Journal of Clinical Epidemiology 121 (2020) 45–54

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Inferential reproduction analysis demonstrated that "paracetamol for acute low back pain" trial conclusions were reproducible

Marco Schreijenberg[a,*], Alessandro Chiarotto[a], Katya A.L. Mauff[b], Chung-Wei Christine Lin[c], Christopher G. Maher[c], Bart W. Koes[a,d]

[a]*Erasmus MC, Department of General Practice, University Medical Center Rotterdam, PO box 2040, 3000 CA, Rotterdam, South Holland, the Netherlands*
[b]*Erasmus MC, Department of Biostatistics, University Medical Center Rotterdam, PO box 2040, 3000 CA, Rotterdam, South Holland, the Netherlands*
[c]*Sydney School of Public Health, Institute for Musculoskeletal Health, The University of Sydney, PO Box M179, Sydney, New South Wales 2050, Australia*
[d]*Center for Muscle and Joint Health, University of Southern Denmark, Odense, South Denmark, Denmark*

Accepted 20 January 2020; Published online 23 January 2020

Abstract

Objectives: The aim of this study was to reanalyze and reinterpret data obtained in Paracetamol in Acute Low Back Pain (PACE), the first large randomized controlled trial evaluating the efficacy of paracetamol in acute low back pain, to assess the inferential reproducibility of the original conclusions.

Study Design and Setting: Mixed effects models were used to reanalyze pain intensity (primary outcome; 11-point Numeric Rating Scale) and physical functioning, health-related quality of life, sleep quality, and time until recovery (as secondary outcomes), according to the intention-to-treat principle. The original authors of the PACE study were not involved in the development of the methods for this reanalysis.

Results: The reproduction analyses indicated no effect of treatment on pain intensity and confidence intervals excluded clinically worthwhile effects (adjusted main effect for regular paracetamol vs. placebo 0.00 [−0.02, 0.01; $P = 0.85$]; adjusted main effect for paracetamol as-needed vs. placebo 0.00 [−0.02, 0.01; $P = 0.92$]). Similar results were obtained for all secondary outcomes.

Conclusion: This study indicates that the conclusions of the PACE trial are inferentially reproducible, even when using a different analytical approach. This reinforces the notion that the management of acute low back pain should focus on providing patients advice and reassurance without the addition of paracetamol.

Keywords: Low back pain; Paracetamol; General practice; Inferential reproduction; Reanalysis

Competing interests: C.G.M. is an investigator on a clinical trial that received FlexEze heat wraps at no cost from the supplier. The other authors declare no competing interests.

Funding: The Paracetamol in Acute Low Back Pain (PACE) trial was an investigator-initiated study funded by a project grant from the National Health and Medical Research Council of Australia. GlaxoSmithKline Australia provided subsequent supplementary funding and the paracetamol and matched placebo. C.M. was supported by a Principal Research Fellowship from Australia's National Health and Medical Research Council (APP1103022). C.-W.C.L. is supported by a Career Development Fellowship from the National Health and Medical Research Council, Australia (APP1061400). This secondary analysis of the PACE data has been supported by a program grant of the Dutch Arthritis Foundation and by the Foundation "De Drie Lichten" in The Netherlands.

Data sharing statement: PACE trial data can be made available to interested researchers on request. Requests can be directed to professor Christopher Maher (christopher.maher@sydney.edu.au). Data cannot be shared publicly because of legal and ethical restraints. Sharing of individual participant data was not included in the informed consent of the PACE trial.

Ethics approval and consent to participate: The University of Sydney Human Research Ethics Committee granted ethical approval of the PACE trial protocol. Written informed consent was provided by all participants. The PACE trial was registered with the Australian and New Zealand Clinical Trial Registry, number ACTN12609000966291.

Authors' contributions: C.L. and C.M. made substantial intellectual contributions to the development of the original study protocol, data collection, statistical analysis, and drafting the original results of the PACE trial. M.S., A.C., K.M., and B.K. all made substantial intellectual contributions to the development of the Inferential Reproduction analysis protocol. M.S. and K.M. conducted the statistical analyses for this trial. M.S. drafted the article, which was revised by A.C., K.M., and B.K. Box 1 with comments from the original authors was provided by C.L. and C.M. All authors have read and approved the final article.

* Corresponding author. Erasmus MC, Department of General Practice, University Medical Center Rotterdam, PO box 2040, 3000 CA, Rotterdam, South Holland, the Netherlands. Tel.: +31107032117; fax: +3110 704 47 66.

E-mail address: m.schreijenberg@erasmusmc.nl (M. Schreijenberg).

https://doi.org/10.1016/j.jclinepi.2020.01.010

What is new?

Key findings

- This inferential reproduction study showed that paracetamol was not more effective than placebo on core outcomes of low back pain, confirming the original results of Paracetamol in Acute Low Back Pain (PACE), the first randomized controlled trial investigating the efficacy of paracetamol for acute low back pain.

What this adds to what was known?

- The present study is the first independent inferential reproduction analysis in the field of low back pain research. Although a different statistical approach was used, our analyses unequivocally confirm the conclusions from the PACE trial.
- As the PACE trial remains the first and only trial showing that paracetamol had no effect on the outcomes of acute low back pain and has, therefore, been highly influential on clinical practice guidelines, the reproducibility of its results is of vital importance.

What is the implication and what should change now?

- The outcome of this study is highly relevant because several national clinical practice guidelines still recommend paracetamol as first-choice analgesic for acute low back pain.
- The available evidence suggests that the management of acute low back pain should focus on providing patients advice and reassurance without the addition of paracetamol.

1. Introduction

Paracetamol (acetaminophen) used to be the first-choice analgesic for acute low back pain (LBP), but several recent clinical practice guidelines have abandoned this recommendation because of new evidence about its lack of efficacy [2−6]. This evidence came from a 2016 Cochrane Review, which mainly based its results on the Paracetamol for Acute Low Back Pain (PACE) trial, the first large placebo-controlled randomized controlled trial (RCT) concerning the efficacy of paracetamol in acute LBP [7,8]. As this trial was highly influential on recent guidelines, the reproducibility of its results is of great importance [9].

Although the importance of reproducibility of scientific results is universally agreed upon, the terminology describing different types of reproducibility is not. In 2016, Goodman et al introduced their ‘‘new lexicon for research reproducibility,’’ in which they described three types of reproducibility: methods reproducibility, results reproducibility, and inferential reproducibility [10,11]. Methods reproducibility refers to the reproduction of an analysis using the same data, analysis plan, and code, with the only difference being the analyst [12,13]. In results reproduction (also called ‘‘replication’’), new data are collected in the same population and consequently analyzed using the same analysis plan [12,13]. Finally, inferential reproducibility is the making of knowledge claims of similar strength from either a study replication or reanalysis of original data [10]. In clinical research, reproduction studies are often the exception rather than the rule. However, early acceptance of scientific claims that are subsequently not reproducible may lead to harms or even to unwarranted ‘‘medical reversal’’ [14]; furthermore, reproduction is important in case only little evidence exists about a certain topic [15].

In the PACE trial, methods reproducibility was already addressed, as *‘‘Two statisticians who were masked to allocation independently did statistical analyses…’’* [8]. Another RCT evaluating the result of the PACE trial (called the PACE Plus trial) was discontinued in 2017 because of insufficient recruitment of participants [16,17]. The primary outcome in the PACE trial was time until recovery from LBP, but this outcome is not among the outcome domains most relevant to patients with LBP [18]. A core outcome set for LBP, published after the PACE trial had already been completed, including pain intensity, physical functioning, health-related quality of life (HRQoL), and number of deaths as core outcome domains [19]. The first three core domains were included in the PACE original analysis as secondary outcomes, while no patients died during trial participation [8]. In the analysis plan of the discontinued PACE Plus trial, pain intensity recorded in the daily pain diary was the primary outcome [16]. The original analysis of the PACE trial reported results for pain intensity at 1, 2, 4, and 12 weeks of follow-up and presented only part of the data from the pain diary (up to 14 days of follow-up) in the appendix; not all collected diary data were used [8]. The aim of this study was to reanalyze the original data obtained in the PACE trial to assess the inferential reproducibility of results obtained in PACE.

2. Methods

2.1. Participants and data collection in the PACE trial

A brief description of participants and data collection in the original PACE trial is provided here; for a detailed description, see the original manuscripts [8,20]. The PACE trial was a randomized, placebo-controlled clinical trial that was conducted from November 2009 until March 2013 in Sydney, Australia. This RCT was conducted in a multicenter setting with a double-dummy design, which means

that in all treatment groups, patients received "regular" medication (or placebo) to take three times a day as well as medication (or placebo) to take as needed for pain. The study protocol, analysis plan, and main results of the PACE trial have been published [8,20,21]. A total of 1,652 patients with a new episode of at least moderate-intensity LBP were randomly allocated to take paracetamol regularly (1,330 mg of modified-release paracetamol three times a day, $n = 550$, which all were analyzed) or as-needed (up to a maximum of 1,000 mg of regular paracetamol four times a day, $n = 549$, of which 546 were analyzed), or to receive placebo ($n = 553$, of which 547 were analyzed). Placebo tablets were identical in appearance to paracetamol tablets but did not contain the active component. Participants were instructed to use study medication until they had experienced seven consecutive days with pain scores of 0 or 1 out of 10 (measured on a numerical rating scale [NRS]), or for a maximum of 4 weeks, whichever occurred first. During the trial, participants, clinicians, and researchers remained blinded to allocation of treatment.

Pain scores and the number of tablets used were recorded by participants into a daily pain and drug diary until recovery or for a maximum of 12 weeks. At 1, 2, 3, and 12 weeks after randomization, follow-up questionnaires were collected.

2.2. Outcomes used in this reanalysis

For this reanalysis, the predefined and published analysis plan from the PACE Plus trial was used [16]. The PACE Plus trial was a randomized, placebo-controlled clinical trial that aimed to reproduce the results obtained in the PACE trial; however, this trial was discontinued because of insufficient patient recruitment [16,17]. As the groups and outcomes were similar but not identical between PACE and PACE Plus, we present primary and secondary outcomes of the current reproduction analysis here. The primary outcome of the PACE Plus trial was LBP-intensity measured with an 11-point NRS (score range 0−10; higher score means more pain) [22]; this outcome was, therefore, used as the primary outcome for this study. As specified in the PACE Plus analysis plan, data from the daily pain and drug diary collected up to 28 days of follow-up were used for the current analyses rather than data from the follow-up questionnaires that were collected after 1, 2, 4, and 12 weeks. Secondary outcome measures from the PACE Plus analysis plan that were also collected in the PACE trial were as follows:

- Time to recovery assessed with the daily LBP severity scores. Recovery is defined as the first day of 0 or 1 pain intensity on a 0−10 pain scale, maintained for seven consecutive days (primary outcome of the PACE trial).
- Physical functioning measured with the Roland Morris Disability Questionnaire (RMDQ; score range 0−24; higher score indicates poorer functioning) [23].
- HRQoL measured with the physical and mental component summary scores of the Short Form 12 (SF-12, range 0−100; higher score indicates better HRQoL) [24].
- Sleep quality measured with a 4-point Likert scale derived from the Pittsburgh Sleep Quality Index. Scores were dichotomized into good sleep quality (score 1: "very good" and 2: "fairly good") and poor sleep quality (score 3: "fairly bad" and 4: "very bad") [25].

2.3. Statistical analysis

The researchers who performed the original analysis of the PACE trial were not involved in the reanalysis of the data; two coauthors of the original trial (CM and CL) involved in this study were only allowed to view the results and to give their comments in a separate box at the end of the article after the reanalysis and interpretation had already been completed. The statistical analysis was performed according to the intention-to-treat (ITT) principle. The software used for the statistical analysis was R version 3.5.3 (R Core Team, 2019 [26]). An overview of differences between the original analysis and the current inferential reproduction analysis can be found in Table 1.

2.4. Primary statistical analysis

For clinical effectiveness, the between-group differences for the primary outcome, LBP intensity, were evaluated using a repeated measurements analysis with Poisson mixed effects models with an adequate specification of the fixed and random effects structures to account for possible nonlinear effects. The covariance structure was unstructured. Poisson mixed effects models rather than linear mixed effects models were used as pain data were found to be zero-inflated and nonnormally distributed (Supplementary Figure 1A); Poisson models have been demonstrated to be more appropriate for the analysis of zero-inflated ordinal data such as data obtained from the NRS [27,28]. The GLMMadaptive R package was used to create the Poisson mixed effects models [29]. Results are presented as corrected coefficients for treatment with corresponding 95% confidence intervals (CIs) and P values. A 20% extra improvement in pain in participants receiving paracetamol compared with those receiving placebo was considered to be the smallest worthwhile effect [30,31].

2.5. Secondary statistical analysis

We used Poisson mixed effect models for physical functioning as data obtained using the RMDQ was found to be zero-inflated and nonnormally distributed (see distribution of data in Supplementary Figure 1B), linear mixed effect

Table 1. Differences between the original analysis by Williams et al and the current inferential reproduction analysis for outcomes of PACE

	Original analysis (Williams et al., Lancet 2014)				Inferential reproduction analysis			
Outcome	**P/S**	**Method**	**Presented outcome**	**SA**	**P/S**	**Method**	**Presented outcome**	**SA**
Time until recovery	P	Cox proportional hazards model; recovery time and status considered after 12 wk of follow-up	Hazard ratios for recovery for overall comparisons between groups after 12 wk of follow-up	No	S	Cox proportional hazards model; recovery time and status considered after 28 days of follow-up	Hazard ratios for recovery for overall comparisons between groups after 28 days of follow-up	Yes
Pain intensity	S	Linear mixed model on pain data at 1-, 2-, 4-, and 12-wk follow-up	Mean and SD in each group at 1-, 2-, 4- and 12-wk follow-up; results for analysis of diary data presented up to 14 days	No	P	Poisson mixed model on pain diary data up to 28 days of follow-up	Coefficients for change in log average pain intensity for overall comparisons between groups	Yes
Physical functioning	S	Linear mixed model	Mean and SD in each group at 1-, 2-, 4-, and 12-wk follow-up	No	S	Poisson mixed model	Coefficients for change in log average physical functioning for overall comparisons between groups	Yes
Sleep quality	S	Log binomial regression	Fractions and percentages of poor sleep quality in each group at 1-, 2-, 4-, and 12-wk follow-up	No	S	Logistic regression	Odds ratios for poor sleep quality for overall comparisons between groups	No
HRQoL	S	Linear mixed model	Mean and SD in each group at 1-, 2-, 4-, and 12-wk follow-up	No	S	Linear mixed model	Coefficients for change in average HRQoL for overall comparisons between groups	No
Global rating of symptom change	S	Linear mixed model	Mean and SD in each group at 1-, 2-, 4-, and 12-wk follow-up	No	NA	-	-	-

Abbreviations: HRQoL, health-related quality of life; NA, not analyzed (not in PACE Plus trial protocol); P, Primary outcome; PACE, Paracetamol in Acute Low Back Pain; S, secondary outcome; SA, subgroup analyses for participants with severe pain intensity (defined as NRS $\geq$ 7) or severe impairment of physical functioning (defined as RMDQ $\geq$ 16) at baseline; SD, standard deviation.

models for HRQoL, a logistic regression model for sleep quality and a Cox proportional hazards model for time until first recovery from LBP to assess between-group differences [27,28]; respective R packages used for the analyses were GLMMadaptive, lme4, Stats, and Survival [26,29,32,33]. Sensitivity to missing data in the recovery analysis was investigated by calculating a best-case scenario and a worst-case scenario for recovery from LBP. In the best-case scenario, we assumed all missing participants recovered after the first day of follow-up. In the worst-case scenario, we assumed none of the missing participants recovered within 28 days of follow-up.

As specified in the PACE Plus study protocol, exploratory subgroup analyses were conducted for participants with severe LBP intensity (defined as NRS $\geq$ 7) or severe impairment of physical functioning (defined as RMDQ $\geq$ 16) at baseline [16]; for these subgroups, estimates were obtained for LBP intensity, physical function, and time until recovery using Poisson mixed effects models and Cox proportional hazard analyses, respectively. The results are presented as corrected coefficients for treatment with corresponding 95% CIs and *P* values. A 20% extra improvement in pain from baseline in participants receiving paracetamol compared with those receiving placebo was considered to be the smallest worthwhile effect [30,31].

3. Results

Reproduced baseline characteristics of participants of the PACE trial can be found in Table 2. Treatment groups were comparable at the start of the trial.

Results for the ITT analysis of the primary and secondary outcomes are presented in Table 3. Comparisons between regular paracetamol and placebo, paracetamol as-needed and placebo, and regular paracetamol and paracetamol as-needed are presented. As an example, the coefficient for regular paracetamol vs. placebo (0.00, 95% CI: −0.02, 0.01) is interpreted as no change in the log average pain intensity for regular paracetamol when compared with placebo, when all other predictors remain constant.

Pain intensity diary data were available for 1,601 participants (538 from the regular paracetamol group, 530 from

Table 2. Patients and episode characteristics

Patient characteristics	Regular group (*N* = 550)	As-needed group (*N* = 546)	Placebo group (*N* = 547)
Age (y)	44.1 (14.8), *N* = 550	45.5 (16.5), *N* = 546	45.4 (15.9), *N* = 546
Women	263/547 (48%)	256/546 (47%)	245/544 (45%)
Private health insurance	275/550 (50%)	240/545 (44%)	248/544 (46%)
Currently employed	424/550 (77%)	403/546 (74%)	389/542 (72%)
Household income per week (per year)			
Negative or no income	19/540 (4%)	11/531 (2%)	22/531 (4%)
AUD 1–649 (1–33,799)	133/540 (25%)	167/531 (31%)	168/531 (32%)
AUD 650–1,699 (33,800-88,399)	243/540 (45%)	243/531 (46%)	226/531 (43%)
AUD 1,700–3,999 (88,400-207,999)	119/540 (22%)	92/531 (17%)	97/531 (18%)
≥AUD 4,000 (≥208,000)	26/540 (5%)	18/531 (3%)	18/531 (3%)
Use of drugs for another disorder	201/550 (37%)	227/543 (42%)	202/544 (37%)
Episode characteristics	Regular group (*N* = 550)	As-needed group (*N* = 546)	Placebo group (*N* = 547)
Days since onset of pain	10.1 (10.1), *N* = 550	9.8 (10.0), *N* = 546	9.7 (9.8), *N* = 546
Number of previous episodes	6.3 (13.7), *N* = 547	7.2 (14.9), *N* = 544	7.2 (16.8), *N* = 544
Presence of pain extending beyond the knee	108/547 (20%)	113/546 (21%)	99/544 (18%)
Number of days reduced usual activity	3.8 (6.5), *N* = 548	3.6 (5.9), *N* = 546	3.4 (5.3), *N* = 545
Physical functioning (RMDQ)	12.8 (5.6), *N* = 543	13.2 (5.4), *N* = 532	13.3 (5.5), *N* = 531
Feelings of depression in last week	3.2 (2.9), *N* = 547	3.1 (2.9), *N* = 546	3.1 (2.9), *N* = 546
Perceived risk of persistent pain	4.6 (2.8), *N* = 548	4.6 (2.8), *N* = 546	4.4 (2.8), *N* = 545
Back pain episode compensable	31/546 (6%)	44/543 (8%)	43/546 (8%)
Pain intensity (NRS)	6.3 (1.9), *N* = 550	6.3 (2.0), *N* = 545	6.2 (1.8), *N* = 546
Global rating of change	0.0 (2.1), *N* = 548	−0.1 (2.2), *N* = 545	−0.1 (2.1), *N* = 546
Poor sleep quality	273/549 (50%)	272/545 (50%)	272/546 (50%)
Function (Nominated Activity)	3.5 (1.7), *N* = 547	3.6 (1.9), *N* = 544	3.7 (1.9), *N* = 545
Quality of life—physical (SF-12)	42.7 (9.1), *N* = 537	41.8 (9.7), *N* = 543	42.1 (9.2), *N* = 538
Quality of life—mental (SF-12)	44.1 (7.7), *N* = 537	44.6 (7.7), *N* = 543	44.4 (7.9), *N* = 538
Credibility score (CEQ)	19.0 (4.9), *N* = 544	18.5 (5.2), *N* = 542	19.4 (4.9), *N* = 540
Expectation score (CEQ)	19.7 (5.3), *N* = 544	19.6 (5.1), *N* = 542	20.2 (5.1), *N* = 542

Abbreviations: AUD, Australian dollars; CEQ, credibility/expectancy questionnaire; LBP, low back pain; NRS, numerical rating scale; PSQI, Pittsburgh Sleep Quality Index; RMDQ, Roland Morris Disability Questionnaire; SF-12, 12-item Short Form Survey.

Data are mean (SD) or n/N (%).

the paracetamol as-needed group, and 533 from the placebo group). All treatment coefficients indicated no effect of treatment on pain intensity during 28 days of follow-up (Table 3A); no estimates exhibited between-group differences (even without correction for multiple testing). Furthermore, CIs for the coefficients were between −0.1 and +0.1 and did not include a clinically worthwhile effect of treatment with paracetamol (taken regularly or as-needed) on pain intensity when compared with placebo.

The estimates for treatment coefficients for physical functioning and HRQoL, odds ratios for poor sleep quality, and hazard ratios for recovery from LBP indicated no effect of treatment without correction for multiple testing (Table 3B). Furthermore, clinically worthwhile differences were not included in the CIs for these estimates.

A graphical representation of the effects of treatment during follow-up is shown in Figure 1; graphs were obtained from uncorrected regression models containing only treatment and time as covariates. The lines for different treatment groups are very close in all graphs (and sometimes nearly indistinguishable), emphasizing no difference in effect between paracetamol and placebo. Pain intensity (Figure 1A) steadily declines over time in all treatment groups. For physical functioning (Fig. 1B), a sharp decline can be observed during the first 4 weeks of follow-up, followed by a stable phase until 12 weeks of follow-up. Although the mental component of HRQoL remained constant during the trial (Fig. 1C), the physical component of HRQoL steadily increased during 12 weeks of follow-up, indicating an improvement of HRQoL over time (Fig. 1D). The probability of poor sleep quality steadily declined during 12 weeks of follow-up.

Figure 1F illustrates the recovery curves as well as median recovery times for the three treatment groups;

Table 3. Coefficients for effect of treatment on log average pain intensity (primary outcome) during 28 days of follow-up and for secondary outcomes during 12 wk of follow-up

A. Primary outcome	Regular paracetamol vs. placebo, β (95% CI)	Paracetamol as needed vs. placebo, β (95% CI)	Regular paracetamol vs. paracetamol as-needed, β (95% CI)
Pain intensity (NRS, scale range 0–10)	0.00 (−0.02, 0.01); $P = 0.85$	0.00 (−0.02, 0.01); $P = 0.92$	0.00 (−0.02, 0.01); $P = 0.92$
B. Secondary outcomes	Regular paracetamol vs. placebo, β (95% CI)	Paracetamol as-needed vs. Placebo, β (95% CI)	Regular paracetamol vs. paracetamol as-needed, β (95% CI)
Physical functioning (RMDQ, scale range 0–24)	−0.06 (−0.13, 0.01); $P = 0.11$	−0.03 (−0.10, 0.04); $P = 0.39$	−0.03 (−0.09, 0.04); $P = 0.46$
HRQoL—mental (SF-12)	−0.13 (−0.72, 0.47); $P = 0.67$	0.17 (−0.42, 0.76); $P = 0.58$	−0.30 (−0.89, 0.30); $P = 0.33$
HRQoL—physical (SF-12)	0.00 (−0.77, 0.77); $P = 1.00$	−0.14 (−0.91, 0.62); $P = 0.71$	0.14 (−0.62, 0.91); $P = 0.71$
Sleep quality (PSQI)	OR 1.03 (0.90, 1.19); $P = 0.62$	OR 1.04 (0.91, 1.19); $P = 0.59$	OR 1.00 (0.87, 1.14); $P = 0.97$
Time until first recovery	HR 1.02 (0.88, 1.18); $P = 0.82$	HR 1.02 (0.88, 1.19); $P = 0.76$	HR 0.99 (0.86, 1.15); $P = 0.93$

Abbreviations: HR, hazard ratio; NRS, numerical rating scale; OR, odds ratio; PSQI, Pittsburgh Sleep Quality Index; RMDQ, Roland Morris Disability Questionnaire; SF-12: Short Form 12.

All numbers rounded to two decimal places. All models were corrected for sex, age, employment status, income, use of medication for other disorders, health insurance status and back pain compensability, days since onset of pain, number of previous episodes, radiating pain beyond the knee, number of days reduced activity, feelings of depression, perceived risk of persistent pain, pain intensity, global rating of symptom change, physical functioning, patient specific function, sleep quality, credibility, expectations and physical and mental health-related quality of life (all measured at baseline).

recovery information could be obtained from pain diary information for 1,601 participants; for 13 additional patients with all pain diary data missing, a recovery date was available, yielding a total of 1,614 patients for the analysis (542 in the regular paracetamol group, 535 in the paracetamol as-needed group, and 537 in the placebo group). Of 1,614 participants, 1,186 (73%) had recovered from LBP after 28 days of follow-up. Median recovery times were 13 days (95% CI: 11–14 days), 14 days (95% CI: 13–15 days), and 12 days (95% CI: 10–14 days) in the regular paracetamol, paracetamol as-needed, and placebo groups, respectively. There was no difference between the three recovery curves (log-rank $P = 0.7$). In Supplementary Figure 2, the results of the analysis for the sensitivity to missing data were presented. The results did not substantially change in the sensitivity analyses when compared with the available data analysis.

In Table 4, results for the subgroups for severe baseline LBP intensity (defined as NRS $\geq$ 7) and severe baseline impairment of physical functioning (defined as RMDQ $\geq$ 16) are displayed. The results did not substantially change in the subgroups when compared with the main analysis. Figure 2 shows the recovery curves for these subgroups. In the severe baseline LBP intensity subgroup, 547 of 776 participants (70%) had recovered from LBP after 28 days of follow-up. Median recovery times were 14 days (95% CI: 13–19 days), 16 days (95% CI: 14–18 days), and 13 days (95% CI: 11–17 days) in the regular paracetamol, paracetamol as-needed, and placebo groups, respectively. There was no difference between the three recovery curves (log-rank $P = 0.8$). In the severe baseline impairment of physical functioning subgroup, 420 of 592 participants (71%) had recovered from LBP after 28 days of follow-up. Median recovery times were 16 days (95% CI: 13–19 days), 16 days (95% CI: 14–19 days), and 14 days (95% CI: 11–21 days) in the regular paracetamol, paracetamol as-needed, and placebo groups, respectively. There was no significant difference between the three recovery curves (log-rank $P = 0.9$).

4. Discussion

We performed an inferential reproduction analysis of data collected in the PACE trial, using the predefined and published analysis plan from the PACE Plus trial; key differences between the original analysis and the current reanalysis include a different primary outcome and different analysis methods, follow-up time points, presented outcomes, and subgroup analyses [8,16]. In our reanalysis of the PACE-trial data, the treatment of patients with acute LBP with paracetamol (taken regularly or as-needed) had no effect on pain intensity, physical functioning, HRQoL, and time until recovery from LBP when compared with placebo; our study thus confirmed the original results of the PACE trial [8].

A strength of this study is the fact that the predefined and published analysis plan from a discontinued replication trial of PACE was used [16]. Furthermore, Poisson mixed models have been demonstrated to be more appropriate for the analysis of zero-inflated ordinal data such as data obtained from the NRS and the RMDQ than linear mixed models [27,28]. A weakness of this study is the fact that the published analysis plan could not be completely used as intended because of differences between the PACE trial and the PACE Plus trial [16]. Although the PACE trial had

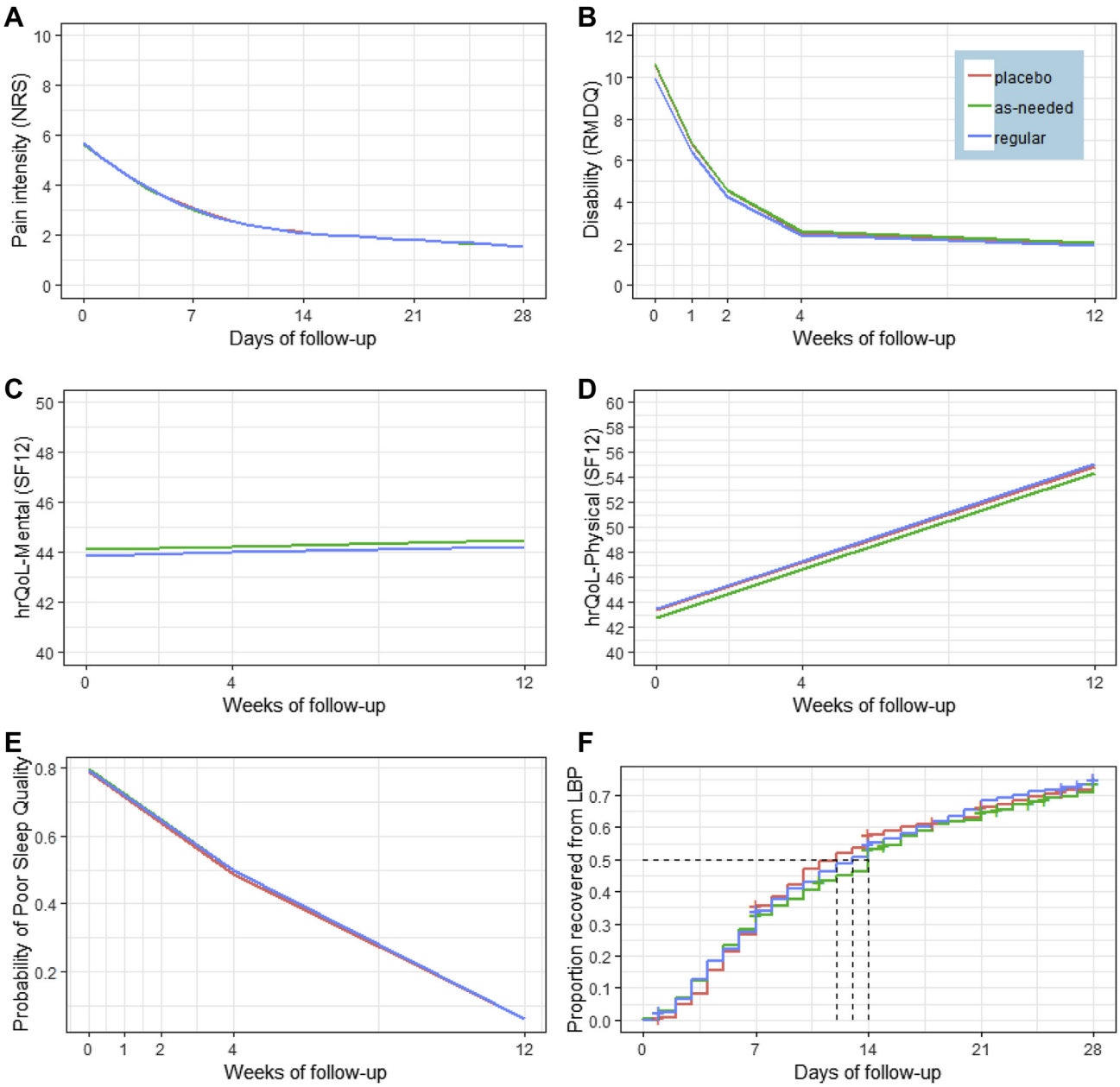

Fig. 1. Effects of treatment on core outcomes of LBP (pain intensity (A), physical functioning (B) and HRQoL (C and D), sleep quality (E), and time until first recovery from LBP (F). Graphs obtained from uncorrected regression models containing only treatment and time as covariates. Y-axis was truncated for plots B, C, D, E, and F to improve visibility of results. Red line indicates placebo group, green line indicates paracetamol as-needed group, blue line indicates regular paracetamol group. HRQoL, health-related quality of life; LBP, low back pain; NRS, numerical rating scale; PSQI, Pittsburgh Sleep Quality Index; RMDQ, Roland Morris Disability Questionnaire; SF12, Short Form 12.

three treatment groups (regular paracetamol, paracetamol as-needed, and placebo), the PACE Plus trial had four treatment groups (regular paracetamol, regular diclofenac, placebo, and advice-only). Furthermore, as mentioned in the "Methods" section, not all outcome domains were the same between both trials, meaning we could only use part of the analysis plan as well as part of the available data collected in the PACE trial; however, despite some differences, the core outcome domains and instruments for LBP were included in the reproduction analysis [19,22]. Second, authors deviated from the original protocol using Poisson mixed effect models rather than the predefined linear mixed effects models, but the nature of the data obligated this change. Finally, this reanalysis of the PACE trial focused on the ITT analysis, which remains the least biased method and, therefore, the most relevant analysis for clinical practice [34]; adherence to medication was not taken into account. However, it has been demonstrated that compliance patterns are unlikely to change the conclusions about the efficacy of paracetamol for acute LBP [35].

Table 4. Coefficients for subgroups for effect of treatment on average pain intensity (primary outcome) and time until first recovery during 28 days of follow-up and on average physical function during 12 wk of follow-up

Subgroup 1: severe baseline LBP intensity (defined as NRS ≥ 7)	Regular paracetamol vs. placebo, β (95% CI)	Paracetamol as-needed vs. placebo, β (95% CI)	Regular paracetamol vs. paracetamol as-needed, β (95% CI)
Pain intensity (NRS, scale range 0–10)	−0.02 (−0.09, 0.05); $P = 0.49$	0.00 (−0.07, 0.07); $P = 0.96$	−0.02 (−0.09, 0.05); $P = 0.53$
Physical functioning (RMDQ, scale range 0–24)	−0.01 (−0.11, 0.08); $P = 0.80$	−0.01 (−0.10, 0.09); $P = 0.88$	−0.01 (−0.10, 0.09); $P = 0.91$
Time until recovery	HR 1.04 (0.83, 1.30) $P = 0.74$	HR 1.09 (0.88, 1.36) $P = 0.44$	HR 0.95 (0.77, 1.19); $P = 0.67$
Subgroup 2: severe baseline impairment of physical functioning (defined as RMDQ ≥ 16)	Regular paracetamol vs. placebo, β (95% CI)	Paracetamol as-needed vs. Placebo, β (95% CI)	Regular paracetamol vs. paracetamol as-needed, β (95% CI)
Pain intensity (NRS, scale range 0–10)	0.00 (−0.10, 0.10); $P = 0.99$	0.03 (−0.07, 0.12); $P = 0.58$	−0.03 (−0.12, 0.07); $P = 0.59$
Physical functioning (RMDQ, scale range 0–24)	0.02 (−0.06, 0.11); $P = 0.56$	−0.03 (−0.11, 0.05); $P = 0.50$	0.05 (−0.03, 0.13); $P = 0.20$
Time until recovery	HR 1.02 (0.79, 1.30); $P = 0.89$	HR 1.08 (0.84, 1.38); $P = 0.53$	HR 0.94 (0.73, 1.21); $P = 0.64$

Abbreviations: HR, hazard ratio; LBP, low back pain; NRS, numerical rating scale; RMDQ, Roland Morris Disability Questionnaire.

Subgroups were severe LBP intensity (defined as NRS ≥ 7) and severe impairment of physical functioning (defined as RMDQ ≥ 16) at baseline. All numbers rounded to two decimal places. All models were corrected for sex, age, employment status, income, use of medication for other disorders, health insurance status and back pain compensability, days since onset of pain, number of previous episodes, radiating pain beyond the knee, number of days reduced activity, feelings of depression, perceived risk of persistent pain, pain intensity, global rating of symptom change, physical functioning, patient-specific function, sleep quality, credibility, expectations and physical and mental health-related quality of life (all measured at baseline).

As the PACE Plus protocol only specified the collection of pain diary data up to 28 days of follow-up (upon which the recovery analysis was based), the authors decided not to use any data gathered in the PACE trial after 28 days of follow-up, as this would not have been available in the PACE Plus study; furthermore, the analysis for this reproduction analysis was conducted on available data with sensitivity analyses for missing data, whereas in the original report, data were imputed to obtain complete groups for the recovery analysis. A consequence of these decisions

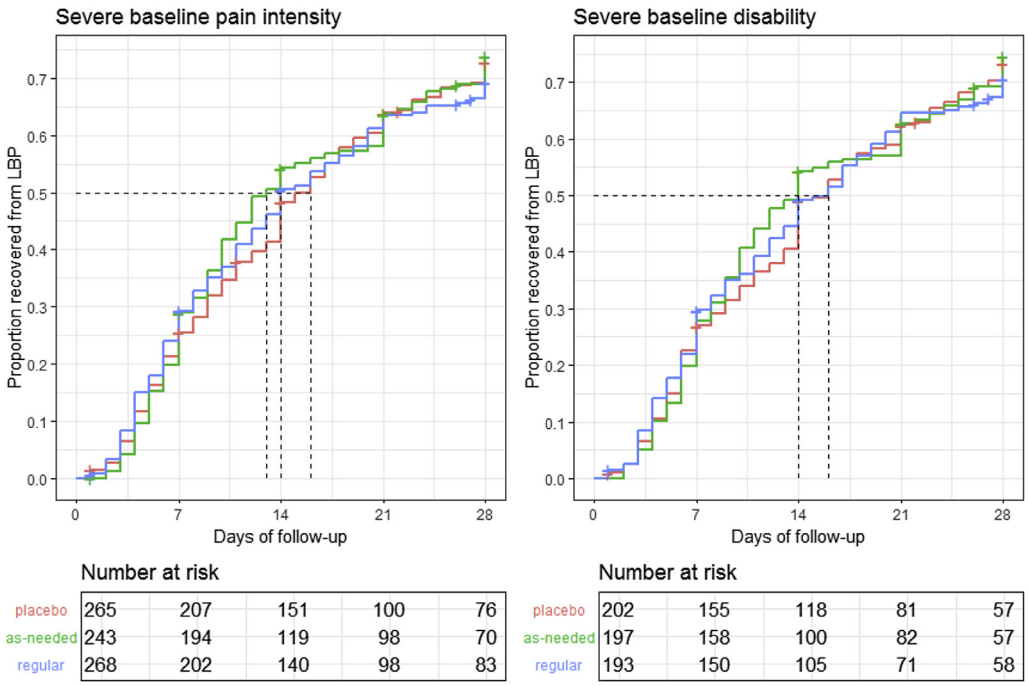

Fig. 2. Survival curves for time until first recovery in subgroups. Subgroups were (A) severe baseline LBP intensity (defined as NRS ≥ 7) and (B) severe baseline physical functioning (defined as RMDQ ≥ 16).

Box 1 Comments on this inferential reproduction analysis of PACE by the original authors

The inferential reproduction analysis of the PACE study, conducted by an independent group based on a predefined statistical analysis plan of a similar study (PACE Plus), agrees with the conclusion from the original PACE analysis—that paracetamol has no effects on pain or other core outcomes compared with placebo in patients with acute LBP.

This study joins other secondary analyses of the PACE study showing the lack of benefits of paracetamol: we have also found that paracetamol did not improve pain intensity even in patients who complied with the regular treatment regimen (article to be published in 2019), and taking paracetamol did not confer any economic benefits in patients with acute LBP [1]. However, we await the most important and currently missing step in definitively confirming the results of PACE—a replication of the PACE study.

We would encourage other trialists to make their datasets available to allow reanalysis of the data by independent groups.

is that patients who recovered after 28 days of follow-up were considered censored in the current version of the recovery analysis; this may be an explanation for the difference in median recovery times (13, 12, and 14 days in the regular paracetamol, paracetamol as-needed, and placebo groups, respectively, vs. 17, 16 and 17 days as reported in the original report).

This reanalysis of the PACE data yielded no substantially different results, and therefore, the interpretation of the PACE trial remains the same: paracetamol (taken regularly or as needed) did not improve outcomes of LBP when compared with placebo. Thus, this study supports the notion that paracetamol has a limited role in the management of acute LBP in general practice. Furthermore, this reanalysis confirms that prognosis of acute LBP is favorable, and that natural course or regression to the mean (Fig. 1), rather than pharmacological treatment, is important factors influencing core outcomes' trajectory in patients with acute LBP.

The clinical implementation of the results of the PACE trial, of course, depends on the acceptability of taking no pain medication, but receiving only reassurance and advice to stay active for patients with acute nonspecific LBP. In practice, this may be feasible if clinicians can invest more time in explaining to patients that the effects of pharmacological interventions for LBP are expected to be limited, although there is a risk of adverse events. Unfortunately, in many countries, GPs have very limited time for each patient, and this may be a barrier to successful implementation. An alternative for clinicians could be to recommend nonpharmacological interventions to patients with LBP (e.g., superficial heat, acupuncture, and spinal manipulative therapy or massage) [4]; such interventions may not have larger effects than pharmacological options, but they are associated with a lower risk of adverse events. Although method reproducibility and inferential reproducibility have now been addressed for the PACE trial, results reproducibility (also called replication) has not [8,10,11]. However, a recent study by Friedman et al compared the combination of the NSAID ibuprofen and paracetamol with the combination of ibuprofen and placebo for patients with acute LBP reporting to two American emergency departments [36]. Although the research question, setting, patient population, and follow-up time were different from the PACE trial, the conclusions regarding the efficacy of paracetamol were the same as those made in the original PACE study. Rather than reproducing the results of PACE in a new RCT, the authors would recommend conducting a meta-analysis of this new study and the PACE trial to more conclusively establish the efficacy of paracetamol for acute LBP.

5. Conclusions

This inferential reproduction analysis indicates that the treatment of patients with acute LBP with paracetamol (taken regularly or as-needed) has no effect on core outcomes of LBP when compared with placebo and thus confirms the original results of the PACE trial [8]. This means the original conclusions of the PACE trial are inferentially reproducible, even when using a different approach to the statistical analysis.

CRediT authorship contribution statement

Marco Schreijenberg: Software, Formal analysis, Writing - original draft. **Alessandro Chiarotto:** Software, Writing - review & editing. **Katya A.L. Mauff:** Software, Formal analysis, Writing - review & editing. **Chung-Wei Christine Lin:** Conceptualization, Investigation, Formal analysis, Writing - original draft. **Christopher G. Maher:** Conceptualization, Investigation, Formal analysis, Writing - original draft. **Bart W. Koes:** Software, Writing - review & editing.

Acknowledgments

The authors thank the participants of this research, the clinicians involved in participant recruitment, and the research support staff involved in site monitoring, data collection, data management, and data analysis.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.01.010.

References

[1] Lin CC, Li Q, Williams CM, Maher CG, Day RO, Hancock MJ, et al. The economic burden of guideline-recommended first line care for acute low back pain. Eur Spine J 2018;27(1):109−16.

[2] Bernstein IA, Malik Q, Carville S, Ward S. Low back pain and sciatica: summary of NICE guidance. BMJ 2017;356:i6748.

[3] Chenot JF, Greitemann B, Kladny B, Petzke F, Pfingsten M, Schorr SG. Non-specific low back pain. Dtsch Arztebl Int 2017; 114(51−52):883−90.

[4] Qaseem A, Wilt TJ, McLean RM, Forciea MA. Clinical Guidelines Committee of the American College of P. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of physicians. Ann Intern Med 2017;166:514−30.

[5] Van Wambeke P, Desomer A, Ailliet L, ABerquin A, Demoulin C, Depreitere B, et al. Low back pain and radicular pain: assessment and management. Brussels: Belgian Health Care Knowledge Centre; 2017.

[6] Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CC, Chenot JF, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2018;27(11):2791−803.

[7] Saragiotto BT, Machado GC, Ferreira ML, Pinheiro MB, Abdel Shaheed C, Maher CG. Paracetamol for low back pain. Cochrane Database Syst Rev 2016CD012230.

[8] Williams CM, Maher CG, Latimer J, McLachlan AJ, Hancock MJ, Day RO, et al. Efficacy of paracetamol for acute low-back pain: a double-blind, randomised controlled trial. Lancet 2014;384:1586−96.

[9] Schreijenberg M, Koes BW, Lin CC. Guideline recommendations on the pharmacological management of non-specific low back pain in primary care - is there a need to change? Expert Rev Clin Pharmacol 2019;12:145−57.

[10] Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? Sci Transl Med 2016;8(341):341ps12.

[11] Plesser HE. Reproducibility vs. Replicability: a brief history of a confused terminology. Front Neuroinform 2017;11:76.

[12] Lee H, Lamb SE, Bagg MK, Toomey E, Cashin AG, Moseley GL. Reproducible and replicable pain research: a critical review. Pain 2018;159(9):1683−9.

[13] Patil P, Peng RD, Leek JT. A statistical definition for reproducibility and replicability. BioRxiV 2017.

[14] Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Intern Med 2011;171:1675−6.

[15] Niven DJ, McCormick TJ, Straus SE, Hemmelgarn BR, Jeffs L, Barnes TRM, et al. Reproducibility of clinical research in critical care: a scoping review. BMC Med 2018;16(1):26.

[16] Schreijenberg M, Luijsterburg PA, Van Trier YD, Rizopoulos D, Koopmanschap MA, Voogt L, et al. Efficacy of paracetamol, diclofenac and advice for acute low back pain in general practice: design of a randomized controlled trial (PACE Plus). BMC Musculoskelet Disord 2017;18(1):56.

[17] Schreijenberg M, Luijsterburg PAJ, Van Trier YDM, Rizopoulos D, Koopmanschap MA, Voogt L, et al. Discontinuation of the PACE Plus trial: problems in patient recruitment in general practice. BMC Musculoskelet Disord 2018;19(1):146.

[18] Froud R, Patterson S, Eldridge S, Seale C, Pincus T, Rajendran D, et al. A systematic review and meta-synthesis of the impact of low back pain on people's lives. BMC Musculoskelet Disord 2014;15:50.

[19] Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J 2015;24(6):1127−42.

[20] Williams CM, Latimer J, Maher CG, McLachlan AJ, Cooper CW, Hancock MJ, et al. PACE–the first placebo controlled trial of paracetamol for acute low back pain: design of a randomised controlled trial. BMC Musculoskelet Disord 2010;11:169.

[21] Williams CM, Maher CG, Latimer J, McLachlan AJ, Hancock MJ, Day RO, et al. PACE–the first placebo controlled trial of paracetamol for acute low back pain: statistical analysis plan. Trials 2013;14:248.

[22] Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 2018;159(3):481−95.

[23] Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8(2):141−4.

[24] Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42:851−9.

[25] Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193−213.

[26] R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.

[27] Goulet J, Buta E, Carroll C, Brandt C. Statistical methods for the analysis of NRS pain data. J Pain 2015;16(4):S7.

[28] Goulet JL, Buta E, Bathulapalli H, Gueorguieva R, Brandt CA. Statistical models for the analysis of zero-inflated pain intensity numeric rating scale data. J Pain 2017;18:340−8.

[29] Rizopoulos D. GLMMadaptive: generalized linear mixed models using adaptive Gaussian quadrature. R package version 0.5-1. 2019. Available at https://CRAN.R-project.org/package=GLMMadaptive. Accessed August 13, 2019.

[30] Ferreira ML, Ferreira PH, Herbert RD, Latimer J. People with low back pain typically need to feel 'much better' to consider intervention worthwhile: an observational study. Aust J Physiother 2009;55(2): 123−7.

[31] Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Grotle M, et al. The smallest worthwhile effect of nonsteroidal anti-inflammatory drugs and physiotherapy for chronic low back pain: a benefit-harm trade-off study. J Clin Epidemiol 2013;66: 1397−404.

[32] Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1−48.

[33] Therneau T. A package for survival analysis in S. version 2.38. 2015. Available at https://CRAN.R-project.org/package=survival. Accessed August 13, 2019.

[34] Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011;2(3):109−12.

[35] Schreijenberg M, Lin CC, McLachlan AJ, Williams CM, Kamper SJ, Koes BW, et al. Paracetamol is ineffective for acute low back pain even for patients who comply with treatment: complier average causal effect analysis of a randomized controlled trial. Pain 2019; 160(12):2848−54.

[36] Friedman BW, Irizarry E, Chertoff A, Feliciano C, Solorzano C, Zias E, et al. Ibuprofen + acetaminophen versus ibuprofen alone for acute low back pain. An ED-based randomized study. Acad Emerg Med 2019.