome1 = . if outcome0 == . // generating missing values for the second
follow-up

reshape long HAQ, i(pat_id) j(time) // reshape to long format
mixed HAQ treatment baseline_HAQ || pat_id: / mixed model analysis
end
simulate _b _se, seed(12345) reps(500): sim2
```

and 3) multiple imputation using predictive mean matching [11,12]. The imputation model included the baseline value of the outcome and the treatment variable and in all situations 20 imputations were performed. Additionally, in the scenarios where the patients who were randomized to the treatment group with missing data on follow-up measurements did not receive the treatment a selective imputation model was used in which data of these patients were imputed as if they were control patients.

For all situations, mixed model analysis was used to estimate treatment effects. Regarding no imputation, two different analyses were performed: 1) a standard mixed model analysis with the follow-up measurements as the outcome adjusting for the baseline value. In this first analysis, the patients with a baseline measurement but without all follow-up measurements were ignored (i.e., these patients were not part of the analysis) (Equation (1)).

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0} \quad (1)$$

where, Y_t = the outcome measured at the follow-up measurement(s), X = treatment variable, β_1 = overall treatment effect, β_2 = is the effect of the outcome variable measured at baseline, and Y_{t0} = outcome variable measured at baseline.

2) an alternative mixed model analysis in which both the baseline value and the follow-up measurements were used as the outcome meaning that also patients with only a baseline value were included in the analysis. This model consisted of time and the interaction between treatment and time [13,15] (Equation (2)).

$$Y_t = \beta_0 + \beta_1 time + \beta_2 time \times X \quad (2)$$

where, Y_t = the outcome measured at the baseline and follow-up measurement(s), X = treatment variable, time = the time variable, β_1 = effect of time for the treatment group coded 0, and β_2 = overall treatment effect.

2.4. Real-life examples

The first real-life example is taken from an intervention study in which the effectiveness of a long-term homocysteine-lowering treatment with folic acid plus pyridoxine in reducing systolic blood pressure was evaluated [19]. In this 2-year, randomized, placebo-controlled trial, a baseline measurement and two follow-up measurements (after 1 year and after 2 years) were performed. At each time-point systolic blood pressure was measured four times and the average value was used in the analysis. In this first example 6% of the patients only had a baseline value.

The second real-life example is taken from a study by Warmerdam et al. [20,21] in which an RCT was performed to compare internet based cognitive behavioral therapy (CBT), internet based problem solving therapy (PST) and a waiting list control group (WL) regarding treatment of patients with depressive symptoms. In this study, besides a baseline measurement, measurements were taken at 5, 8 and 12 weeks after baseline and both depression and anxiety were used as outcome variables. In this second example 22% of the patients only had a baseline

Table 1
Results (bias and coverage) of the simulation analysis: missing completely at random.

		No imputation		Imputation	
% missing	Real values[a]	Standard mixed model	Alternative mixed model	Single (BVCF)	Multiple (PMM)
5%	0.723 (0.153)	0.722 (0.157)	0.698 (0.148)	0.689 (0.154)	0.719 (0.157)
Bias		−0.001	−0.025	−0.034	−0.004
Coverage		94%	93%	94%	93%
10%	0.723 (0.153)	0.722 (0.161)	0.698 (0.152)	0.654 (0.155)	0.724 (0.161)
Bias		−0.001	−0.025	−0.069	0.001
Coverage		95%	93%	91%	94%
20%	0.723 (0.153)	0.720 (0.171)	0.696 (0.161)	0.578 (0.154)	0.727 (0.169)
Bias		−0.003	−0.027	−0.145	0.004
Coverage		94%	92%	82%	93%
40%	0.723 (0.153)	0.722 (0.196)	0.696 (0.187)	0.436 (0.147)	0.701 (0.191)
Bias		−0.001	−0.027	−0.287	−0.022
Coverage		93%	93%	47%	91%

BVCF: baseline value carries forward; PMM: predictive mean matching.
[a] Numbers are regression coefficients and standard errors (between brackets).

value.

Regarding the real-life data examples the same statistical methods were applied as for the simulated datasets and all analyses were performed with STATA (version 15).

3. Results

3.1. Simulations

Tables 1–4 show the results of the simulations. It should be realised that the real treatment effect is basically an empirical true value and not equal to 1. This is because it is based on an analysis adjusted for the baseline value of the outcome. When missing is completely at random, the standard mixed model analyses and multiple imputation performed well in all situations, while the alternative mixed model analyses performed slightly worse and showed underestimated regression coefficients. Single imputation based on the baseline value carried forward performed much worse (Table 1). When missingness was associated with the baseline value, both mixed model analyses performed equally well, although the estimates with standard mixed model analyses were slightly overestimated, while the estimates with the alternative mixed model analyses were slightly underestimated. Multiple imputation performed slightly better than both mixed model analyses, although the differences were small. Single imputation based on the baseline value carried forward, again, did not perform well (Table 2).

In the situations in which the subjects with missing data in the treatment group did not receive the treatment the results of the simulations were quite different (Tables 3 and 4). Both mixed model analyses lead to highly overestimated regression coefficients, irrespective whether the missing values were completely at random or related to the baseline value. The performance became worse with increasing percentage of missing data. In all situations, the alternative mixed model analyses performed slightly better than the standard mixed model analyses. Multiple imputation also lead to overestimated regression coefficients, but performed well in the situation where the missing values were related to the baseline value. Single imputation based on the baseline value carried forward, on the other hand, performed better than both mixed model analyses and multiple imputation in all situations. Selective multiple imputation, taking into account that the patients in the treatment group did not received the treatment, performed best in all situations. However, this method also gave slightly overestimated regression coefficients in the situation in which the missing values were related to the baseline values.

3.2. Real-life examples

Tables 5 and 6 show the results of the real-life examples. In the first example, 6% of the patients had only a baseline value, but there was a huge difference between the baseline values of the two groups. Nevertheless, the regression coefficients for the standard mixed model analysis and multiple imputation were highly comparable. The alternative mixed model analysis on the other hand showed a slightly lower regression coefficient. The regression coefficient obtained from the analysis with single imputation based on the baseline value carried forward was much lower than the other estimated regression coefficients (Table 5).

In the second real-life example, 22% of the patients had only a baseline value. The baseline values of the group with only a baseline value were slightly higher than the baseline values of the patients with

Table 2
Results (bias and coverage) of the simulation analysis: missing at random, in which missing was associated with the baseline value.

		No imputation		Imputation	
% missing	Real values[a]	Standard mixed model	Alternative mixed model	Single (BVCF)	Multiple (PMM)
5%	0.723 (0.153)	0.734 (0.158)	0.709 (0.149)	0.695 (0.153)	0.725 (0.157)
Bias		0.011	−0.014	−0.028	0.002
Coverage		94%	93%	94%	95%
10%	0.723 (0.153)	0.735 (0.162)	0.709 (0.152)	0.664 (0.152)	0.736 (0.161)
Bias		0.012	−0.014	−0.059	0.013
Coverage		94%	93%	93%	93%
20%	0.723 (0.153)	0.737 (0.173)	0.711 (0.152)	0.577 (0.149)	0.719 (0.170)
Bias		0.014	−0.012	−0.146	−0.004
Coverage		94%	92%	83%	96%
40%	0.723 (0.153)	0.738 (0.196)	0.711 (0.186)	0.448 (0.144)	0.700 (0.190)
Bias		0.015	−0.012	−0.275	−0.023
Coverage		93%	93%	51%	94%

BVCF: baseline value carries forward; PMM: predictive mean matching.
[a] Numbers are regression coefficients and standard errors (between brackets).

Table 3
Results (bias and coverage) of the simulation analysis: missing completely at random. Subjects missing in the intervention group did not perform the intervention.

		No imputation		Imputation		
% missing	Real values[a]	Standard mixed model	Alternative mixed model	Single (BVCF)	Multiple (PMM)	Multiple selective (PMM)
5%	0.684 (0.154)	0.722 (0.157)	0.698 (0.148)	0.667 (0.155)	0.717 (0.156)	0.682 (0.158)
bias		0.038	0.015	−0.017	0.033	−0.002
coverage		94%	93%	94%	92%	95%
10%	0.645 (0.154)	0.722 (0.161)	0.698 (0.152)	0.614 (0.156)	0.727 (0.161)	0.649 (0.163)
bias		0.077	0.053	−0.031	0.082	0.004
coverage		92%	93%	93%	92%	98%
20%	0.568 (0.155)	0.720 (0.171)	0.696 (0.161)	0.508 (0.156)	0.729 (0.170)	0.569 (0.170)
bias		0.152	0.128	−0.060	0.161	0.001
coverage		86%	87%	93%	84%	97%
40%	0.421 (0.155)	0.722 (0.196)	0.697 (0.186)	0.338 (0.148)	0.708 (0.192)	0.418 (0.182)
bias		0.301	0.276	−0.083	0.287	−0.003
coverage		66%	69%	92%	67%	99%

BVCF: baseline value carries forward; PMM: predictive mean matching.
[a] Numbers are regression coefficients and standard errors (between brackets).

Table 4
Results (bias and coverage) of the simulation analysis: missing at random, in which missing was associated with the baseline value. Subjects missing in the intervention group did not perform the intervention.

		No imputation		Imputation		
% missing	Real values[a]	Standard mixed model	Alternative mixed model	Single (BVCF)	Multiple (PMM)	Multiple selective (PMM)
5%	0.675 (0.148)	0.725 (0.157)	0.700 (0.148)	0.685 (0.152)	0.722 (0.157)	0.691 (0.159)
bias		0.050	0.025	0.010	0.047	0.016
coverage		93%	93%	94%	94%	96%
10%	0.633 (0.150)	0.724 (0.161)	0.700 (0.152)	0.654 (0.151)	0.656 (0.153)	0.656 (0.162)
bias		0.091	0.067	0.021	0.023	0.023
coverage		92%	92%	95%	97%	97%
20%	0.517 (0.161)	0.728 (0.172)	0.702 (0.163)	0.570 (0.149)	0.575 (0.173)	0.569 (0.173)
bias		0.201	0.185	0.053	0.057	0.052
coverage		74%	75%	94%	97%	98%
40%	0.342 (0.163)	0.738 (0.195)	0.712 (0.185)	0.452 (0.143)	0.436 (0.183)	0.440 (0.185)
bias		0.396	0.370	0.110	0.094	0.098
coverage		46%	48%	88%	98%	97%

BVCF: baseline value carries forward; PMM: predictive mean matching.
[a] Numbers are regression coefficients and standard errors (between brackets).

Table 5
Baseline values and results of the different analyses to estimate the treatment effect in the blood pressure example study.

	Subjects with at least one follow-up (N = 130)		Subjects with only baseline (N = 9)
Baseline blood pressure	128.4 (15.4)[a]		132.2 (15.1)
No imputation		Imputation	
Standard mixed model	Alternative mixed model	Single (BCVF)	Multiple (PMM)
−3.71 (−6.78 to −0.63)[b]	−3.47 (−6.63 to −0.30)	−3.17 (−6.10 to −0.24)	−3.71 (−6.71 to −0.71)

BVCF: baseline value carries forward; PMM: predictive mean matching.
[a] Numbers are mean values and standard deviations (between brackets).
[b] Numbers are regression coefficients and 95% confidence intervals (between brackets).

one or more follow-up measurement for both outcome variables. As in the first real-life data example, the two mixed model analyses and multiple imputation showed comparable results. There is no clear pattern in the differences between the methods for the two outcome variables and the two conditions. Also comparable to the first real-life data example, the single imputation based on the baseline value carried forward, showed much lower regression coefficients (Table 6).

4. Discussion

The purpose of this paper was to compare different methods to deal with the problem of missing all follow-up measurements in an RCT while the baseline value of the outcome is available. It was shown that in most situations, the mixed model analyses were comparable to multiple imputation. Only in the situation where it is known that subjects allocated to the treatment did not perform the treatment, a (selective) multiple imputation is preferable above mixed model analyses without imputation. Surprisingly, the alternative mixed model analysis did not perform better than the standard mixed model analysis, although it is suggested that the alternative mixed model analysis should be preferred above the standard mixed model analysis in this situation. This because in the alternative mixed model analysis, all subjects are included in the analysis, while in the standard mixed model analysis the subjects with only a baseline measurement are not included in the analysis.

Although selective imputation seems to be an acceptable approach [5], it should be realised that it is questionable whether or not the analysis with selective imputation can be classified as intention-to-treat. In the selective multiple imputation method, the patients in the treatment group with missing data were imputed as if they were allocated to the control condition. And although the analysis is on an

Table 6
Baseline values and results of the different analyses to estimate the treatment effect for PST and CBT in the internet example study[a].

	Subjects with at least one follow-up (N = 205)			Subjects with only baseline (N = 58)
Baseline depression	31.34 (7.40)[b]			32.98 (7.94)
Baseline anxiety	10.45 (3.42)			11.40 (3.17)
	No imputation		Imputation	
	Standard mixed model	Alternative mixed model	Single (BVCF)	Multiple (PMM)
Depression				
- PST	−4.53 (−7.33 to −1.73)[c]	−4.66 (−6.72 to −2.61)	−2.47 (−5.08 to 0.15)	−4.64 (−7.44 to −1.83)
- CBT	−5.21 (−8.08 to −2.34)	−5.46 (−7.56 to −3.36)	−2.10 (−4.71 to 0.51)	−5.05 (−8.06 to −2.03)
Anxiety				
- PST	−1.47 (−2.46 to −0.48)	−1.39 (−2.18 to −0.61)	−0.73 (−1.64 to 0.19)	−1.68 (−2.74 to −0.62)
- CBT	−1.47 (−2.50 to −0.44)	−1.37 (−2.18 to −0.57)	−0.29 (−1.21 to 0.62)	−1.57 (−2.72 to −0.42)

PST: internet based problem solving therapy; CBT: internet based cognitive behavioral therapy.
BVCF: baseline value carries forward; PMM: predictive mean matching.
[a] The waiting list control group is used as reference group.
[b] Numbers are mean values and standard deviations (between brackets).
[c] Numbers are regression coefficients and 95% confidence intervals (between brackets).

intention-to-treat basis, i.e. all patients allocated to the treatment group are analysed as treatment, the imputation is not. On the other hand, when it is ignored that the patients with missing data on follow-up measurements did not receive the treatment, the effect estimates were highly overestimated. An overestimation of the effect estimate is not what you may expect from an intention-to-treat analysis. It should also be realised that a potential limitation of selective imputation is the fact that the patients who did not receive the treatment were imputed as they were control patients. It is however questionable whether that is correct, because in the literature there are some examples showing that treatment non adherers act differently than control patients [22,23]. So therefore, the results of the analyses with the selective imputation can be slightly invalid.

Surprisingly, the single imputation method based on the baseline value carried forward performed well in the situation where the subjects allocated to the treatment group did not receive the treatment. From the baseline value carried forward method it is often suggested that it leads to a conservative estimation of the treatment effect and is therefore an acceptable method to analyse RCT data. This is, however, not always true, but depends highly on the setting of the study. Suppose, an intervention is performed to reduce the decline in physical functioning in elderly people, the baseline value carried forward assumes no decline, which is actually a positive outcome. When, on the other hand, an intervention is performed to reduce blood pressure in hypertensive patients, the baseline value carried forward assuming no reduction can be classified as a negative outcome.

In the present study, predictive mean matching was used for multiple imputation. The advantage of predictive mean matching, when compared to other multiple imputation techniques, is that it imputes values that are observed in the dataset and are therefore much alike real values. And although it is not an issue in the present study, this makes the method suitable for normally distributed outcome variables as well as non-normally distributed outcome variables [20,21]. The general idea behind predictive mean matching is that first predictive values are generated for all cases (including the cases with no missing data). Secondly, based on the predictive values, a group of cases without missing data (in the present study a group of five cases is used) is selected that are close to the predictive values of a case with missing data. From this group of cases, one case is randomly selected and the observed value of this particular case is used for the imputation. This procedure is then repeated for all (in the present study 20) datasets.

In the real-life data examples the missing follow-up data was (highly) related to the baseline value. Nevertheless, the effect estimates of the standard mixed model analyses, the alternative mixed model analyses and multiple imputation were only slightly different. In the second real-life data example the difference in effect estimates were higher due to the fact that the percentage of missing data was much higher. Because it is not clear which of the effect estimates reflects the real treatment effect, it is suggested that sensitivity analyses should be included in the analysis of RCT data in order to obtain a more robust effect estimation [5]. Surprisingly, the results of sensitivity analyses on RCT data are almost never reported in the scientific literature and when they are reported, they are mostly performed to show the robustness of the analysis against different assumptions underlying the statistical analysis. However, sensitivity analyses can also be performed with different statistical methods and based on the real-life data examples it seems to be appropriate to report the results of different statistical methods as sensitivity analyses, especially in situations where the percentage of missing follow-up data is relatively high.

In the present paper, different methods were used to analyse RCT data in which patients had a baseline value and no follow-up data. Although less common, it is also possible that patients were not measured at baseline but do have follow-up data. In an analysis adjusted for the baseline value, these subjects are (of course) also excluded from the analysis. It is expected that the probability of having a missing baseline value is not related to the follow-up measurements, so this situation can be considered as missing completely at random (MCAR). Probably, in this situation, standard mixed model analysis, alternative mixed model analysis and multiple imputation will not lead to very different effect estimates. Also because the percentage of patients without a baseline value but with follow-up measurements will be relatively low.

In the present paper, RCT data with more than one follow-up measurement was used in the simulations and in the real-life data examples. Therefore, linear mixed model analyses were used to analyse the data. Of course, it is also possible that an RCT has only one follow-up measurement. Theoretically, it is expected that the results obtained in the present study will be the same in the situation with only one follow-up measurement. However, to investigate this we performed the same analyses on simulated data with only one follow-up measurement and on the real life data examples only using the first follow-up measurement. As expected the results of these analyses were comparable to the ones provided by the analyses on RCT data with more than one follow-up measurement. It should be noted, however, that the only difference is that in (most of) the analyses, only one follow-up measurement is analysed (adjusted for the baseline value) and therefore, standard linear regression analyses can be used instead of linear mixed model analyses.

It should be noted that in the examples used in the present paper missing data was either completely at random (MCAR) or related to the baseline value of the outcome (i.e., MAR). The assumption of using either mixed model analysis or multiple imputation is that missing is MAR. In real-life data, MCAR and MAR are probably the most common, however, missing data can also be not at random (MNAR). Although it is not possible to evaluate whether missing data is MNAR or MAR, there

are methods available that claim to appropriately take into account missing data which is MNAR. These methods, such as pattern mixture models, selection models and shared parameter models [24–28] are complicated and difficult to interpret and therefore, not often used in practice. It should also be noted that in the examples used in the present paper, missingness was only related to the baseline value of the outcome and therefore, only the baseline value was used for the imputation models. Although it is possible that missingness is related to other variables, the baseline value of the outcome is mostly by far the best predictor of the missing outcomes. So, adding other variables to the imputation models would not add much information to the imputation models and would therefore probably not change the results of the simulations presented in this paper.

5. Conclusion

In a situation where it is not clear whether the patients with missing follow-up data allocated to the treatment group actually received treatment, it seems that using a standard mixed model analysis is appropriate as an intention-to-treat analysis. Using an alternative mixed model analysis or multiple imputation does not lead to more valid results. This is irrespective whether missingness was related to the baseline value or not. However, when the patients with missing follow-up data allocated to the treatment group actually did not received treatment it is advised to use a selective imputation method using this information, although the results should be interpreted with caution. In general, it should be encouraged to perform and report the results of different sensitivity analyses.

Declaration of competing interest

None.

References

[1] S. Hollis, F. Campbell, What is meant by intention-to-treat analysis? Survey of published randomized controlled trials, BMJ 319 (1999) 670–674.

[2] C.C. Wright, J. Sim, Intention-to-treat approach to data from randomized controlled trials : a sensitivity analysis, J. Clin. Epidemiol. 56 (2003) 833–842.

[3] J. Gravel, L. Opatrny, S. Shapiro, The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? Clin. Trials 4 (2007) 350–356.

[4] P. Flyer, J. Hirman, Missing data in confirmatory clinical trials, J. Biopharm. Stat. 19 (2009) 969–979.

[5] European Medicines Agency, Guideline on missing data in confirmatory clinical trials. https://www.ema.europa.eu/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf, 2010.

[6] I.R. White, J. Carpenter, N.J. Horton, Including all individuals is not enough: lessons for intention-to-treat analysis, Clin. Trials 9 (2012) 396–407.

[7] T. Brody, Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines Is a Practical Guidebook for Those Engaged in Clinical Trial Design, Elsevier Inc, 2012, https://doi.org/10.1016/C2011-0-04245-X. Chapter 8: Intent to treat analysis vs. per protocol analysis.

[8] R. Joseph, J. Sim, R. Ogollah, M. Lewis, A systematic review finds variable use of the intention-to-treat prinicple in musculoskeletal randomized controlled trials with missing data, Journal of Clinical Epidmiology 68 (2015) 15–24.

[9] M. Mukaka, S.A. White, D.J. Terlouw, V. Mwapasa, L. Kalilani-Phiri, Is using multiple imputation better than a complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing, Trials 17 (2016) 341.

[10] S. Lydersen, Last observation carried forward, Tidsskrift for Den Norske Legeforening (Journal of the Norwegian Medical Association) 139 (2019) 845.

[11] S van Buuren, Multiple imputation of discrete and continuous data by fully conditional specification, Stat. Methods Med. Res. 16 (2007) 219–242.

[12] T.P. Morris, I.R. White, P. Royston, Tuning multiple imputation by predictive mean matching and local residual draws, BMC Med. Res. Methodol. 14 (2014) 75–87.

[13] J. Twisk, L. Bosman, T. Hoekstra, J. Rijnhart, M. Welten, M. Heymans, Different ways to estimate treatment effects in randomised controlled trials, Contemporary Clinical Trials Communications 10 (2018) 80–85.

[14] J. Twisk, Applied Mixed Model Analysis, Cambridge University Press, Cambridge, UK, 2018.

[15] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Data Analysis, Wiley, Hoboken, NJ, 2004.

[16] J.W.R. Twisk, in: Applied Longitudinal Data Analysis for Epidemiology, second ed., Cambridge University Press, UK, 2013.

[17] J. Twisk, M. de Boer, W. de Vente, M. Heymans, Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis, J. Clin. Epidemiol. 66 (2013) 1022–1028.

[18] A. Burton, D.G. Altman, P. Royston, R.L. Holder, The design of simulation studies in medical statistics, Stat. Med. 25 (2006) 4279–4292.

[19] R.A. van Dijk, J.A. Rauwerda, M. Steyn, J.W. Twisk, C.D. Stehouwer, Long-term homocysteine-lowering treatment with folic acid plus pyridoxine is associated with decreased blood pressure but not with improved brachial artery endothelium-dependent vasodilation or carotid artery stiffness: a 2-year, randomized, placebo-controlled trial, Arterioscler. Thromb. Vasc. Biol. 21 (12) (2001) 2072–2079.

[20] L. Warmerdam, A van Straten, J. Twisk, H. Riper, P. Cuijpers, Internet-based treatment for adults with depressive symptoms: randomized controlled trial, J. Med. Internet Res. 10 (2008) e44.

[21] L. Warmerdam, A van Straten, J. Jongsma, J. Twisk, P. Cuijpers, Online cognitive behavorial therapy and problem-solving therapy for depressive symptoms: exploring mechanisms of change, J. Behav. Ther. Exp. Psychiatr. 41 (2010) 64–70.

[22] I.B. Wilson, Adherence, placebo effects, and mortality, J. Gen. Intern. Med. 25 (2010) 1270–1272.

[23] E.J. Murray, B.L. Claggett, B. Granger, S.D. Solomon, Adherence-adjustment in placebo-controlled randomized trials: an application to the candesartan in heart faillure randomized trial, Contemp. Clin. Trials 90 (2020) 105937.

[24] R.J.A. Little, Pattern-mixture models for multivariate incomplete data, J. Am. Stat. Assoc. 88 (1993) 125–134.

[25] R.J.A. Little, A class of pattern-mixture models for normal incomplete data, Biometrika 81 (1994) 471–483.

[26] G. Molenberghs, M.G. Kenward, Missing Data in Clinical Studies, John Wiley & Sons, Ltd, Hoboken, New Jersey, VS, 2007.

[27] R. Tsonaka, G. Verbeke, E. Lesaffre, A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness, Biometrics 65 (2009) 81–87.

[28] M.H. Fiero, C.-H. Hsu, M.L. Bell, A pattern-mixture model approach for handling missing continuous outcome data in longitudinal cluster randomized trials, Stat. Med. 36 (2017) 4094–4105.