Journal of Clinical Epidemiology 143 (2022) 105–117

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Analyses of repeatedly measured continuous outcomes in randomized controlled trials needed substantial improvements

Yan Ren[a,b,c], Yuanjin Zhang[a,b,c], Yulong Jia[a,b,c], Yunxiang Huang[a,b,c], Minghong Yao[a,b,c], Ling Li[a,b,c], Guowei Li[d], Qianrui Li[a,e], Min Yang[f], Peijing Yan[f], Yuning Wang[a,b,c], Kang Zou[a,b,c], Xin Sun[a,b,c,*]

[a] *Chinese Evidence-based Medicine Center, Cochrane China Center, National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China*

[b] *NMPA Key Laboratory for Real World Data Research and Evaluation in Hainan, Chengdu, Sichuan 610041, China*

[c] *Sichuan Center of Technology Innovation for Real World Data, Chengdu, Sichuan 610041, China*

[d] *Center for Clinical Epidemiology and Methodology (CCEM), Guangdong Second Provincial General Hospital, Guangzhou, Guangdong 510000, China*

[e] *Department of Nuclear Medicine, West China Hospital of Sichuan University, Chengdu, Sichuan 610041, China*

[f] *Department of Epidemiology and Biostatistics, West China School of Public Health, Sichuan University, Chengdu, Sichuan 610041, China*

Accepted 2 December 2021; Available online 8 December 2021

Abstract

Objectives: Systematic understanding is lacking regarding how current trials handle repeated measure data and the extent to which appropriate statistical methods are used for such data set. This study investigated the current practice of analyzing the repeated measure data among randomized controlled trials (RCTs).

Study Design and Setting: We searched the Core Clinical Journals indexed in PubMed for RCTs published in 2019 and included a continuous primary outcome with repeated measures. We randomly sampled RCTs from the eligible trials. Team of methods trained investigators screened studies for eligibility and collected data using the pilot-tested, standardized questionnaires. We thoroughly documented statistical analyses of the continuous primary outcome with repeated measures and particularly recorded how statistically advanced methods were used to handle these repeated measures.

Results: In total, 200 trials were included. Of these trials, the mean number of repeated measures for the continuous primary outcome was 5.46 (SD = 3.4); 58 (29.0%) trials did not specify the time point of primary outcome in the method; 113 (56.5%) trials did not use statistically advanced methods for handling repeated measure data in the primary analyses. Among187 trials included the baseline values, 88 (47.1%) trials did not adjust for outcome value at baseline. Among 87 trials using statistically advanced methods, 49 (56.3%) did not specify correlation structure for model.

Conclusions: The statistical analyses of repeatedly measured continuous outcomes in RCTs need substantial improvements. Careful planning of the primary outcome and the use of statistically advanced methods for analyzing data are warranted.

Keywords: Randomized controlled trials; Continuous outcomes; Repeated measure data; Analysis; Statistical methods; Cross-sectional survey

1. Introduction

Repeated measures are common in the randomized controlled trials (RCTs) [1,2] and are often used to investigate the treatment effect [3,4]. However, repeated measure data from the same patient are often correlated and many commonly used statistical methods, such as *t*-test, analysis of variance (ANOVA), and linear regression models are not appropriate for handling repeated measure data [5]. Although repeated measure ANOVA may address the issue of correlation, it carries the strong assumption of sphericity,

Abbreviations: RCT, randomized controlled trial; CI, confidence interval; PRO, patient-reported outcome; GEE, generalized estimating equations; ANOVA, analysis of variance; CONSORT, Consolidated Standards of Reporting Trials; ANCOVA, analysis of covariance; BMJ, British Medical Journal; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; SPIRIT, Standard Protocol Items; MMRM, mixed-effect model for repeated measures.

Conflict of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[*] Corresponding author.

E-mail address: sunxin@wchscu.cn (X. Sun).

https://doi.org/10.1016/j.jclinepi.2021.12.007

What is new?

Key findings:

- More than half of RCTs did not use statistically advanced methods, such as the mixed-effect model for repeated measures analysis or generalized estimating equations (GEE), for handling repeatedly measured continuous outcomes. Nearly half of the trials did not adjust for outcome value at baseline.
- More than half of trials using statistically advanced methods failed to specify the details about the statistical model, including the assumption of correlation structure, the inclusion of variables, and adjustment for baseline outcome values.
- Trials published in the higher impact journals and the involvement of methodologists were more likely to use statistically advanced methods.

What this adds to what is known

- Although statistically advanced methods have been recommended for analyzing repeated measures data, it remains unclear how current trials handle such data in their actual research practice. Through a comprehensive survey, our study identified several critical methodological issues in journals of various scientific impacts. The results suggested that statistical analyses of repeatedly measured continuous outcomes in current RCTs still require substantial improvement, moreover the use of statistically advanced methods is highly warranted. .

What is the implication, what should change now

- When planning and analyzing repeatedly measured continuous outcomes in RCTs, use of statistically advanced methods is desirable.
- Pre-specification of the methods for analyzing repeated measurement data and statistical details (e.g. assumption of the correlation structure, variables inclusion in the statistical model, and adjustment for the baseline outcome value) are warranted to improve the methodological quality of statistical analyses as well as to promote the transparency.

which is often unrealistic in medical research. Situations may become even more complicated in case of missing outcome data [6,7].

For solving these issues, statistically advanced models such as the mixed-effect model for repeated measures (MMRM) analysis and generalized estimating equations (GEE) are developed to handle complex repeated measure data [2,3,8–10]. Because these models can utilize all available data from repeated measures and fully consider the interaction between treatment group and the time, they often have superior statistical performance over traditional statistical methods, not only helpful for investigating the main effect across multiple time points or the change of treatment effects over time, but also for investigating treatment effect at a specific time point [4]. Nevertheless, the adaption of these methods is relatively low and trial investigators may remain interested in using traditional methods. The omission of repeated measures may reduce the power of analysis and sometimes leads to the biased estimates.

Despite statistical recommendations for repeated measures are available [11–13], it remains unclear that how current trials utilized repeated measures and whether they used statistically advanced methods in the analyses. The lack of this important information may prevent the trial community from further improving study design and analysis of repeated measures. We thus conducted a cross-sectional survey of current RCTs involving repeatedly measured continuous outcomes to examine the current practice and identify critical methodological issues for improvements.

2. Methods

2.1. Definitions

We classified statistical analysis approaches into statistically advanced methods and conventional methods, according to (1) whether all available repeated measure data were fully used and (2) whether missing data were dealt flexibly. The handling of missing data was considered ’flexible’ if a model readily accommodated the missing data (such as a linear mixed-effect model), in which case additional imputation was not necessary [4,9,14]. Statistically advanced methods included mixed-effect model for repeated measures analysis and generalized estimation equation. Conventional methods were defined as the use of *t*-test, ANOVA, analysis of covariance (ANCOVA), linear regression models and repeated measure ANOVA. In addition, we defined patient-reported outcome (PRO) as an outcome reported by the patients, such as assessments of health status, quality of life, and symptoms [15].

2.2. Eligibility criteria

We included a study if it was an RCT with at least one repeatedly measured continuous outcome as the primary outcome. The primary outcome had to be clearly described in the abstract, method, or results of the RCT. We included studies using repeated measured data with at least 3 measurements to fit for statistically advanced methods (e.g., mixed-effect model for repeated measures analysis or generalized estimation equation).

Studies were excluded if they were explicitly labelled as a phase I trial or phase I/II trial, an unparalleled RCT (including a factorial RCT, a crossover RCT, a stepped wedge cluster RCT, or an n-of-1 trial), reported a continuous variable as the primary outcomes but analyzed as a categorical variable, a subgroup analysis of RCTs or a study protocol.

2.3. Literature search

We searched PubMed to identify RCTs published in Core Clinical Journals between January 1 and December 31, 2019. The Core Clinical Journals were previously known as the Abridged Index Medicus and included 118 journals covering all specialties of clinical medicine and public health sciences [16]. In developing the search strategy, we used search terms related to the randomized controlled trial. The full search strategy is outlined in Appendix A.

2.4. Study process

Teams of paired reviewers (Y.R., Y.J., Y.Z., and Y.H.), trained in clinical epidemiology or medical statistics with practical skills of repeated measurement analysis, undertook the study selection and the data extraction. First, they independently screened titles and abstracts to identify reports of RCTs. Then they independently screened full texts using the pre-defined eligibility criteria and extracted data from eligible studies using pilot-tested, structured forms with detailed written instructions. Any disagreements were resolved by the discussion, if needed, adjudication by a third reviewer (X.S.).

2.5. Data collection

We developed a questionnaire to collect data from eligible RCTs. First, the initial version was developed by two investigators (Y.R. and M.Y.) trained in medical statistics with reference to Consolidated Standards of Reporting Trials (CONSORT) [17] and the other methodological studies [18–21]. Subsequently, they discussed with two experienced methodologists (L.L. and X.S.) to decide if each of the questions were appropriate, or dropped the items otherwise. Then, two investigators (Y.R. and Y.J.) conducted pilot testing by collecting data from 10 trials, and documented issues about the appropriateness and accuracy of the questionnaire. Finally, on the basis of the pilot testing, the team of ten investigators convened to discuss and determine if each of the questions needs to be included or if the questions should be further refined.

For each RCT, we pre-defined a continuous primary outcome with repeated measures, using the following rules: (1) if the trial clearly specified a continuous primary outcome with repeated measures in the abstract, method or result, we selected that outcome as the primary one; and (2) If the trial included more than one continuous primary outcomes with repeated measures, we selected the one that was firstly reported in the abstract.

We collected general study characteristics from each RCT, including the first-author name, journal, number of centers (single center, multicenter: number of trial sites $\geq$ 2, not reported), international trial (i.e., number of countries involved $\geq$ 2), type of intervention (drugs, medical devices, surgical, behavioral intervention, rehabilitation, invasive nonsurgical procedure, other), type of control (standard care, placebo, drugs, medical devices, surgical, invasive nonsurgical procedure, rehabilitation, psychological, behavioral intervention, other), source of funding (government, private for profit, private not for profit, no funding, not reported), sample size, length of follow-up, trial registration, protocol availability, provision of statistical analysis plan, number of treatment arms, use of blinding, and involvement of a methodologist. We assessed the protocol availability as to whether the study protocol was publicly available. We judged that a methodologist was involved if any authors declared an affiliation with a department of epidemiology or statistics, or if a methodologist was clearly acknowledged in the paper.

For the pre-defined primary outcome, we collected the information including number of repeated measures, whether baseline measure was included, type of the primary outcome (i.e., laboratory examinations, symptoms, quality of life, functional status, other), whether it was a patient-reported outcome, form of the primary outcome (i.e., raw value, absolute change from baseline, percent change from baseline, other), was the time point of primary outcome pre-specified, which time points were used for the primary outcome, if single time point was used for the primary outcome, whether multiple time points of the outcome variable were included for secondary analyses.

We also collected the information details regarding statistical analyses of the pre-defined primary outcome, as below: (1) which statistical methods were used (i.e., *t*-test or Mann-Whitney U test or ANOVA, ANCOVA, repeated measure ANOVA, mixed-effect model for repeated measures analysis, generalized estimation equation, other, not specified); (2) if an advanced method was used, whether correlation structure was assumed and whether the fit of correlation structure was assessed; (3) if an advanced method was used, then whether the trial investigators described independent variables (e.g., intervention, the timing of measurement, and their interactions); (4) whether any sensitivity analysis was conducted using advanced methods; (5) whether the baseline outcome data was adjusted, and which method was used for adjusting; (6) whether the primary analysis was the same as the statistical analysis plan.

2.6. Sample size and random sampling

We estimated 200 papers to achieve the desired confidence interval (0.43, 0.57) around the proportion of RCTs that used advanced methods regarding repeatedly measured continuous outcome, where the proportion was set as 0.5 since it was the most conservative situation. A total of 200 papers could achieve a sufficient power when assessing current practice, in line with similar studies [18,21].

We stratified journals into higher impact factor groups (Journal of the American Medical Association, The Lancet,

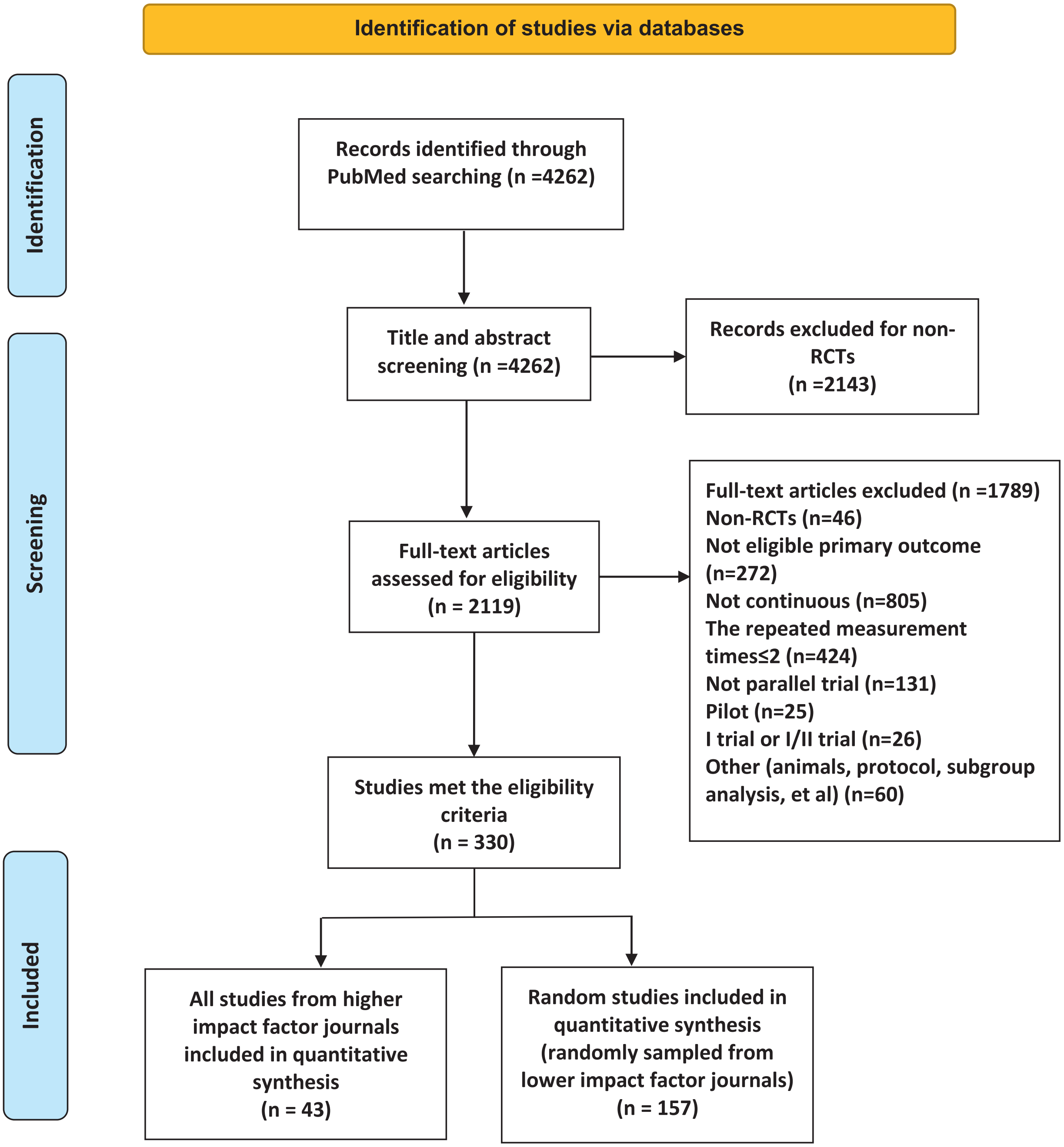
Identification of studies via databases
Identification
Screening
Included
Records identified through PubMed searching (n =4262)
Title and abstract screening (n =4262)
Records excluded for non-RCTs (n =2143)
Full-text articles assessed for eligibility (n = 2119)
Full-text articles excluded (n =1789)
Non-RCTs (n=46)
Not eligible primary outcome (n=272)
Not continuous (n=805)
The repeated measurement times≤2 (n=424)
Not parallel trial (n=131)
Pilot (n=25)
I trial or I/II trial (n=26)
Other (animals, protocol, subgroup analysis, et al) (n=60)
Studies met the eligibility criteria (n = 330)
All studies from higher impact factor journals included in quantitative synthesis (n = 43)
Random studies included in quantitative synthesis (randomly sampled from lower impact factor journals) (n = 157)

Fig. 1. Flow chart of study selection. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

New England Journal of Medicine, British Medical Journal (BMJ)) and lower impact factor groups (the remaining Core Clinical journals) according to 2019 impact factor from the Institute for Scientific Information (ISI) Web of Knowledge Journal Citation Reports. We included all identified publications from the higher impact journals and a random sample from the eligible studies published in lower impact journals. We assigned a unique ID number for each searched record. For eligible studies in lower impact journals, we randomly sampled from this unique ID number using SPSS 24.0 software.

2.7. Analysis

For all descriptive analyses, we used percentages for categorical variables, and mean (standard deviation) or median (interquartile range) for continuous variables. We compared general study characteristics, characteristics about primary outcome, and statistical analysis methods between RCTs published in the higher impact journals versus the lower impact journals, using the Chi-Square test or Fisher's exact test for categorical variables, and *t*-test or Mann Whitney U test for continuous variables.

Table 1. General characteristics of included randomized controlled trials.

Characteristics	All (N, %)	Higher impact journals (N, %)	Lower impact journals (N, %)	*P* value
Sample sizes[a]	148 (66, 249)	305 (168, 663)	100 (60, 194)	<0.001
≥ 100	118 (59.0)	37 (86.0)	81 (51.6)	
< 100	82 (41.0)	6 (14.0)	76 (48.4)	
Length of follow-up[a]	180 (56, 365)	364 (180, 560)	174 (42, 365)	0.002
Number of centers				<0.001
Single center	86 (43.0)	2 (4.7)	84 (53.5)	
Multicenter	101 (50.5)	40 (93.0)	61 (38.9)	
Not reported	13 (6.5)	1 (2.3)	12 (7.6)	
International study	24 (12.0)	16 (37.2)	8 (5.1)	<0.001
Type of intervention				0.095
Drugs	98 (49.0)	28 (65.1)	70 (44.6)	
Medical devices	23 (11.5)	3 (7.0)	20 (12.7)	
Surgical	17 (8.5)	3 (7.0)	14 (8.9)	
Behavioral intervention	23 (11.5)	7 (16.3)	16 (10.2)	
Rehabilitation	18 (9.0)	1 (2.3)	17 (10.8)	
Invasive nonsurgical procedure	6 (3.0)	0 (0)	6 (3.8)	
Other	15 (7.5)	1 (2.3)	14 (8.9)	
Type of control				0.703
Standard care	36 (18.0)	9 (20.9)	27 (17.2)	
Placebo	75 (37.5)	21 (48.8)	54 (34.4)	
Drugs	39 (19.5)	7 (16.3)	32 (20.4)	
Medical devices	12 (6.0)	2 (4.7)	10 (6.4)	
Surgical	11 (5.5)	1 (1.3)	10 (6.4)	
Behavioral intervention	10 (5.0)	2 (4.7)	8 (5.1)	
Rehabilitation	8 (4.0)	0 (0)	8 (5.1)	
Invasive nonsurgical procedure	1 (0.5)	0 (0)	1 (0.6)	
Other	8 (4.0)	1 (2.3)	7 (4.5)	
Number of treatment arms				0.405
2	158 (79.0)	32 (74.4)	126 (80.3)	
3 or more	42 (21.0)	11 (25.6)	31 (19.7)	
Source of funding[b]				
Government funding	96 (48.0)	19 (44.2)	77 (49.0)	0.572
Private for profit	45 (22.5)	19 (44.2)	26 (16.6)	<0.001
Private not for profit	64 (32.0)	14 (32.6)	50 (31.9)	0.929
No funding	9 (4.5)	0 (0)	9 (5.7)	0.209
Not reported	18 (9.0)	0 (0)	18 (11.5)	0.015
Was the trial registered?	186 (93.0)	43 (100.0)	143 (91.1)	0.196
Was trial protocol publicly available?	100 (50.0)	43 (100.0)	57 (36.3)	<0.001
Source of trial protocol				<0.001
Published article	26 (26.0)	6(14.0)	20 (35.1)	
Trial registry platform	24 (24.0)	4 (9.3)	20 (35.1)	
Supplemental content	50 (50.0)	33 (76.7)	17 (29.8)	
Was statistical analysis plan available?	97 (48.5)	42 (97.7)	55 (35.0)	<0.001
Source of statistical analysis plan				0.001
Protocol	72 (74.2)	32 (76.2)	40 (72.7)	

(*continued on next page*)

Table 1 (*continued*)

Characteristics	All (N, %)	Higher impact journals (N, %)	Lower impact journals (N, %)	*P* value
Trial registry platform	16 (16.5)	2 (4.8)	14 (25.5)	
Supplemental content	9 (9.3)	8 (19.0)	1 (1.8)	
Methodologist involved	61 (30.5)	15 (34.8)	46 (29.3)	0.481
Blinding				0.253
Single-blinded	46 (23.0)	6 (13.9)	40 (25.5)	
Double-blinded	99 (49.5)	26 (60.5)	73 (46.5)	
Multi-blinded	10 (5.0)	1 (2.3)	9 (5.7)	
Un-blinded	45 (22.5)	10 (23.3)	35 (22.3)	

[a] Median (interquartile range),

[b] 32 trials have more than one financial support.

We planned to use a logistic regression to examine the association of study characteristics with using vs. not using statistically advanced methods. In our regression analysis, we included 6 prespecified study characteristics, i.e., 6 variables, including sample size ($\geq$ 100 vs. $<$ 100), involvement of methodologists, type of funding (government vs. other), journal type (higher impact vs. lower impact), type of intervention (pharmaceutical vs. others), and protocol availability. Data from these analyses were reported as OR with 95% confidence intervals (CIs). Our *a priori* hypotheses based on expert opinions and similar methodological studies [21,22] were as follows: trials with larger sample sizes, involvement of methodologists, government funding, higher impact journal, drug intervention and availability of trial protocol were more likely to use statistically advanced methods. For all statistical tests, a two-tailed α level of 0.05 was used. Statistical comparisons with $P < 0.05$ were considered statistically significant. Statistical analyses were undertaken in SAS 9.4 software.

3. Results

Through search of PubMed, 4262 records were identified, and 330 studies met the eligibility criteria. Using a stratified sampling strategy, we included all the 43 trials published in the higher impact journals and a random sample of 157 trials from lower impact journals (Fig. 1, Appendix B).

Of the 200 trials, 101 (50.5%) were multicenter studies, 24 (12.0%) were international studies, and 98 (49.0%) assessed drug effects (Table 1). The median sample size was 148 (interquartile range (IQR) 66–249), and 118 (59.0%) trials had sample sizes less than 100. The median follow-up was 180 days (IQR 56–365). A total of 186 (93.0%) trials were registered, 100 (50.0%) trial protocols were publicly available, and 97 (48.5%) provided a statistical analysis plan. Most trials received financial support, among which 96 (48.0%) received government funding and 45 (22.5%) received industry funding (Table 1).

3.1. Characteristics of the repeatedly measured continuous primary outcomes

Table 2 presents the characteristics of repeatedly measured continuous primary outcomes in the included RCTs. The mean number of repeated measures was 5.46, and 187 (93.5%) trials included baseline measures. Laboratory examinations (n = 78, 39.0%), symptoms (n = 77, 38.5%) and functional status (n = 27, 13.5%) were the most frequently used primary outcomes.

Of the 200 trials, 98 (49.0%) used PROs as the primary outcome; 125 (62.5%) used the raw value of repeated measures for primary outcome as opposed to that 68 (34.0%) used transformed data including absolute or percentage change from baseline. One hundred forty-two (71.0%) trials prespecified time point of the primary outcome in the method. 104 (52.0%) used multi-time points for primary outcomes.

Trials published in the higher impact journals were less likely to use the raw value of repeated measures (37.2% vs. 69.4%; $P < 0.001$), were more likely to specify the time point of primary outcome in the method (100.0% vs. 63.1%; $P < 0.001$), and were less likely to use multiple time points for defining primary outcome (16.3% vs. 61.8%; $P < 0.001$).

3.2. Statistical analyses of the repeatedly measured primary outcomes

Among the 200 trials, 96 (48.0 %) assessed treatment effect at a single point, and 104 (52.0%) assessed treatment effect over a span of multiple observations (Fig. 2 and Table 3). Of the 96 trials assessing treatment effect at a single time point, 39 (41.0%) used statistically advanced methods, and the others used conventional methods. Of the 104 trials for assessing treatment effect across a span of multiple observations, 48 (46.1%) used statistically advanced methods (Table 3). In total, 113 (56.5%) trials did not use statistically advanced methods in the primary analyses, in which cases only 10 (8.9%) used statistically ad-

Table 2. Characteristics of the continuous primary outcomes with repeated measuresa

Characteristics	All (N, %)	Higher impact journals (N, %)	Lower impact journals (N, %)	*P* value
Number of repeated measures: Mean (SD)	5.46 (3.4)	6.2 (3.3)	5.3 (3.4)	0.109
Included baseline measure of primary outcome	187 (93.5)	42 (97.7)	145 (92.4)	0.306
Type of the primary outcomes				<0.001
Laboratory Examinations	78 (39.0)	22 (51.2)	56 (35.7)	
Symptoms	77 (38.5)	9 (20.9)	68 (43.3)	
Quality of life	11 (5.5)	7 (16.3)	4 (2.6)	
Functional status	27 (13.5)	3 (7.0)	24 (15.3)	
Other	7 (3.5)	2 (4.7)	5 (3.2)	
Whether the primary outcome was a patient-reported outcome (PRO)				0.476
Yes	98 (49.0)	19 (44.2)	79 (50.3)	
No	102 (51.0)	24 (55.8)	78 (49.7)	
Form of the primary outcome				<0.001
Raw value	125 (62.5)	16 (37.2)	109 (69.4)	
Absolute change from baseline	58 (29.0)	21 (48.8)	37 (23.6)	
Percent change from baseline	10 (5.0)	4 (9.3)	6 (3.8)	
Other	7 (3.5)	2 (4.7)	5 (3.2)	
Was the time point specified for primary outcome in the method?				<0.001
Yes	142 (71.0)	43 (100.0)	99 (63.1)	
No	58 (29.0)	0 (0)	58 (36.9)	
Which time points were used for primary outcomes?				<0.001
Multiples time points	104 (52.0)	7 (16.3)	97 (61.8)	
Single time point only	96 (48.0)	36 (83.7)	60 (38.2)	
If single time point was used for primary outcome, whether multiple time points of the outcome variable were included for secondary analyses				0.241
Yes	29 (30.2)	13 (36.1)	16 (26.7)	
No	67 (69.8)	23 (63.9)	44 (73.3)	

vanced methods for sensitivity analyses. A total of 187 trials included baseline value of the primary outcome, of which 88 (47.1%) did not adjust for baseline value. Of those 99 trials adjusting for baseline value, 68 (68.7%) included baseline value as a covariate and 31 (31.3%) as a response variable for adjusting.

Trials published in the higher impact journals were more likely to use statistically advanced methods (treatment effect at a single time point: 55.6% vs. 31.7%, $P = 0.021$; treatment effect across multiple time points: 85.7% vs. 43.3%, $P = 0.047$), use statistically advanced methods for sensitivity analyses (41.2% vs. 3.1%; $P < 0.001$), and adjust for baseline value (83.3% vs. 44.1%; $P < 0.001$).

Additionally, we found that among the 97 trials reporting statistical analysis plan, the primary analyses were inconsistent with the plan in 15 (15.5%) trials. Five of these trials with protocol deviation were published in higher impact journals, in which these trials originally planned to use traditional methods but changed to use statistically advanced methods in the final analyses (Appendix Table 1).

3.3. Statistical details among trials using statistically advanced methods

In further analyzing the 87 trials using statistically advanced methods for primary analysis, 49 (56.3%) did not assume correlation structure for model, most of which (65.8%) did not assess the fitting of correlation structure, and most of which (55.3%) assumed unstructured correlation structure. In total, 76 (87.4%) trials described independent variables (i.e., treatment, time, or treatment*time), 67 (77.0%) assessed the interaction between treatment and

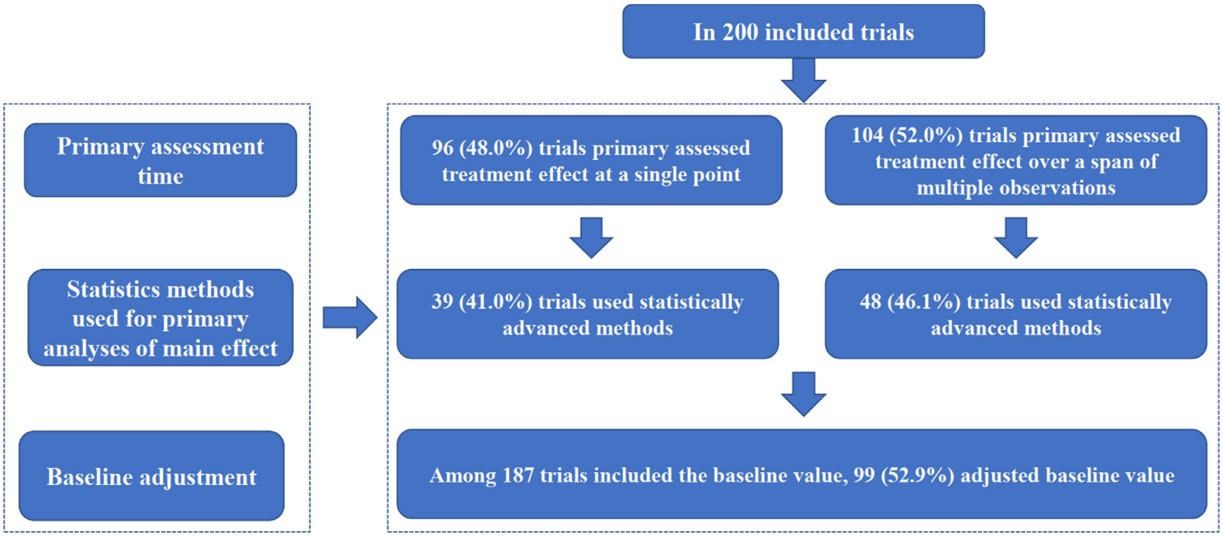
In 200 included trials
Primary assessment time
Statistics methods used for primary analyses of main effect
Baseline adjustment
96 (48.0%) trials primary assessed treatment effect at a single point
104 (52.0%) trials primary assessed treatment effect over a span of multiple observations
39 (41.0%) trials used statistically advanced methods
48 (46.1%) trials used statistically advanced methods
Among 187 trials included the baseline value, 99 (52.9%) adjusted baseline value

Fig. 2. The reporting of key methodological information by included studies (N = 200). For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

time variable. Of the 73 (83.9%) trials included time variable, 44 (60.3%) used categorical time variable, 18 (24.6%) were continuous and 11 (15.1%) were unclear. Among 18 trials including a continuous time variable, 9 (50.0%) assessed if a non-linear relationship over time was present (Table 4). Trials published in higher impact journals were more likely to specify correlation structure (61.5% vs. 36.1%, $P = 0.028$) and consider the interaction (92.3% vs. 70.5%, $P = 0.027$).

3.4. Characteristics associated with the use of statistically advanced methods

Our multivariable logistic regression analysis suggested that involvement of methodologists (68.9% vs. 39.6%, OR = 2.92, 95% CI: 1.45–5.85, $P = 0.003$), publication in higher impact journals (76.7% vs. 40.8%, OR = 3.24, 95% CI: 1.25–8.43, $P = 0.016$), and larger sample sizes (60.2% vs. 31.7%, OR = 2.21, 95% CI: 1.15–4.22, $P = 0.017$) were more likely to use statistically advanced methods (Table 5).

4. Discussion

4.1. Findings and interpretations

In this study, we found that statistical practices for handling repeated measure continuous outcomes varied substantially among current RCTs. For instance, in the initial treatment of such data, while most of the trials (62.5%) used the raw values of repeated measures for analyses, a relatively large proportion (34%) of trials used transformed data, including absolute or percentage change from baseline. We also found that the interest in assessing treatment effect over a span of time points vs. that on a specific time point is nearly an equal split among trial investigators. However, looking into details, we found that trials in top medical journals were much more likely to focus on treatment effects at a single time point and those in lower impact journals are more inclined to examine treatment effect over a span of multiple time points.

We also found that the planning and statistical analyses of continuous primary outcomes with repeated measures warrant substantial improvements. For instance, nearly one third of trials did not specify the time points for the continuous primary outcome in the method, and this issue was in particular a case in lower impact journals. In analyzing the repeated measure data, more than half of trials (56.5%) did not use statistically advanced methods in primary analysis. This is particularly concerning even if the interest was to examine treatment effect over multiple time points, and we identified 56 trials were in this case which were almost all published in lower impact journals. One more serious issue was that approximately half (45.5%) of trials published in lower impact journals used *t*-test or similar approaches to conduct multiple tests at each time point, which substantially increased false-positive findings.

When assessing treatment effect at a single time point, trials published in higher impact journals are more likely to use statistically advanced methods. Even if traditional methods were used in such trials, the ANCOVA model became the primarily used statistical method. Additionally, these trials were more likely to include statistically advanced methods in the sensitivity analyses if they were not used in the primary ones.

When analyzing continuous outcomes with repeated measures, baseline values were often correlated with follow-up measures; adjusting for baseline value has been shown to remove conditional bias for assessing treatment effect and improve efficiency over unadjusted compar-

Table 3. Statistical analyses of repeatedly measured continuous primary outcome.

Items	All (N, %)	Higher impact journals (N, %)	Lower impact journals (N, %)	*P* value
Statistical methods used for primary analyses of main effect				
Treatment effect at a single time point (n = 96)				0.021
Using statistically advanced methods	39 (41.0)	20 (55.6)	19 (31.7)	
MMRM	35 (89.7)	19 (95.0)	16 (84.2)	
GEE	4 (10.3)	1 (5.0)	3 (15.8)	
Using conventional methods	57 (59.0)	16 (44.4)	41 (68.3)	
t-test or Mann-Whitney U test or ANOVA	20 (35.1)	1 (6.3)	19 (46.3)	
Repeated measures ANOVA with Bonferroni correction	1 (1.8)	0 (0)	1 (1.8)	
ANCOVA	25 (43.9)	11(68.7)	14 (34.2)	
Other[1]	11 (19.3))	4 (25.0))	7 (17.1)	
Treatment effect across multiple time points (n = 104)				0.047
Using statistically advanced methods	48 (46.1)	6 (85.7)	42 (43.3)	
MMRM	42 (87.5)	6 (100.0)	36 (85.7)	
GEE	6 (12.5)	0 (0)	6 (14.3)	
Using conventional methods	56 (53.9)	1 (14.3)	55 (56.7)	
t-test or Mann-Whitney U test or ANOVA for each time point	25 (44.6)	0 (0)	25 (45.5)	
Repeated measures ANOVA	23 (41.1)	0 (0)	23 (41.8)	
ANCOVA	3 (5.4)	0 (0)	3 (5.4)	
Other[2]	5 (8.9)	1 (100.0)	4 (7.3)	
Whether sensitivity analyses used advanced methods if they were not used in primary analyses (n = 113)				<0.001
Yes	10 (8.9)	7 (41.2)	3 (3.1)	
No	103 (91.1)	10 (58.8)	92 (96.9)	
Whether baseline value was adjusted (n = 187[a])				<0.001
Yes	99 (52.9)	35 (83.3)	64 (44.1)	
No	88 (47.1)	7 (16.7)	81 (55.9)	
If yes, methods for adjusting baseline value (n = 99)				<0.001
Included baseline as a covariate	68 (68.7)	30 (85.7)	38 (59.4)	
Included baseline as a response	31 (31.3)	5 (14.3)	26 (40.6)	
Whether the primary analysis was consistent with statistical analysis plan (n = 97)				0.779
Yes	82(84.5)	36 (85.7)	46 (83.6)	
No	15(15.5)	6(14.3)	9 (16.4)	

MMRM, mixed-effect model for repeated measures; GEE, generalized estimating equation.

Note: Other[1] included mixed model for center only, linear regression, mean difference with 95% CI.

Other[2] included Bayesian analysis, mixed model for center only, linear regression.

[a] 187 RCTs included the baseline value.

isons [23,24]. However, in our study, nearly half of RCTs (47.1%) did not adjust for baseline outcome value, and those published in lower impact journals were more likely to neglect this adjustment. This finding is similar to the analysis of longitudinal trials in rehabilitation post-stroke [25].

In thoroughly examining the statistical details among trials that used statistically advanced methods, we have identified several methodological issues that need further improvements, including the specification of correlation structure and assessment of the fitting of correlation structure. We also found that handling of key independent variables – including treatment, time, and their interactions – needed further improvement.

All the above findings clearly suggested that the patterns of handling continuous primary outcomes with re-

Table 4. Statistical details among 87 trials using statistically advanced methods for primary analyses.

Items	All (N, %)	Higher impact journals (N, %)	Lower impact journals (N, %)	*P* value
Whether correlation structure was assumed				0.028
Yes	38 (43.7)	16 (61.5)	22 (36.1)	
No	49 (56.3)	10 (38.5)	39 (63.9)	
Type of correlation structure specified				0.715
Unstructured correlation structure	21 (55.3)	11 (68.8)	10 (45.5)	
Autoregressive correlation structure	7 (18.4)	2 (12.5)	5 (22.7)	
Exchangeable correlation structure	2 (5.3)	1 (6.2)	1 (4.6)	
Compound symmetry correlation structure	4 (10.5)	1 (6.2)	3 (13.6)	
Correlation structure based on AIC/BIC	4 (10.5)	1 (6.2)	3 (13.6)	
Whether the fit of correlation structure was assessed				0.743
Yes	13 (34.2)	5 (31.3)	8 (36.4)	
No	25 (65.8)	11 (68.7)	14 (63.6)	
Whether the independent variables were described in methods				0.107
Yes	76 (87.4)	25 (96.2)	51 (83.6)	
No	11 (12.6)	1 (3.8)	10 (16.4)	
Independent variables				
Treatment	76(87.4)	25(96.2)	51(83.6)	0.107
Time	73(83.9)	24(92.3)	49(80.3)	0.164
Treatment*time	67(77.0)	24(92.3)	43(70.5)	0.027
Type of time variable				0.115
Categorical	44(60.3)	18(75.0)	26(53.1)	
Continuous	18(24.6)	5(20.8)	13(26.5)	
Unclear	11(15.1)	1(4.2)	10(20.4)	
If continuous, whether a non-linear development overtime was considered				0.599
Yes	9(50.0)	2(40.0)	7(53.8)	
No	9(50.0)	3(60.0)	6(46.2)	

Table 5. Factors associated with the use of statistically advanced methods.

Study characteristics	Frequency	OR (95% CI)	*P* value
Journal type (higher impact vs. lower impact)	76.7% vs. 40.8%	3.24 (1.25, 8.43)	0.016
Involvement of methodologist (Yes vs. No)	68.9% vs. 39.6%	2.92 (1.45, 5.85)	0.003
Sample size ($\geq$ 100 vs. $<$100)	60.2% vs. 31.7%	2.21 (1.15, 4.22)	0.017
Type of funding (government vs. other)	53.1% vs. 44.2%	1.40 (0.75, 2.62)	0.290
Type of intervention (pharmaceutical vs. others)	52.0% vs. 45.1%	1.22 (0.65, 2.30)	0.538
Protocol reported	61.0% vs. 36.0%	1.29 (0.63, 2.64)	0.488

peated measures are often heterogeneous across trials and use of statistical methods for analyzing these outcomes is far from ideal. The first issue is that many trials continued to use less desirable statistical methods, such as *t*-test, in the presence of complex repeated measure data and statistically advanced methods were less used. The second issue is that the sophistication of advanced methods seemed to have become an important obstacle for its wide use. Often the appropriate use of such method often requires deep understanding of statistical theory and strong expertise, for which methodologists are needed. Our findings also confirmed that involvement of a methodologist was associated with better use of advanced statistical methods.

4.2. Implications for research

In planning and analyzing trials that included a continuous primary outcome with repeated measures, trial investigators should prespecify analytical strategies. Clearly, use of statistically advanced methods would be more appropriate for repeated measures, because they can appropriately account for data correlations, make full use of all available data, and are flexible in dealing with missing data [4,6,21,26,27]. Furthermore, mixed-effect model for repeated measures analysis is more popular than GEE, as it could provide more information, such as estimation of the variations between individuals. Trial investigators should always bear in mind about the potential impact of baseline values; adjusting for baseline value using statistical method, regardless how continuous outcome is presented – either as the raw data or transformed into the change from baseline. The adjustment will prevent effect estimates from being biased [23,28,29].

One additional issue implicated from our study findings was that reporting of statistical details about continuous primary outcomes were suboptimal. This is partly due to the restriction of the space of academic journals. To address this issue, inclusion of more statistical details (e.g., estimation method, methods for handling missing data, model assumption and its rationale, the inclusion of variables, adjustment for baseline, sensitivity analysis) in the method of an RCT and its proposal is recommended. Additionally, CONSORT/SPIRIT (Standard Protocol Items) (and its extended versions) also need to be further refined by adding relevant items. The inclusion of such methodological details is critical for the transparent presentation of methods for all the stakeholders.

Given the sophistication of the statistically advanced methods for the continuous primary outcome, we strongly encourage collaboration between clinicians and methodologists. In all cases, trial investigators should carefully consider and report the information in the Box.

Box: Suggested approaches to handling continuous primary outcomes with repeated measures

- Prespecify the methods for analyzing repeated measure data, including the estimation method (e.g., maximum likelihood), methods for handling missing data, model assumptions, type of correlation structure (e.g., unstructured, autoregressive), and the rationale (e.g., use of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)) [30,31].
- Prespecify all variables included in the statistical model, including dependent variable and independent variables [32]. The approaches for handling time should also be clearly predefined (i.e., continuous and modeled as a piecewise linear or a polynomial versus categorical with separate dummy variables) [33,34]. The potential interaction between time and treatment should also be examined.
- Prespecify the methods for adjusting the baseline value of repeatedly measured outcomes [35].
- Conduct sensitivity analyses using alternative methods to ensure the robustness of results [36].

4.3. Strengths and limitations

To the best of our knowledge, this was the first study that systematically investigated current practice for analyzing repeatedly measured continuous outcomes in RCTs. In this study, we included all studies from higher impact journals and a random sample of studies from lower impact journals, which was a representative of all eligible studies. We used rigorous methods for searching, selecting, and data collection from a representative sample. We provided practical recommendations to help trial investigators appropriately using advanced methods.

Our study also has limitations. First, our study and inference were based on a sample from trials published in 2019. Nevertheless, it is unlikely to incur bias in our findings due to publication time. Second, we did not investigate analysis of continuous secondary outcomes; therefore our findings may not be applicable to secondary outcomes. However, it is less possible that analyses of secondary outcomes would be superior to those for primary outcomes. Third, we restricted our survey to RCTs with parallel design, thus limiting the generalizability of findings to other RCTs designs.

5. Conclusions

In summary, the current practices of handling continuous primary outcomes with repeated measures vary substantially across trials and the use of statistical methods for analyzing these outcomes is far from ideal. We have identified a number of issues about analyzing continuous primary outcomes with repeated measures and offered recommendations for improving the statistical practice about analyzing these data. The trial research community should pay more efforts to improve the planning and analysis of continuous primary outcomes with repeated measures.

Authors' contributions

R.Y. and S.X. conceived and designed the study. R.Y. and L.L. conducted the literature search. R.Y., J.Y.L., Z.Y.J., H.Y.X., Y.M.H., and Y.P.J. screened the articles and extracted the data. R.Y., Y.M. and J.Y.L. conducted the analysis and interpreted the data. R.Y. and S.X. drafted the manuscript. All authors reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by National Natural Science Foundation of China (Grant No. 72004149), Sichuan Youth

Science and Technology Innovation Research Team (Grant No. 2020JDTD0015), and 1•3•5 project for disciplines of excellence, West China Hospital, Sichuan University (Grant No. ZYYC08003).

Availability of data and materials

Details of the search strategy and data extracted from the papers are included in the additional files.

Ethics approval and consent to participate

Not applicable. Our study only used published results of the eligible randomized controlled trials.

Consent for publication

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2021.12.007.

References

[1] Pocock NS, Hughes CM, Lee BJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987;317:426–32.

[2] De Livera AM, Zaloumis S, Simpson JA. Models for the analysis of repeated continuous outcome measures in clinical trials. Respirology 2014;19:155–61.

[3] Newgard CD, Lewis RJ. Analyzing repeated measurements using mixed models. JAMA 2015;314:940.

[4] Zou B, Jin B, Koch GG, Zhou H, Borst SE, Menon S, et al. On model selections for repeated measurement data in clinical studies. Stats Med 2015;34:1621–33.

[5] Albert PS. Analysis: Longitudinal Data Analysis (Repeated Measures) in Clinical Trials: Tutorials in Biostatistics.

[6] Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the archives of general psychiatry. Arch Gen Psychiatry 2004;61:310–17.

[7] Nicolas H, André B. Violation of the sphericity assumption and its effect on Type-I error rates in repeated measures ANOVA and multi-level linear models (MLM). Front Psychol 2017;8:1841.

[8] Fitzmaurice GM, Ravichandran C. A primer in longitudinal data analysis. Circulation 2008;118:2005–10.

[9] Ma Y, Mazumdar M, Memtsoudis SG. Beyond repeated-measures analysis of variance. Reg Anesth Pain Med 2012;37:99–105.

[10] Chen G, Saad ZS, Britton JC, Pine DS, Cox RW. Linear mixed–effects modeling approach to FMRI group analysis. Neuroimage 2013;73:176–90.

[11] Mallinckrod CH, Lane PW, Dan S, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal 2008;42:303–19.

[12] Armstrong, Richard A. Recommendations for analysis of repeated-measures designs: testing and correcting for sphericity and use of manova and mixed model analysis. Ophthalmic Physiol Opt 2017;37:585–93.

[13] Maurissen JP, Vidmar TJ. Repeated-measure analyses: which one? A survey of statistical models and recommendations for reporting. Neurotoxicol Teratol 2016;59:78–84.

[14] Mascha EJ, Sessler DI. Equivalence and noninferiority testing in regression models and repeated-measures designs. Anesth Analg 2011;112:678–87.

[15] Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ 2019;364:k5267. doi:10.1136/bmj.k5267.

[16] U.S. National Library of Medicine. Abridged Index Medicus (AIM or "core clinical") journal titles. Available at http://www.nlm.nih.gov/bsd/aim.html. Accessed July 27, 2020.

[17] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55.

[18] Rombach I, Knight R, Peckham N, Stokes JR, Cook JA. Current practice in analysing and reporting binary outcome data—a review of randomised controlled trial reports. BMC Med 2020;18:147. doi:10.1186/s12916-020-01598-7.

[19] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. BMJ Br Med J 2011;342:748.

[20] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ 2012;344:e1553.

[21] Zhang Y, Flórez ID, Colunga Lozano LE, Aloweni FAB, Kennedy SA, Li A, et al. A systematic survey on reporting and methods for handling missing participant data for continuous outcomes in randomized controlled trials. J Clin Epidemiol 2017;88:57–66.

[22] Yan P, Yao L, Li H, Zhang M, Xun Y, Li M, et al. The methodological quality of robotic surgical meta-analyses needed to be improved: a cross-sectional study. J Clin Epidemiol 2019;109:20–9.

[23] Liu GF, Lu K, Mogg R, Mallick M, Mehrotra DV. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? Stats Med 2009;28:2509–30.

[24] Sauzet O, Kleine M, Williams JE. Data in longitudinal randomised controlled trials in cancer pain: is there any loss of the information available in the data? Results of a systematic literature review and guideline for reporting. BMC Cancer 2016;16:771.

[25] Sauzet O, Kleine M, Menzel-Begemann A, Exner AK. Longitudinal randomised controlled trials in rehabilitation post-stroke: a systematic review on the quality of reporting and use of baseline outcome values. BMC Neurol 2015;15:99.

[26] Schober Patrick, Vetter Thomas R. Repeated measures designs and analysis of longitudinal data: if at first you do not succeed-try, try again. Anesth Analg 2018;127:1.

[27] Ashbeck EL, Bell ML. Single time point comparisons in longitudinal randomized controlled trials: power and bias in the presence of missing data. BMC Med Res Methodol 2016;16:43.

[28] Nash R, Bunce C, Freemantle N, Doré CJ, Rogers CA, Cairns D, et al. Ophthalmic Statistics Note 4: analysing data from randomised controlled trials with baseline and follow-up measurements. Br J Ophthalmol 2014;98:1467–9.

[29] Clifton L, Clifton DA. The correlation between baseline score and post-intervention score, and its implications for statistical analysis. Trials 2019;20.

[30] Vallejo G, Fernández MP, Livacic-Rojas PE, Tuero-Herrero E. Selecting the best unbalanced repeated measures model. Behav Res Methods 2010;43:18–36.

[31] Lu K, Mehrotra DV. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. Stats Med 2010;29:474–88.

[32] Althouse AD, Below JE, Claggett BL, Cox NJ, de Lemos JA, Deo RC, et al. Recommendations for Statistical Reporting in Cardiovascular Medicine: a special report From the American

Heart Association. Circulation 2021;144:e70–91 CIRCULATIONAHA121055393. doi:10.1161/CIRCULATIONAHA.121.055393.

[33] Harring JR, Blozis SA. Fitting correlated residual error structures in nonlinear mixed-effects models using SAS PROC NLMIXED. Behav Res Methods 2014;46:372–84.

[34] Jang H, Conklin DJ, Kong M. Piecewise nonlinear mixed-effects models for modeling cardiac function and assessing treatment effects. Comput Methods Programs Biomed 2013;110:240–52.

[35] Committee for Proprietary Medicinal PCommittee for Proprietary Medicinal Products (CPMP): points to consider on adjustment for baseline covariates. Stat Med 2004;23:701–9.

[36] de Souza RJ, Eisen RB, Perera S, Bantoto B, Bawor M, Dennis BB, et al. Best (but oft-forgotten) practices: sensitivity analyses in randomized controlled trials. Am J Clin Nutr 2016;103:5–17.