Journal of Clinical Epidemiology 68 (2015) 15–24

Journal of Clinical Epidemiology

A systematic review finds variable use of the intention-to-treat principle in musculoskeletal randomized controlled trials with missing data

Royes Joseph, Julius Sim, Reuben Ogollah, Martyn Lewis*

Arthritis Research UK Primary Care Centre, Research Institute of Primary Care and Health Sciences, Keele University, Keele, Staffordshire, ST5 5BG, UK

Accepted 3 September 2014; Published online 7 October 2014

Abstract

Objectives: In randomized trials, the primary analysis should be consistent with the intention-to-treat (ITT) principle and should address missing data appropriately to draw valid inferences. This review focuses on current practices relating to the ITT principle and methods to handle missing data in the major musculoskeletal journals.

Study Design and Setting: A systematic review of randomized trials published in 2010 and 2011 in five musculoskeletal journals was performed.

Results: We reviewed 91 trials: 38% performed a full ITT analysis (analyzing outcome data for all randomized participants) and 31% performed a partial ITT analysis (excluding participants with no follow-up data). The overall median dropout was 12%; 60% of trials had more than 10% dropouts, and 32% of trials had more than 20% dropouts. Among those that performed an ITT analysis, the majority adopted a form of single imputation; last observation carried forward was the designated approach in most cases. Mixed models for repeated measures and/or multiple imputations were limited to eight trials.

Conclusion: It appears that many trials reporting missing data are inappropriately analyzed and may therefore be prone to biased estimates and invalid inferences.

Keywords: Randomized controlled trial; Musculoskeletal conditions; Intention-to-treat; Missing data; Dropout; Sensitivity analysis; Systematic review

1. Introduction

Intention-to-treat (ITT) analysis is the preferred method for randomized controlled trials (RCTs) with a superiority design. The ITT principle states that an analysis should include all study participants in the groups to which they were randomized, regardless of any departures from the original assigned group [1]. This principle helps to preserve the benefits of randomization, which is intended to ensure that differences in outcome observed between groups are solely the result of the treatment [2,3], and to reduce the risk of selection bias [4,5]. In an ideal setting, all subjects enrolled in an RCT would follow the study protocol and complete their allocated treatment as detailed therein, thus contributing data that are complete in all respects [6]. However, this is rarely achieved in practice—particularly under pragmatic trial conditions [7]. Moreover, to provide an unbiased estimate of treatment effect, randomization alone is insufficient—it is also important to obtain complete data on all randomized subjects and include these in the analysis [8]. Some authors, however, describe an analysis as ITT without regard to this requirement to include data for all randomized participants in the analysis [9]. We refer to an approach that deviates from a full ITT (FITT) analysis in this way—by retaining treatment group membership as per random allocation but excluding participants with no follow-up data—as a partial ITT (PITT) analysis. (The term ‘‘modified intention-to-treat’’ has frequently been used to describe this approach [9], but this term has been criticized for being ambiguous and lacking clarity regarding the exclusion of data [10,11].)

Because of a perceived misuse of the term ‘‘intention-to-treat’’ [10–12], item 16 in the 2010 CONSORT statement was updated to include a more explicit request for group-wise details on the number of participants included in each analysis and whether the analysis was randomized by groups [12]. Non-ITT analyses such as an ‘‘as-treated’’ (AT) analysis, which groups participants according to treatment received rather than according to randomization, and a ‘‘per-protocol’’ (PP) analysis, which omits participants who do not follow the study protocol, are not protected

Funding: None.

Conflict of interest: None.

* Corresponding author. Tel.: +44-1782734849; fax: +44-1782734719.

E-mail address: a.m.lewis@keele.ac.uk (M. Lewis).

http://dx.doi.org/10.1016/j.jclinepi.2014.09.002

What is new?

Key findings

- In accordance with the intention-to-treat (ITT) principle, most trials analyzed data by the groups to which subjects were randomized regardless of the intervention received. However, many failed to obtain outcome data for all randomized subjects and/or include all subjects in the primary analysis.
- On average, the dropout rate was a little over 10%, and because most trials failed to use appropriate statistical methods to account for missing data, it is likely that descriptive data and inferential estimates of treatment effect were biased, given that missing data probably differ from reported data.

What this adds to what was known?

- Many trials are not carrying out an ITT analysis as recommended by guidelines. The violation of the ITT approach largely concerns inappropriate handling of missing data.
- The present study found sensitivity analyses to be infrequently and inappropriately used and insufficiently reported.
- It appears that only modest progress has been made, subsequent to previous reviews, in reducing the large proportion of trials that are inappropriately analyzed.

What is the implication and what should change now?

- ITT is the gold standard approach to the analysis of randomized clinical trials with hypothesis testing in respect of superiority of treatment. However, deviation from the ITT approach is common, particularly in respect of analysis of incomplete data, which may result in biased estimates and give rise to invalid inferences. Trialists should ensure that missing values are handled judiciously and apply methods of analysis that make appropriate assumptions about the missing data.

by randomization and thus may be affected by imbalance in baseline variables [13].

The basic issue in an analysis of trial data with missing values is the selection of an ITT analysis data set. White et al. [14] stated that a true ITT analysis is possible only when there is no missing outcome data. However, in practice, no matter how well designed and implemented a study, missing data are almost inevitable [15]. Hence, the benefits of randomization may be compromised; any statistical inferences, therefore, rely on additional assumptions. Incomplete outcome data can lead to problems such as loss of efficiency due to reduced sample size and—if data are missing disproportionally in each arm and/or for different reasons—bias in the estimate of treatment effect due to differences between the observed and unobserved data [16]. Therefore, a full data set requires either imputation of missing values or modeling of unobserved data [17]. Any analysis of RCTs with incomplete data is based on specific assumptions on the missing data mechanism, such as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) [18,19]. Under the MCAR mechanism, missingness is independent of both observed data (eg, baseline covariates and observed responses) and unobserved data (those observations that would have been recorded if the patients had stayed in the study). Under the MAR mechanism, missingness depends on observed data but not on unobserved data. Under the MNAR mechanism, missingness depends on unobserved data.

As trials with missing data may not retain the balance of randomization, the basis for statistical inference is lost [6,20], and there is no longer a statistical rationale to guarantee lack of bias for the estimation of the parameter and its associated confidence interval—even if the study is assumed to be free of other risks of bias, such as nonmasked evaluation. Identification of the underlying missing data mechanism is important to carry out appropriate formal analyses of data with missing values; however, it is impossible to identify this mechanism with certainty based on the observed data alone [21]. Missing data should therefore be considered at the design, conduct, and analysis stages of a trial [14,22,23]. First, trialists should attempt to minimize missing data in the first instance by following up all randomized subjects, even if they withdraw from an allocated intervention. Second, analysts should perform a primary analysis with a plausible assumption on the mechanism of missing data. Third, sensitivity analyses should explore the robustness of the results to a range of alternative plausible assumptions regarding missingness.

A few studies [24–28] have examined practices regarding the use of the ITT principle and/or the reporting and handling of missing data in RCTs published in general medical journals. Additionally, two studies have assessed these issues in RCTs in musculoskeletal conditions [29,30].

These studies found many instances in which analyses were poorly defined and described and noted variation in practice regarding the ITT principle and the handling of missing data. For example, Gravel et al. [27] evaluated 403 reports of RCTs published in 2002 in 10 medical journals and reported that 62% of the trials analyzed their primary outcome on an ITT basis. However, only 39% of trials analyzed all subjects as randomized. The study also reported that 60% of trials had at least some missing data and most of these trials (59%) excluded subjects with missing data from the primary analysis. In the musculoskeletal field, Baron et al. [29] examined the use of the ITT

approach and the rate of missing data in 81 reports of superiority RCTs assessing structural outcomes in rheumatic diseases published between 1994 and 2003. These authors reported that an FITT analysis—one that includes all subjects as randomized—was only applied in 7% of these RCTs. They also noted that almost 68% of 63 articles in which missing data information had been reported had more than 10% missing data and approximately one-third had more than 20%. However, only a quarter of the articles reported statistical methods for handling missing data.

Henschke et al. [30] reported the trend over time (1980–2008) in the quality of 157 RCTs of interventions for chronic low back pain. Their evaluation was based on 11 criteria described by Koes et al. [31], of which two are closely linked to the ITT principle. Criterion 9 is fulfilled if the percentage of withdrawals and dropouts does not exceed 20% for short-term follow-up and 30% for long-term follow-up and if such dropouts are described with reasons; for most years of the study period, more than one-third of the reviewed RCTs published in the year concerned failed to fulfill the criterion. Criterion 11 is fulfilled if all randomized patients are analyzed in the group to which they were allocated by randomization for the primary effect measurement, minus missing values, irrespective of noncompliance, and cointerventions; fewer than 60% of the RCTs fulfilled this criterion.

Previous reviews [26,28,29] have reported that listwise deletion of cases with one or more missing values was the most common approach in the primary analysis. This approach is likely to provide a biased result (unless the mechanism is MCAR), inefficient estimates (ie, estimates that have wide confidence intervals through lack of precision), and loss of statistical power [22]. These reviews also found that single imputations were widely used to create a full data set. Carrying the last observation forward was a frequently used approach among the single imputations; however, the assumption of zero change after dropout is not justifiable in most trials [22]. Several guidelines and recommendations issued on missing data in RCTs [17,23,32] have advocated, as a starting point, methods of analysis that are valid under MAR—such as methods based on multiple imputations (MI), likelihood-based methods (eg, mixed models for repeated measures [MMRMs] [33]), or moment-based methods such as weighted generalized estimating equations (weighted GEE) [34,35]. However, previous reviews [26,28,29,36] on methods of handling missing data in RCTs have found that these recommended approaches were limited to a small fraction of trials.

Pragmatic RCTs differ from the archetypal clinical trial that uses placebo-controlled methods to ensure blinding, rigid treatment procedures, and objective measurement of outcome [37]. One clinical area where pragmatic trials abound is in musculoskeletal conditions, where the aims of treatment are mainly to reduce the burden of disease and disability under everyday clinical conditions (ie, the effectiveness rather than the efficacy of a treatment); accordingly, the outcome measures are mostly subjective and relate to participants' well-being, not the quantification of laboratory data. Owing to their chronic nature, many musculoskeletal conditions necessitate long-term trials, which are prone to loss to follow-up. Each of these features may predispose to missing values.

Against this background, our study sought to examine current practice relating to ITT analysis and methods to handle missing data in published trials in musculoskeletal conditions. Specifically, the study had the following objectives:

1. To describe the extent of adherence to random allocation,
2. To describe the extent of reported dropout and the appropriateness of the analytical methods used to handle missing data,
3. To assess the use of sensitivity analyses.

2. Methods

2.1. Selection of studies

Five journals (*Annals of the Rheumatic Diseases*, *Arthritis & Rheumatism*, *Journal of Rheumatology*, *Pain*, and *Rheumatology*) were selected as sources of RCTs in the areas of arthritis and musculoskeletal conditions. Journals with high impact factors were targeted, as it was considered important to evaluate "best" statistical practice in this field, for which the impact factor was taken as a proxy. The impact factors ranged from 3.6 (*Journal of Rheumatology*) to 9.1 (*Annals of the Rheumatic Diseases*).

Only parallel-arm individual-level RCTs using a superiority design and reporting on the primary outcome of the study were included in this study. Additional exclusion criteria were as follows:

1. Pilot/feasibility studies, as these mainly aim to demonstrate the feasibility and/or affordability of subsequently conducting a large similar study, rather than to detect a true between-group difference with sufficient power.
2. Trials with fewer than 50 randomized subjects, as the small sample size could impose limitations on possible methods of analysis.
3. Publications based on an interim analysis (ie, where the primary analysis was centered on an outcome measured at a time point earlier than the designated primary end point).
4. Extended follow-up studies (those only reporting outcomes beyond the primary end point)
5. Studies with survival outcomes, as standard survival models take into account missingness through noninformative censoring.

2.2. Search strategy and data extraction

A search was performed for all reports of RCTs published between January 1, 2010, and December 31, 2011, focusing

on keywords "clinical trial", "randomization", "randomisation", "randomized", "randomised", "randomly", or "random" in the titles or abstracts to identify relevant citations. Fig. 1 illustrates the selection procedure and indicates the reasons for exclusions.

Using a proforma based on recent recommendations [12,23,38], data were extracted from each of the 91 eligible trial reports. Information was obtained on basic characteristics of the trials, participant exclusions and withdrawals, sample size calculation, measurement of outcomes, protocol deviations, methods used to handle missing data, statistical analyses performed, and sensitivity analyses conducted. To check consistency of data extraction, the process was repeated after a period of several months, and the data extraction was verified by a second investigator in 20% of randomly selected reports.

Data extraction centered on the primary outcome at the study primary end point. The primary outcome was identified from the definition given in the report (eg, in the study objectives) or from the details on the sample size calculation. If more than one primary outcome was reported, the first one listed was used. If, in the case of multiple follow-ups, the primary end point was not explicitly identified, it was taken to be the final measurement.

2.3. Dropout rate

Dropouts were subjects who did not complete the primary outcome assessment at the primary end point, whereas completers were those who completed the assessment. Dropouts include individuals lost to follow-up (through nonresponse) and those not followed up because of protocol violations such as ineligibility or treatment crossover. The dropout rate was calculated as the difference between the number randomized and the number remaining in the trial (completers) at the primary end point, divided by the number randomized. In trials with repeated follow-ups, the dropouts were classified as either "early dropouts," defined as subjects who did not complete any follow-up assessment on the primary outcome, or "late dropouts," defined as those who completed at least one follow-up assessment before dropping out.

2.4. Classification of analysis strategies

The analysis strategy used in the reviewed reports was categorized as FITT, PITT, complete case (CC), PP, or AT. The definition of each category is given in Table 1. FITT is an analysis of data as randomized and includes data on all randomized subjects through either imputation or modeling of any missing data. PITT denotes an analysis restricted to a subset of the FITT sample where the subsample excludes early dropouts (in trials with repeated follow-ups). The purpose of this classification is to highlight the exclusion of early dropouts from the primary analysis. In trials with a single follow-up, the exclusions lead to a CC analysis as there is no scope for further follow-up data. Such exclusions of dropouts with no follow-up data may

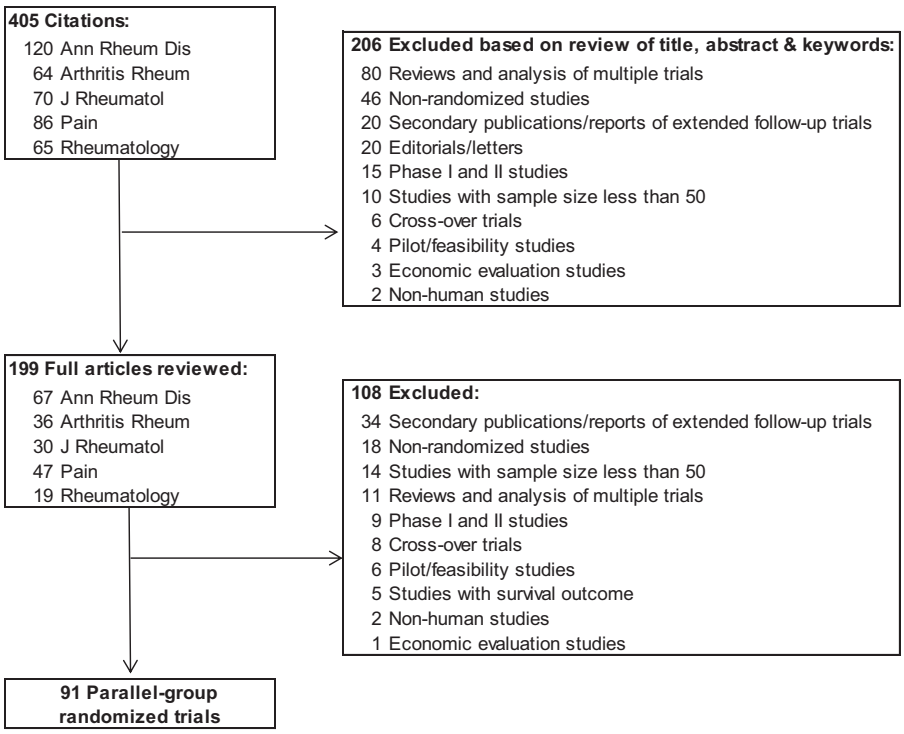

Fig. 1. Identification of randomized trials from January 2010 to December 2011.

Table 1. Classification of analysis strategy used in trial reports

Statistical analysis strategy	Explanation
Full ITT (FITT) analysis	All randomized subjects included in the analysis and analyzed as randomized.
Partial ITT (PITT) analysis	Analysis excludes only those randomized subjects who did not provide any follow-up data from an FITT sample in trials with repeated follow-ups.
Complete-case (CC) analysis	Analysis includes only those randomized subjects who completed the primary outcome measurement at the primary end point (ie, this analysis excludes subjects with missing data at the primary end point), and analysis is performed on the basis of randomized group allocation.
As-treated (AT) analysis	Subjects analyzed as treated, regardless of the treatment to which they were assigned.
Per-protocol (PP) analysis	Analysis includes subjects who completed the trial in full accordance with the study protocol.

Abbreviation: ITT, intention to treat.

be reasonable if we can assume that the process of exclusion is protected against biased selection of an outcome or a data-driven preference for a particular analysis. Such protection is afforded when the criteria for exclusion of participants from the analysis are prespecified in the protocol and are not based on information related to either treatment allocation or events or outcomes that occurred after randomization [3]. However, such exclusions should be limited to avoid selection bias [3]. Ideally, such decisions should also be made by a blind or independent observer. White et al. [39] argued that including all randomized subjects in an analysis of an outcome with missing data is insufficient; one should also consider an appropriate method to handle the missing data.

2.5. Classification of methods to handle missing data

The method used, if any, to deal with missing data was classified as

1. Listwise deletion
2. Single imputation, such as baseline observation carried forward (BOCF), last observation carried forward (LOCF), worst observation carried forward, nonresponder (ie, treatment failure) imputation, regression method, or linear extrapolation method
3. MI, whereby missing values are replaced by a set of values generated from the posterior predictive distribution of missing data
4. Statistical models that can include all randomized subjects without imputation of missing values (eg, MMRM)

The validity of the aforementioned methods is, as previously noted, dependent on the plausibility of the missing data assumption.

3. Results

3.1. Characteristics of included trials

A description of the 91 trials (list provided in Appendix at www.jclinepi.com) included in the review is presented in Table 2. The majority included a numerical primary outcome measure. Among the 72 trials that detailed a sample size calculation, only 28 (39%) made adjustment for attrition. In 18 of 72 trials (25%), the number of subjects randomized was less than the calculated sample size (the shortfall ranged from 1% to 53%; median, 5%). Most of the trials (79 of 91; 87%) in this review followed the CONSORT statement [40] in reporting armwise flow of participants through the trial.

3.2. Analysis strategy

Table 3 indicates the analysis strategy followed in the primary analysis of the trials. In all but one trial, subjects were analyzed as randomized (though 10 trials reported crossover of treatment after randomization). An FITT analysis was performed in 34 trials (37%); in four trials, there were no missing data at the primary end point, and in the remaining 30 trials, all randomized subjects were included in the analysis through either imputation of missing values

Table 2. Description of trials (n = 91) included in the study

Description of the trials	No. of trials (%)
Journal	
Annals of the Rheumatic Diseases	31 (34.1)
Arthritis & Rheumatism	16 (17.6)
Journal of Rheumatology	11 (12.1)
Pain	27 (29.6)
Rheumatology	6 (6.6)
Year of publication	
2010	38 (41.8)
2011	53 (58.2)
Multicenter trials	52 (57.1)
Number of subjects per trial[a]	
<100	36 (39.6)
100–499	41 (45.0)
500 and above	14 (15.4)
Number of arms per trial	
2	67 (73.6)
3	15 (16.5)
>3	9 (9.9)
Type of primary outcome measure	
Categorical	19 (20.9)
Numerical[b]	72 (79.1)
Number of follow-up assessments	
Single	11 (12.1)
Repeated	80 (87.9)

[a] Number of subjects randomized.
[b] Thirteen trials analyzed these outcomes as categorical.

Table 3. Analysis strategy followed in the primary analysis

Analysis strategy	Trials with single follow-up (n = 11)	Trials with repeated follow-ups (n = 80)	Total (n = 91)
Full ITT	2 (18.2)	32 (40.0)	34 (37.4)
Partial ITT	N/A	28 (35.0)	28 (30.7)
Complete-case analysis	7 (63.6)	17 (21.3)	24 (26.4)
As-treated analysis	0 (0.0)	1 (1.2)	1 (1.1)
Per-protocol analysis	2 (18.2)	2 (2.5)	4 (4.4)

Abbreviations: ITT, intention to treat; N/A, not applicable.
Data are counts (%).

or appropriate modeling of incomplete data. In detail, 21% of trials (17 of 80) with repeated follow-ups and 64% of trials (7 of 11) with a single follow-up excluded all dropouts from the primary analysis. Other protocol violations were reported in 22 trials, and four of them followed a PP strategy as the primary analysis. A PITT analysis (excluding early dropouts) was performed in 28 of 80 trials (35%) with repeated follow-ups.

3.3. Dropouts

Fig. 2 displays the percentage of trials with various levels of dropout. Eighty-six trials (95%) had some subjects with missing outcome data at the primary end point. The median dropout rate was 12% [interquartile range (IQR), 7–24%; range, 0–51%]. The median (IQR) dropout rate was 11% (5–21%) for trials using FITT and CC analysis methods and 20% (11–31%) for trials using PITT.

Among 11 trials with single follow-up, 10 (91%) reported dropouts [median (IQR) dropout rate of 9% (6–12%)]. Among 80 trials with repeated follow-ups, 39 (49%) reported early dropouts [median (IQR) dropout rate of 3% (1–9%)], and 75 trials (94%) reported the presence of late dropouts [median (IQR) dropout rate of 10% (6–21%)].

3.4. Dealing with dropouts

Among the 10 trials with single follow-up reporting dropouts, nine excluded these dropouts from analysis, whereas the remaining trial used BOCF. Among the 39 trials with repeated follow-ups reporting early dropouts, 36 (92%) excluded these dropouts from the analysis (28 followed PITT, 7 followed CC, and 1 followed AT analysis) and the remaining three used BOCF, MI, or MMRM to handle the missing data.

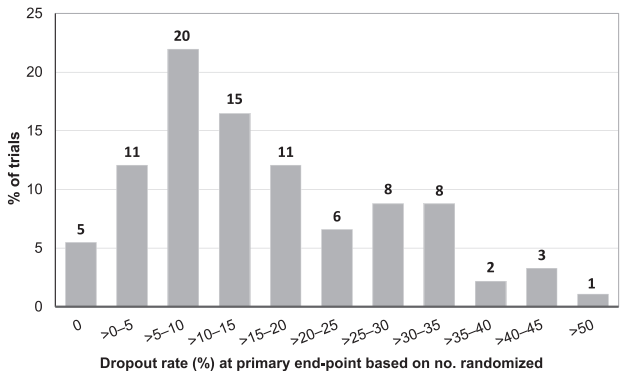

Fig. 2. The distribution of the trials (n = 90) based on the percentage of dropouts. The number above each bar indicates the number of trials. One trial did not report the dropout rate.

Among the 75 trials with late dropouts, 46 (61%) performed some sort of imputation (Table 4). Among the 26 trials (35%) that did not use any imputation, eight used analysis methods that make full use of all available data through modeling of incomplete repeated measures data (MMRM in six and GEE in two) and the remaining 18 excluded the dropouts from the primary analysis. LOCF was the most frequent imputation approach, used in 23 trials (31%).

3.5. Sensitivity analyses

Eighteen (21%) of 86 trials with missing outcome values at the primary end point reported a sensitivity analysis to assess the robustness of inferences from the primary analysis to a range of alternative plausible assumptions regarding missingness. The sensitivity analyses were performed in trials with relatively high proportions of missing data (median, 24%; IQR, 17–33%). Either exclusion of subjects with missing data (ie, listwise deletion) or a single

Table 4. Methods used to handle late dropouts who had completed at least one follow-up assessment

Method used in primary analysis	Percentage of late dropouts			
	>0–10%	>10–20%	>20%	Total
No imputation				
Excluded	10 (29.5)	6 (33.3)	2 (9.1)	18 (24.0)
MMRM	2 (5.9)	1 (5.6)	3 (13.6)	6 (8.0)
GEE	1 (2.9)		1 (4.6)	2 (2.7)
Single imputation				
LOCF[a]	7 (20.6)	7 (38.8)	8 (36.3)	23 (30.7)
NRI	7 (20.6)	2 (11.1)	3 (13.6)	12 (16.0)
BOCF	1 (2.9)		3 (13.6)	4 (5.3)
Regression imputation	1 (2.9)	1 (5.6)		2 (2.7)
LOCF + NRI[b]		1 (5.6)		1 (1.3)
LOCF + WOCF[c]			1 (4.6)	1 (1.3)
Linear extrapolation	1 (2.9)			1 (1.3)
Multiple imputation	2 (5.9)			2 (2.7)
No details	2 (5.9)		1 (4.6)	3 (4.0)
Total	34 (100)	18 (100)	22 (100)	75 (100)

Abbreviations: MMRM, mixed model for repeated measures; GEE, generalized estimating equations; LOCF, last observation carried forward; NRI, nonresponder imputation; BOCF, baseline observation carried forward; WOCF, worst observation carried forward.

Data are counts (%).

[a] Dropout rate not reported in one trial.

[b] Trials in which subjects dropping out were treated as nonresponders when dropout is due to adverse events or lack of effectiveness, otherwise imputed with LOCF.

[c] Trials imputed with WOCF when dropout is due to adverse events or inefficacy, otherwise imputed with LOCF.

imputation method was the designated sensitivity analysis. Very few trials (6 of 18; 33%) presented the results of their sensitivity analysis, whereas the others just reported that a sensitivity analysis had been performed and indicated that the findings from the primary analysis were supported by those of the sensitivity analysis.

3.6. Cautionary notes on missing data

We further reviewed all 55 trial reports with 10% or more missing data to determine how the authors addressed the uncertainty due to missing data in their interpretation of results. Among those trials that did not report a sensitivity analysis, none attempted to highlight the uncertainty around their findings due to the missing data, apart from a few instances in which dropout was briefly identified as a limitation of the study.

4. Discussion

In accordance with the ITT principle, most trials in the review analyzed data by the groups to which subjects were randomized, regardless of the intervention received. However, many of these trials failed to obtain outcome data on all randomized subjects and/or include all subjects in the primary analysis. The average dropout rate was a little over 10%, and because most trials failed to use appropriate statistical methods to account for missingness, it is likely that descriptive data and inferential estimates of treatment effect were biased, given that missing data likely differ from reported data. In total, only one-third of trials used an FITT approach for the primary analysis. This proportion is comparable with those reported by Kruse et al. [25] and Gravel et al. [27], but higher than that reported by Baron et al. [29]. The lower proportion noted by Baron et al. [29] may reflect changing practice since 2003 or may be due to their specific focus on structural outcomes in trials of rheumatic diseases without regard to whether these were a primary outcome.

Early dropouts are a major challenge to performing an FITT analysis. Nearly half of the trials with repeated follow-ups in this review reported early dropout after randomization; nearly a quarter of them had more than 10%. In such cases, the most commonly favored approach was to exclude those dropouts from the analysis; thus, 28 of 39 trials (72%) with early dropouts performed a PITT analysis and another seven (18%) performed a CC analysis. Late dropouts are also a challenge to an FITT analysis. Most trials with repeated follow-ups (75 of 80; 94%) reported late dropouts; more than half had more than 10% late dropouts, and a quarter had more than 20%. A quarter of trials (18 of 75) with late dropouts excluded them from the primary analysis. In a pragmatic trial, exclusion of participants may limit interpretation of findings and an FITT analysis is therefore recommended [1,3], but many trials in this review chose instead to exclude participants who had dropped out. Conceptually, if the analyzed-as-randomized principle is disturbed in any way, and for whatever reason, the chance of an imbalance in baseline variables increases.

Trials in this review frequently reported the reasons for the dropouts but failed to justify the assumptions made regarding missingness. Definitive testing of the assumptions regarding missingness is not possible—certainly in regard to the assumption of the missing data mechanism being nonignorable. However, some degree of testing is achievable; for example, in a stratified (by treatment group) comparison of observed data, differences between responders and nonresponders may help to reject the possibility of an MCAR assumption [21]. However, the assumption of ignorable missing data (as assumed by most statistical approaches) is difficult to test formally, as it is impossible to know with certainty whether the reason for missingness is somehow related to the fact that the data are missing or are related to any unobserved covariates. Clearly, it is important to gather as much information as possible on the reasons for missing data and on patient characteristics that can predict missingness [23], as such information can help justify (or not) an ignorable missing data assumption. Trialists should consider the availability of secondary sources (eg, general practice records) to obtain outcome data when there are missing data from the primary source. Furthermore, none of the trials considered the possibility of MNAR, and many failed to adopt methods that are appropriate and valid under an MAR assumption, which is a recommended neutral starting point in many settings [14,22,23]. Importantly, the most commonly applied method (LOCF) may not even be valid under MCAR [17,23].

4.1. Power calculation in anticipation of dropouts

In the review, 21% of trials (19 of 91) failed to report a formal sample size calculation, contrary to CONSORT recommendations [41]. Sample size should relate to a predetermined primary outcome. Specification of a primary outcome variable and primary end point guards against changing the planned outcome and placing undue emphasis on one that was not the original primary outcome. Of the 72 trials reporting a sample size calculation, 18 (25%) failed to achieve adequate numbers at randomization and 62 (86%) did not meet the target set for the primary end point. Particularly important in relation to missing data is how to account for loss of power due to dropouts in hypothesis tests or confidence intervals [42]. One approach is to inflate the sample size to take account of missing values; 28 of 72 trials (39%) reporting a sample size calculation adjusted the sample size in this way. Such inflation of sample size assumes that the loss of nominal power is proportional to the amount of missing data. However, this assumption may not be warranted. Little et al. [42] point out that inflation of the sample size does not necessarily reduce bias and

is valid only under an MCAR assumption, but this assumption is very rarely justified. Therefore, power analyses that account for missing data should be based on more plausible MAR or MNAR assumptions [42].

4.2. Baseline comparison

Exclusion of randomized subjects can affect the baseline balance achieved through randomization. However, although most trials reported the dropout rate between arms and provided reasons for dropouts, many failed to compare key baseline variables between subjects with and without missing outcome data. Moreover, most trials evaluated the differences in baseline characteristics between arms based on number of subjects randomized rather than on number analyzed, despite discrepancies between these numbers. This oversight fails to locate (and hence leads to failure to adjust for) any imbalance in baseline characteristics between arms in the analysis data set. Many reports highlighted the equality in the dropout rate between arms, suggesting a view that an equal dropout rate would not lead to a biased estimate of treatment effect. However, bias is a function of both the frequency of and the reasons for the missing values in each arm [20].

4.3. Handling missing data

Although most trials with dropouts included them in the primary analysis, most of these trials used inappropriate methods to account for missing outcome data in the analysis, and inferences from these analyses may therefore be biased. A CC analysis only includes subjects with complete data. In trials with repeated follow-ups, standard statistical methods, such as analysis of covariance, exclude subjects for whom some intermediate measurements are available. A quarter (18 of 75) of trials with repeated follow-ups that reported late dropouts excluded them from the primary analysis.

Several single imputation strategies are common in RCTs [19]. These methods are generally not recommended because they inadequately account for uncertainty in the data and may produce biased estimates [22]. Many trials (44 of 75; 60%) with repeated follow-ups used some sort of single imputation. The findings of these trials are doubtful, as many of these methods are not valid even under an MCAR assumption [17,23]. In particular, the LOCF approach makes a very strong assumption, which is unlikely to be true, that the value of an outcome remains constant after dropout. This was the most frequently used imputation method in the review (in more than a third of the trials with late dropouts) despite recommendations against its use [22,23,43]. LOCF has been shown to be common elsewhere [44,45].

A substantial proportion of trials used imputations that require extreme assumptions; for example, the assumption that dropouts are ''nonresponders'' or have no change from baseline. In 16% (12 of 75) of trials with late dropouts, the dropouts were simply classified as ''failure'' where the primary outcome was analyzed on a dichotomous scale (''success'' or ''failure''). Additionally, in another four trials with late dropouts, the missing data were replaced by baseline data (a BOCF approach) where the primary outcome was analyzed on a continuous scale. These imputations sometimes provide informative bounds on the effect of the missing data but rarely produce unbiased estimates of treatment effect.

Recently, MI has received substantial attention in the literature as it helps to overcome the limitations of single imputations [46]. The MI technique uses several stochastic imputations to incorporate the uncertainty surrounding the missing value and gives valid standard errors under MAR. Despite evidence of increasing use of MI in the analysis of RCTs [36], only two trials in this review [47,48] performed MI-based analyses and reported the results. Another three trials claimed to have performed MI as a sensitivity analysis but failed to report the results. Importantly, trials that performed MI failed to report the procedure adequately. Sterne et al. [46] suggested guidelines for reporting analyses based on MI to avoid pitfalls with its application and aid interpretation of its results.

Similarly, analysis methods that make use of all available data in the presence of dropouts were infrequently used. Likelihood-based mixed-effects models such as MMRM can use all available longitudinal data without a need to impute values and are valid when the dropout mechanism is ignorable [33]. Only 8% (6 of 75) of trials with late dropouts used MMRM to analyze longitudinal outcome data. Additionally, MMRM was performed after imputation of missing values using LOCF in two trials [49,50], BOCF in one trial [51], and MI in another one [48]. The use of imputations such as LOCF and BOCF undermines the benefits of MMRM, and the results may not be valid under MAR. Also, the use of MI before performing the MMRM is unnecessary, as there is no obvious gain from doing so [52].

Semiparametric regression-based methods such as GEE can also use all available data in the presence of dropouts and were used by two trials. However, standard GEE methods are valid only under MCAR [53]. Weighted GEE (where weight is assigned at the subject level and is calculated as the inverse of the probability for dropping out at the observed time of dropout) or MI-based GEE (where MI is used before performing GEE) are preferred over standard GEE because these methods can provide a valid estimate of treatment effect under MAR [54]. No trials in this study presented results based on these methods.

4.4. Sensitivity analysis

Many researchers agree that analysis based on an MAR assumption is often a reasonable starting point [14,22,23].

However, one should always be open to the possibility that the data are MNAR. One should, therefore, evaluate the sensitivity of results to possible departures from the MAR assumption by assuming a range of plausible MNAR mechanisms. Our study found sensitivity analyses to be infrequently and inappropriately used and insufficiently reported. Moreover, trials that performed a sensitivity analysis used CC or single imputation methods, which may not be justified given that MCAR is an unlikely scenario. Sensitivity analyses that test the robustness of the primary and other key outcome estimate(s) to plausible deviations from underlying assumptions are strongly recommended. Along with judicious sensitivity analyses, discussion around interpretation and validity of the primary analysis findings is highly recommended and should be an integral part of any statistical report, unless perhaps the amount of missing data is minimal [23].

5. Conclusion

This review focuses specifically on recent trial publications for musculoskeletal conditions, although similar dropout figures have been reported for RCTs covering a broad spectrum of clinical areas [15]. Conceptually, extrapolation of the findings to other areas using more explanatory designs is questionable, given the different methodological issues raised in connection with pragmatic trials [7].

The findings of this study are comparable with those of previous reviews [24−27] of trials published in general medical journals. It appears that progress has not been made in reducing the large proportion of trials that are inappropriately analyzed and that may therefore be prone to erroneous estimates and conclusions. Given that missing outcome data are not avoidable in most musculoskeletal trials, researchers should ensure that missing values are handled judiciously, in line with current best practice, and should use methods of analysis that make appropriate assumptions about missingness. Equally, such trials should be reported in line with current guidelines. Because trials with a large proportion of missing data are highly sensitive to deviation from simple assumptions like MCAR and the assumptions cannot be fully justified from the data, reporting of sensitivity analyses is advised.

Author contributions

R.J. carried out the review, with independent assessment also performed by R.O. on 20% of collated manuscripts. Any disagreement was resolved by consensus. R.J. drafted the manuscript and carried out statistical evaluation of the findings. All authors were involved in the conception of the study, interpretation of the findings, revisions to the original draft, and have approved submission of the final article.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2014.09.002.

References

[1] Chan A, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013;346:e7586.
[2] Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ 2001;165(10):1339−41.
[3] Heritier SR, Gebski VJ, Keech AC. Inclusion of patients in clinical trial analysis: the intention-to-treat principle. Med J Aust 2003;179:438−40.
[4] Altman DG. Missing outcomes in randomized trials: addressing the dilemma. Open Med 2009;3(2):e51−3.
[5] Fleming TR. Addressing missing data in clinical trials. Ann Intern Med 2011;154:113−7.
[6] Lewis JA, Machin D. Intention to treat—who should use ITT? Br J Cancer 1993;68:647−50.
[7] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390.
[8] Lachin JM. Statistical consideration in the intention-to-treat principle. Control Clin Trials 2000;21:167−89.
[9] Higgins JPT, Altman DG. Chapter 8: assessing risk of bias in included studies. In: Higgins JPT, Green S, Eds. Cochrane handbook for systematic reviews of interventions, Version 5.1.0: The Cochrane Collaboration. Available at www.cochrane-handbook.org. Accessed June 8, 2014.
[10] Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781−8.
[11] Abraha L, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ 2010;340:c2697.
[12] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
[13] McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. Stat Med 2009;28:2639−52.
[14] White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ 2011;342:d40.
[15] Crutzen R, Viechtbauer W, Kotz D, Spigt M. No differential attrition was found in randomized controlled trials published in general medical journals: a meta-analysis. J Clin Epidemiol 2013;66:948−54.
[16] Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing data. Am Stat 2001;55:244−54.
[17] European Medicines Agency. Guideline on missing data in confirmatory clinical trials. London, UK: Committee for Medical Products for Human Use; 2010: Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf. Accessed September 20, 2013.
[18] Rubin DB. Inference and missing data. Biometrika 1976;63(3):581−92.
[19] Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken, NJ: Wiley; 2002.
[20] Wright CC, Sim J. Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. J Clin Epidemiol 2003;56:833−42.
[21] Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. Health Qual Life Outcomes 2009;7:57.

[22] Molenberghs G, Kenward MG. Missing data in clinical studies. Chichester, UK: Wiley; 2007.
[23] National Research Council. The prevention and treatment of missing data in clinical trials. Panel on handling missing data in clinical trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press; 2010.
[24] Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999;319:670.
[25] Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: Who is in? Who is out? J Fam Pract 2002;51:969−71.
[26] Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials 2004;1:368−76.
[27] Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? Clin Trials 2007;4:350−6.
[28] Fielding S, Maclennan G, Cook JA, Ramsay CR. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. Trial 2008;9:51.
[29] Baron G, Boutron I, Giraudeau B, Ravaud P. Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. Arthritis Rheum 2005;52:1858−65.
[30] Henschke N, Kuijpers T, Rubinstein SM, van Middelkoop M, Ostelo R, Verhagen A, et al. Trends over time in the size and quality of randomised controlled trials of interventions for chronic low-back pain. Eur Spine J 2012;21(3):375−81.
[31] Koes BW, Malmivaara A, van Tulder MW. Trend in methodological quality of randomised clinical trials in low back pain. Best Pract Res Clin Rheumatol 2005;19(4):529−39.
[32] Food and Drug Administration. Guidance for sponsors, clinical investigators, and IRBs: data retention when subjects withdraw from FDA-regulated clinical trials 2008: Available at http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126489.pdf. Accessed October 17, 2013.
[33] Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Stat Med 1997;16:2349−80.
[34] Hardin JW, Hilbe JM. Generalized estimating equations. Boca Raton, FL: Chapman & Hall/CRC; 2003.
[35] Preisser JS, Lohman KK, Rathouz PJ. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. Stat Med 2002;21:3035−54.
[36] Mackinnon A. The use and reporting of multiple imputation in medical research—a review. J Intern Med 2010;268(6):586−93.
[37] Akai M. Musculoskeletal disorders. Pharmaceutical sciences encyclopedia 2010:1−23.
[38] Hopewell S, Hirst A, Collins G, Mallett S, Yu L, Altman D. Reporting of participant flow diagrams in published reports of randomized trials. Trials 2011;12:253.
[39] White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. Clin Trials 2012;9:396−407.
[40] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191−4.
[41] Moher D, Hopewell S, Kenneth FS, Montori V, Peter CG, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869.
[42] Little RJ, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Neaton JD, et al. The design and conduct of clinical trials to limit missing data. Stat Med 2012;31:3433−43.
[43] Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm Stat 2008;7(2):93−106.
[44] Kim M. Statistical methods in arthritis and rheumatism: current trends. Arthritis Rheum 2006;54:3741−9.
[45] Woolley SB, Cardoni AA, Goethe JW. Last-observation-carried-forward imputation method in clinical efficacy trials: review of 352 antidepressant studies. Pharmacotherapy 2009;29(12):1408−16.
[46] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.
[47] Beaudreuil J, Lasbleiz S, Richette P, Seguin G, Rastel C, Aout M, et al. Assessment of dynamic humeral centering in shoulder pain with impingement syndrome: a randomised clinical trial. Ann Rheum Dis 2011;70(9):1613−8.
[48] Thorn BE, Day MA, Burns J, Kuhajda MC, Gaskins SW, Sweeney K, et al. Randomized trial of group cognitive behavioral therapy compared with a pain education control for low-literacy rural people with chronic pain. Pain 2011;152(12):2710−20.
[49] Genevay S, Viatte S, Finckh A, Zufferey P, Balagué F, Gabay C. Adalimumab in severe and acute sciatica: a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2010;62:2339−46.
[50] Fary RE, Carroll GJ, Briffa TG, Briffa NK. The effectiveness of pulsed electrical stimulation in the management of osteoarthritis of the knee: results of a double-blind, randomized, placebo-controlled, repeated-measures trial. Arthritis Rheum 2011;63:1333−42.
[51] Bliddal H, Leeds AR, Stigsgaard L, Astrup A, Christensen R. Weight loss as treatment for knee osteoarthritis symptoms in obese patients: 1-year results from a randomised controlled trial. Ann Rheum Dis 2011;70(10):1798−803.
[52] Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J Clin Epidemiol 2013;66:1022−8.
[53] Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13−22.
[54] Birhanu T, Molenberghs G, Sotto C, Kenward MG. Doubly robust and multiple-imputation-based generalized estimating equations. J Biopharm Stat 2011;21(2):202−25.