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

Marco Schreijenberg, Chung-Wei Christine Lin, Andrew J. Mclachlan, Christopher M. Williams, Steven J. Kamper, Bart W. Koes, Christopher G. Maher, Laurent Billot.

PAIN, 2019

ABSTRACT

Introduction

In 2014, the PACE trial demonstrated that paracetamol had no effect compared to placebo in acute low back pain (LBP). However, non-compliance was a potential limitation of this trial. The aim of this study was to investigate the efficacy of paracetamol in acute low back pain among compliers.

Methods

Using individual participant data from the PACE trial (ACTN12609000966291), Complier Average Causal Effects (CACE), Intention-to-treat (ITT) and Per Protocol (PP) estimates were calculated for pain intensity (primary), and disability, global rating of symptom change and function (all secondary) after two weeks of follow-up. Compliance was defined as intake of an average of at least four of the prescribed six tablets of regular paracetamol per day (2660 milligrams in total) during the first two weeks after enrolment. Exploratory analyses using alternative time points and definitions of compliance were conducted.

Results

Mean between-group differences in pain intensity on a 0-10 scale using the primary time point and definition of compliance were not clinically relevant (propensity weighted CACE 0.07 (-0.37, 0.50) p = 0.76; joint modelling CACE 0.23 (-0.16, 0.62) p = 0.24; ITT 0.11 (-0.20, 0.42) p = 0.49; PP 0.29 (-0.07, 0.65) p = 0.12); results for secondary outcomes and for exploratory analyses were similar.

Conclusions

Paracetamol is ineffective for acute low back pain even for patients who comply with treatment. This reinforces the notion that management of acute low back pain should focus on providing patients advice and reassurance without the addition of paracetamol.

INTRODUCTION

The Paracetamol for Acute Low Back Pain (PACE) trial was the first placebo-controlled randomized controlled trial (RCT) investigating the efficacy of paracetamol (acetaminophen) for acute low back pain (LBP) (1-3). In this RCT, 1652 people seeking care for LBP were randomized to take paracetamol regularly, paracetamol ***as needed for pain,*** or placebo using a blinded double-dummy design. The unexpected result that paracetamol had no effect compared to placebo on pain intensity, time until recovery, disability and function in acute LBP received worldwide attention in the medical literature and the lay-press. Nonadherence to study medication was identified as a potential limitation in the original publication of the PACE results as well as in a number of commentaries discussing the impact of the trial (1, 4-6); in a descriptive analysis of nonadherence in PACE, 70% of patients were found to be non-adherent over the four-week treatment period, and overall adherence to guideline-recommended care for acute LBP was described as 'poor' (5). In RCTs, noncompliance has always been an issue and may even influence their results (7). However, the question as to whether there is benefit of an intervention in participants who adequately adhere to treatment is difficult to answer using conventional techniques used in the analysis of RCTs (i.e. intention to treat analysis and per protocol analysis).

Complier average causal effects (CACE) analysis involves comparing participants who were randomized to the intervention and complied, to participants from the control group who would have complied to the intervention had they been randomized to the intervention (so called 'would be compliers'). As participants in the control group are never offered the active treatment in reality, there is no observed data in the control group for adherence to active treatment. CACE analysis is therefore essentially a missing data problem. CACE analyses have been used to assess the efficacy among compliers of intervention programs in substance abuse, behavioral interventions and a multifactorial intervention in physiotherapy (8-16). In the field of LBP, CACE analysis has been used to assess the influence of non-compliance on effectiveness of a cognitive behavioral intervention (17).

This analysis aims to investigate the efficacy of paracetamol in acute LBP among participants who complied with regular paracetamol treatment in the PACE trial using a CACE analysis, to address the uncertainty that compliance may have influenced drug efficacy (14, 15). Additionally, we conducted intention to treat analysis and per protocol analysis to compare to the CACE analysis.

METHODS

Ethics

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.

Participants and procedures

The PACE trial was a multicenter, double-dummy, randomized, placebo-controlled trial that was conducted from November 2009 until March 2013. The study protocol, analysis plan and main outcomes have been published (1-3). In summary, 4606 people seeking care for acute non-specific low-back pain or responding to a community advertisement were screened by 235 primary care clinicians across Sydney, Australia. The trial included 1652 participants with a new episode of moderate, or severe-intensity low back pain with or without leg pain. Participants were randomly allocated (in a 1:1:1 ratio) to receive 2 tablets of 665 mg modified-release paracetamol tablets 3 times a day regularly (n = 550), or 2 tablets of 500 mg immediate-release paracetamol tablets up to 4 times a day ***as-needed*** for pain (n = 549), or placebo (n = 553). Participants, clinicians and researchers were blinded to allocation of treatment during the trial. Participants were instructed to use study medication until they had experienced 7 consecutive days with pain scores of 0 or 1 out of 10 (measured on a numerical pain rating scale (NRS)), or for a maximum of four weeks, whichever occurred first. Participants were asked to return to their clinician for review after 1 week, at which time the use of study medication was reviewed. Rescue medication (naproxen 250 mg) was available for participants with continuing ongoing pain as required.

Participants recorded pain scores and number of tablets taken in a daily pain and drug diary until recovery or for a maximum of four weeks. Follow-up data was collected at 1, 2, 4 and 12 weeks after randomization. Data were either entered directly by the participant into an online database or recorded by participants in a booklet and transcribed to a case report form during a telephone interview with research staff. Returned tablets were counted by research staff to confirm self-reported compliance. In this CACE analysis, data from the ***as-needed*** treatment group were not used because the 'need' to take medication would have been different for each individual participant, preventing the use of one universal definition of compliance in this treatment group.

Outcome measures

For this CACE analysis, pain intensity measured on a NRS from 0 (no pain) to 10 (worst possible pain) was the primary outcome; analyses were also performed for disability (Ro-

land Morris Disability Questionnaire, scored from 0 (no disability) to 24 (high disability)), Global rating of symptom change (scored from-5 (vastly worse) to +5 (completely recovered)) and function (Patient Specific Function Scale, with the average of 3 items scored from 0 (unable to perform) to 10 (able to perform at preinjury level), these outcomes represent two of the three core outcome domains for non-specific LBP (18). Although measurements were conducted in the PACE trial for the third core outcome domain (health-related quality of life), this outcome was omitted from the CACE analysis because of missing data (the Short Form 12 (SF12)), which we expected would compromise the CACE estimation. Time until recovery, the primary outcome of the original PACE analysis, was omitted as methods for survival CACE analysis have not yet been developed.

Definitions of compliance to the study intervention and time points

Compliance was defined as taking an average of at least 4 tablets per day (approximately 66% of the prescribed dosage or 2660 mg per day) of modified-release paracetamol until recovery or for a maximum of 2 weeks for the primary outcome of the CACE analysis (pain intensity at 2 weeks of follow-up).

Two alternative cut-off points for compliance were defined ***a priori*** to assess whether the treatment effect differed according to the level of compliance: taking an average of 5 tablets per day (83% of the prescribed dosage or 3325 mg per day) and taking 6 tablets per day (100% of the prescribed dosage or 3990 mg per day). The two-week questionnaire was chosen as the primary time point as this was closest to the median recovery time (1); exploratory analyses were performed at 1 week and 4 weeks follow-up for pain intensity only. For the exploratory analysis of pain intensity at 4 weeks, the definition of compliance was expanded to 'until recovery or for a maximum of 4 weeks'.

Statistical analysis

Using individual participant data from the PACE trial, baseline participant and back pain episode characteristics were compared between observed compliers and observed non-compliers in the regular paracetamol treatment group, using standardized differences (St.Diffs). For binary variables, the St.Diff was calculated as the difference in proportions divided by the standard deviation i.e. (p1 – p2) / sqrt ([p1 (1 – p1) + p2 (1 – p2)] / 2). For categorical variables with more than 2 levels, we used a method proposed by Yang and Dalton based on a multivariate Mahalanobis distance method which generalizes the St.Diff metric (19). St.Diffs larger than 0.1 were considered to be relevant and were reported in the results section.

We calculated ITT, CACE and Per Protocol (PP) estimates for the 4 outcomes of interest (pain intensity, disability, global rating of symptom change, and function). ITT analyses were performed consistent with the original analysis of the PACE trial, comparing outcomes between all participants randomized to the regular paracetamol group and

all patients randomized to the placebo group using linear mixed models adjusted for all baseline characteristics (1, 3). Based on our definition of compliance, we created a dichotomous variable indicating observed compliance status. We used this dichotomous variable for the PP analysis, where we compared outcomes of observed compliers from the regular paracetamol group to outcomes of observed compliers in the placebo group using linear mixed models adjusted for all baseline characteristics. Outcomes of the PP analysis are not included in the main results of this article, but are added to the supplementary information. The reason for this is that we were interested in comparing results of the CACE analysis to results of a PP analysis, which may provide biased estimates of efficacy for compliers, as the reasons for noncompliance could be different for the regular paracetamol group than for the placebo group. For example, noncompliance in the regular paracetamol group could be related to side effects despite efficacy, whereas noncompliance in the placebo group may be due to lack of efficacy (20). In the Supplementary Information, the difference between PP and CACE analyses is discussed in more detail.

As the underlying assumptions for CACE analysis are untestable, we obtained CACE estimates using both a propensity weighted estimation approach and a joint modeling estimation approach, which serve as each other's sensitivity analysis (15). More information about the underlying assumptions for these CACE estimation techniques can be found in the Supplementary Information. For the propensity weighted CACE estimation, compliance to regular paracetamol was predicted on baseline covariates using logistic regression with a dichotomous variable indicating the observed compliance status. The prediction model was developed using only data from the regular paracetamol group. This model was then used to calculate the likelihood of compliance (propensity score) in the placebo group. To prevent missing propensity scores due to missing baseline data, missing baseline variables were imputed once using fully conditional specification (i.e. imputation on a variable-by-variable basis in an iterative fashion, with an imputation model specified for each incomplete baseline variable (21)). The imputed dataset was used to predict the propensity score. Once derived, the propensity scores were added back to the original non-imputed baseline data set and each participant was weighted as follows: in the regular paracetamol treatment arm, compliers received a weight of 1 and non-compliers a weight of 0; in the placebo treatment arm, the weight was calculated as the odds of the propensity score p (odds=p/(1-p)). We investigated if any residual imbalances existed after weighting by calculating St.Diffs between baseline variables between compliers in the regular paracetamol group and weighted placebo group participants (see Supplementary Information). Finally, we performed an analysis comparing compliers in the regular paracetamol group to odds-weighted patients in the placebo group. Propensity weighted CACE analyses were adjusted for all baseline characteristics in order to correct for residual imbalances. To assess a potential "dose-response" ef-

fect we performed a pre-specified subgroup analysis according to quintiles of likelihood of compliance (using the propensity scores created for the propensity weighted CACE analysis). For this subgroup analysis, the primary cut-off point for compliance (taking an average of at least 4 tablets of modified-release paracetamol per day) and primary time point (two weeks of follow-up) were used); for each quintile group, a mean difference and corresponding confidence interval was calculated.

For the CACE analysis using joint modeling, 2 models were simultaneously estimated: a model for compliance and a model for the outcome (pain intensity). Estimates were adjusted for all baseline characteristics. This estimation approach resulted in a comparison between observed compliers in the regular paracetamol group to inferred compliers (would-be-compliers) in the placebo group.

Results of all the analyses (ITT, CACE propensity and CACE joint modeling and PP) are presented as mean differences between paracetamol and placebo groups with 95% confidence intervals and corresponding p-values. ITT, PP and propensity weighted CACE analyses were performed in SAS version 9.4 (SAS Institute, Inc., Cary, NC), joint modeling CACE estimation was performed in Mplus version 7 (22).

RESULTS

Characteristics of compliers to regular paracetamol

The baseline characteristics of participants in the regular paracetamol group are presented in Table 1; participants were split into compliers and non-compliers based on our main definition of compliance (an average of at least 4 tablets of 665 mg regular paracetamol per day during the first 2 weeks). Table 1 also shows St.Diffs between observed compliers and non-compliers. At the primary time point of the CACE analysis (2 weeks), 394 out of 550 participants in the paracetamol group (72%) were classified as compliers.

When comparing compliers and non-compliers, compliers tended to be somewhat older (44.9 vs 42.4 years, St.Diff 0.17); were more likely to be male (54% vs 46%, St.Diff 0.15); were more likely to have private health insurance (52% vs 46%, St.Diff 0.12); had a different distribution of household income (St.Diff 0.23); were less likely to have pain extending beyond the knee (17% vs 26%, St.Diff 0.22); had a longer period of reduced usual activity (4.1 vs 3.2 days, St.Diff 0.13); scored higher for feelings of depression (3.4 vs 2.8, St.Diff 0.18); reported a higher perceived risk of persistent pain (4.8 vs 4.1 out of 10, St.Diff0.22); more often reported poor sleep quality (51% vs 46%, St.Diff 0.10); scored lower on function (3.4 vs 3.7, St.Diff 0.15) and scored lower for physical quality of life (42.4 vs 43.3, St.Diff 0.11).

Table 1: Baseline characteristics for observed compliers and non-compliers in the regular paracetamol group, including standardized mean differences between observed compliers and observed non-compliers.

Patient characteristics	Regular Paracetamol (N = 550)		Standardized differences
	Observed compliers (N = 394)	Observed non-compliers (N = 142)	
Age (years)	44.9 (14.9) N = 394	42.4 (14.5) N = 142	0.171*
Women	182/393 (46%)	75/140 (54%)	0.146*
Private health insurance	203/394 (52%)	65/142 (46%)	0.115*
Currently employed	305/394 (77%)	107/142 (75%)	0.048
Household income per week (per year)			0.342*
Negative or no income	13/384 (3%)	6/142 (4%)	
AUD 1-649 (1-33799)	89/384 (23%)	42/142 (30%)	
AUD 650-1699 (33800-88399)	174/384 (45%)	59/142 (42%)	
AUD 1700-3999 (88400-207999)	86/384 (22.4%)	32/142 (23%)	
≥AUD 4000 (≥208000)	22/384 (6%)	3/142 (2%)	
Use of drugs for another disorder	148/394 (38%)	49/142 (35%)	0.064
LBP Episode characteristics			
Days since onset of pain	10.2 (10.3) N = 394	9.8 (9.6) N = 142	0.037
Number of previous episodes	6.4 (12.8) N = 392	6.5 (16.4) N = 141	0.009
Presence of pain extending beyond the knee	68/392 (17%)	37/141 (26%)	0.217*
Number of days reduced usual activity	4.1 (7.0) N = 393	3.2 (4.9) N = 141	0.134*
Disability (RMDQ)	12.7 (5.5) N = 390	12.9 (5.9) N = 139	0.027
Feelings of depression in last week	3.4 (2.9) N = 392	2.8 (3.0) N = 141	0.175*
Perceived risk of persistent pain	4.8 (2.7) N = 392	4.1 (2.9) N = 142	0.224*
Back pain episode compensable	20/392 (5%)	10/140 (7%)	0.085
Pain intensity (NRS)	6.3 (1.9) N = 394	6.2 (2.0) N = 142	0.039
Global rating of symptom change	0.0 (2.1) N = 393	0.1 (2.0) N = 141	0.054
Poor sleep quality	200/393 (51%)	65/142 (46%)	0.103*
Function (Nominated Activity)	3.4 (1.7) N = 392	3.7 (1.9) N = 141	0.151*
Quality of life – physical (SF-12)	42.4 (9.0) N = 384	43.3 (9.4) N = 140	0.112*
Quality of life – mental (SF-12)	44.3 (7.7) N = 384	43.7 (7.8) N = 140	0.071
Credibility score (CEQ)	19.1 (4.9) N = 390	18.8 (4.8) N = 140	0.064
Expectation score (CEQ)	19.8 (5.4) N = 389	19.4 (5.3) N = 141	0.080

St.Diffs: Standardized Differences; Data are mean (SD) or n/N (%);* under standardized differences indicate St.Diffs > 0.1. AUD: Australian Dollar; LBP: Low Back Pain; NRS: Numerical Rating Scale; RMDQ: Roland Morris Disability Questionnaire; SF-12; 12-item Short Form Survey; CEQ: Credibility/Expectancy Questionnaire.

Estimates of the CACE models

Table 2 presents ITT and CACE estimates for pain intensity, disability, global rating of symptom change, and function in the PACE trial at week 2 with compliance defined as an average intake of at least 4 tablets per day during the first 2 weeks.

For the primary outcome measure, none of the analyses indicated a difference in pain intensity (ITT: mean difference 0.11 (-0.20, 0.42) p = 0.49; joint modeling CACE: mean difference 0.23 (-0.16, 0.62) p = 0.24; propensity weighted CACE: mean difference 0.07 (-0.37, 0.50) p = 0.76). Similar results were obtained for the secondary outcomes disability, global rating of symptom change, and function. Confidence intervals of estimates for pain intensity, global rating of symptom change and function were all between-1 and 1 and therefore exclude clinically meaningful differences; the confidence interval of the estimate of disability exceeded 1 in both the propensity weighted CACE estimation and the joint modelling CACE estimation; however, this difference is still smaller than the minimal clinically important difference (MCID) of 30% change from baseline (in PACE, approximately 4 points) (23).

Table 2: Outcomes of PACE trial (Pain Intensity, Disability, Global Rating of Symptom Change and Function) at week 2 with compliance defined as an average intake of ≥4 tablets per day for regular paracetamol group vs placebo group.

Outcome	ITT	Propensity weighted CACE	Joint Modeling CACE
Pain Intensity (NRS) (scale range 0-10)	0.11 (-0.20, 0.42) p = 0.49	0.068 (-0.37, 0.50) p = 0.76	0.23 (-0.16, 0.62) p = 0.24
Disability (RMDQ) (scale range 0-24)	0.11 (-0.60, 0.82) p = 0.76	0.054 (-0.93, 1.04) p = 0.91	0.37 (-0.55, 1.30) p = 0.43
Global Rating of Symptom Change (scale range-5 to +5)	0.0019 (-0.26, 0.27) p = 0.99	0.059 (-0.33, 0.44) p = 0.76	-0.083 (-0.42, 0.25) p = 0.62
Function (Patient Specific Function Scale) (scale range 0-10)	-0.069 (-0.38, 0.24) p = 0.67	0.0043 (-0.45, 0.45) p = 0.99	-0.28 (-0.67, 0.11) p = 0.16

All values represent mean difference (lower limit of 95% CI, upper limit of 95% CI) p value; mean differences calculated by subtracting placebo group mean from regular paracetamol group mean. All analyses were adjusted for gender and baseline age, private health insurance, employment status, household income, use of drugs for another disorder, days since onset of pain, number of previous episodes, presence of pain extending beyond the knee, number of days reduced usual activity, disability (RMDQ), feelings of depression, perceived risk of persistent pain, back pain episode compensability, pain intensity, global rating of symptom change, sleep quality, function, quality of life (mental and physical components of the 12 item short form survey (SF-12)) and credibility and expectation scores (CEQ). Values rounded to 2 significant figures. Abbreviations: CACE: Complier Average Causal Effect; CEQ: Credibility/Expectancy Questionnaire; ITT: Intention-to-Treat; NRS: Numerical Rating Scale; RMDQ: Roland Morris Disability Questionnaire.

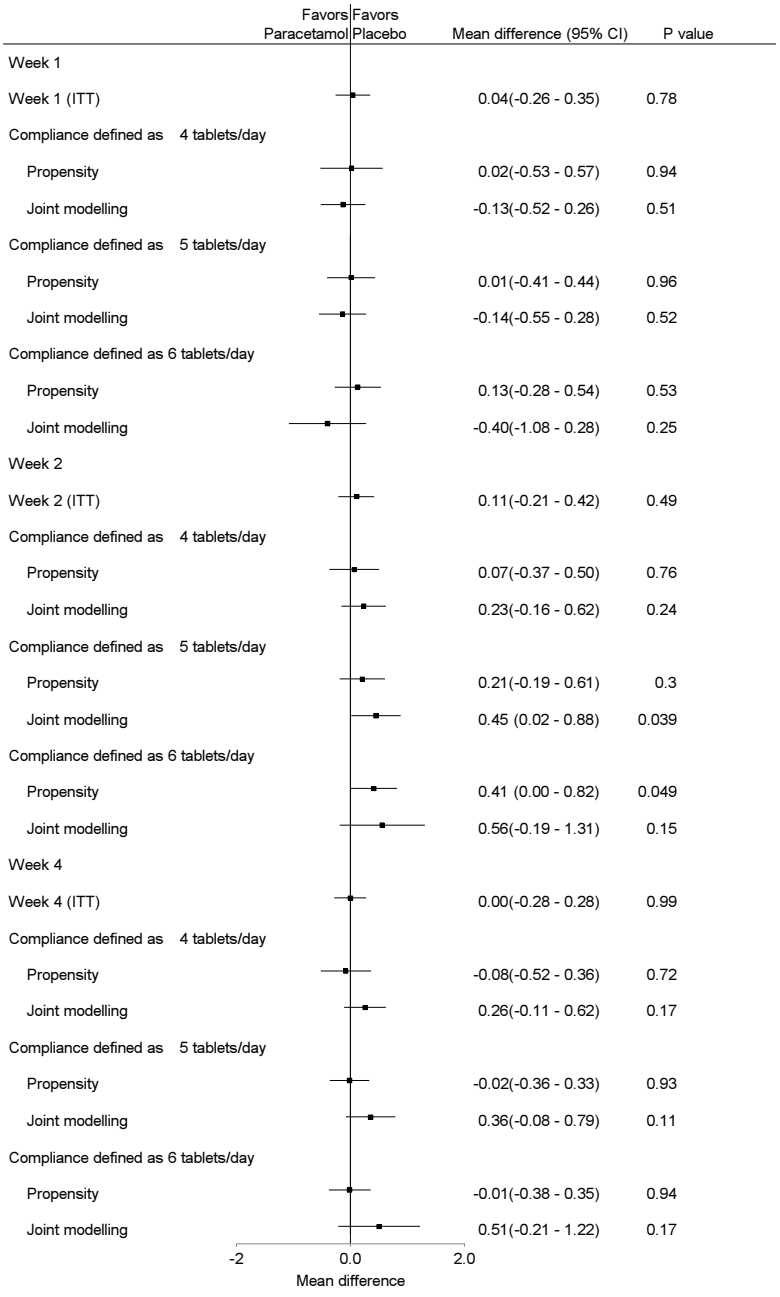


Figure 1: Exploratory ITT and CACE analyses for pain intensity including both primary and alternative cut-off points for compliance (an average of at least 5 tablets per day and 6 tablets per day, calculated over the periods of interest) as well as primary and alternative time points (1 week and 4 weeks). Values rounded to 2 significant figures. ITT: intention to treat, CACE: Complier Average Causal Effect, CI: Confidence Interval.

Exploratory analyses

Figure 1 shows results of the exploratory ITT and CACE analyses using primary and alternative cut-off points for compliance (an average of at least 5 tablets per day, and 6 tablets per day) and primary and alternative time points (1 week and 4 weeks). Mean differences in pain intensity between regular paracetamol and placebo were calculated for 3 definitions of compliance at 3 time points using 3 analysis techniques, yielding a total of 21 estimates.

Minimal differences in pain intensity were only found for 2 of the 21 analyses: the joint modeling CACE estimate after 2 weeks with compliance defined as an average of at least 5 paracetamol tablets per day (mean difference 0.45 (0.02, 0.88), p = 0.039) and for the propensity weighted CACE estimate after 2 weeks with compliance defined as 6 paracetamol tablets per day (mean difference 0.41 (0.00, 0.82) p = 0.049); however, no correction was made for multiple testing. Furthermore, the confidence intervals for these estimates do not include clinically meaningful differences. For all other time points, no differences in pain intensity were found.

Results of the ITT analysis for pain intensity at 2 weeks for quintiles of compliance (defined as an average of at least 4 tablets per day over 2 weeks) are depicted in Figure 2. No difference in pain intensity was found between regular paracetamol and placebo for any of the compliance subgroups. There appears to be no clear dose-response relationship between compliance and effect of paracetamol.

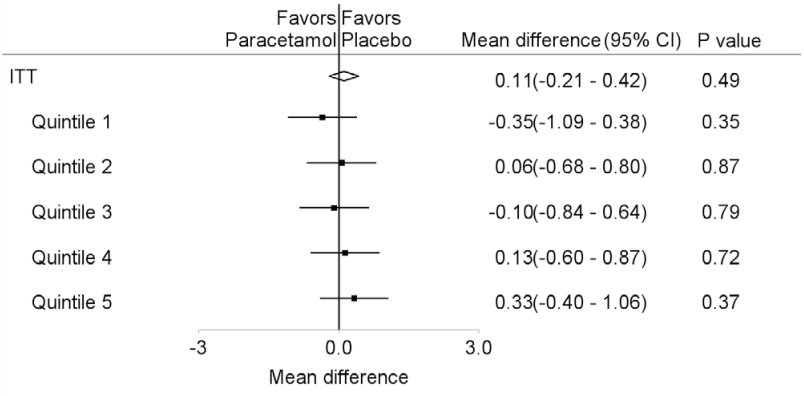


Figure 2: Exploratory ITT analysis for pain intensity at 2 weeks for quintiles of likelihood of compliance (with compliance defined as taking an average of at least 4 tablets of modified-release paracetamol per day during 2 weeks of follow-up). Quintile groups are presented in order of increasing likelihood of being compliant, with Quintile 1 representing the group that was least likely to be compliant and Quintile 5 representing the group that was most likely to be compliant. ITT: intention to treat, CI: Confidence Interval.

DISCUSSION

In this secondary analysis of the PACE trial we found that paracetamol had no clinically-meaningful effect when compared to placebo on pain intensity, disability, global rating of symptom change and function in people with acute LBP who complied with regular paracetamol.

The CACE analysis technique produces robust estimates of efficacy amongst compliers; furthermore, we applied 2 distinct methods to estimate complier average causal effects, which serve as each other's sensitivity analysis (15). The credibility of our findings is supported by the fact that no large differences exist between these 2 estimation techniques (15). Data used in this analysis were collected in a large and well-conducted RCT (1, 24).

The CACE analysis technique has two main weaknesses. First, no universally accepted definition of compliance to paracetamol for low back pain exists. Using our main definition of compliance, 72% of participants in the regular paracetamol group were classified as compliers. We explored stricter definitions of compliance and found results consistent with the primary analysis; however, as the percentage of compliers was lower using these definitions, CACE estimates using these definitions are less robust. Second, CACE estimates were based on patient-reported compliance filled out in paper drug diaries, which may not have perfectly represented actual consumption of tablets. However, counts of returned medicines and results from the brief adherence rating scale were consistent with patient-reported compliance (1).

The findings of this secondary analysis should be placed in context of the original analysis of the PACE trial, which is still the only RCT that has assessed the efficacy of paracetamol for acute LBP and is considered to be the best available evidence (24). As mentioned in the introduction, non-compliance to study medication was considered a potential limitation of the PACE results (1, 4-6). The results of this analysis suggest this is not the case and thus support the conclusion from the original analysis of the PACE trial that paracetamol is ineffective for acute LBP when compared to placebo. It is important to note that CACE analysis is a technique that accounts for a very specific participant group, namely those who comply with treatment. Although this analysis technique may be useful in trials where non-compliance is an issue, results of the ITT analysis remain the most relevant to clinical practice.

After a lack of efficacy of paracetamol for acute LBP was demonstrated by the PACE trial, paracetamol was no longer recommended as first choice analgesic in four out of eight recently published national clinical practice guidelines (25-28). However, other recent guidelines still endorse the prescription of paracetamol for acute LBP (29-32). One possible justification was that paracetamol may be effective in those who comply with the dosing regimen. Our CACE analyses have demonstrated that the efficacy of paracetamol is unlikely to change even in patients with total compliance to the regular

regimen, reinforcing that management of acute low back pain should focus on providing patients advice and reassurance without the addition of paracetamol.

In conclusion, paracetamol is not more effective than placebo for acute LBP in compliers of the treatment regimen. CACE analyses using different cut points showed that paracetamol had no effect on pain intensity and secondary outcomes when compared to placebo for participants that complied to regular paracetamol in the PACE trial. These results support the original findings of the PACE trial.

REFERENCES:

1. 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(9954):1586-96.
2. 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.
3. 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 [Electronic Resource]. 2013;14:248.
4. Koes BW, Enthoven WT. Do patients with acute low-back pain need paracetamol? Lancet. 2014;384(9954):1556-7.
5. Bier JD, Kamper SJ, Verhagen AP, Maher CG, Williams CM. Patient Nonadherence to Guideline-Recommended Care in Acute Low Back Pain. Arch Phys Med Rehabil. 2017.
6. Kress HG, Untersteiner G. Clinical update on benefit versus risks of oral paracetamol alone or with codeine: still a good option? Curr Med Res Opin. 2017;33(2):289-304.
7. Pullar T, Kumar S, Feely M. Compliance in clinical trials. Ann Rheum Dis. 1989;48(10):871-5.
8. Berg JK, Bradshaw CP, Jo B, Ialongo NS. Using Complier Average Causal Effect Estimation to Determine the Impacts of the Good Behavior Game Preventive Intervention on Teacher Implementers. Adm Policy Ment Health. 2017;44(4):558-71.
9. Hesser H, Hedman E, Lindfors P, Andersson E, Ljotsson B. The specific effect of systematic exposure in irritable bowel syndrome: complier average causal effect analysis using growth mixture modeling. Psychol Med. 2017;47(15):2653-62.
10. Connell AM. Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. Am J Drug Alcohol Abuse. 2009;35(4):253-9.
11. Huang S, Cordova D, Estrada Y, Brincks AM, Asfour LS, Prado G. An application of the Complier Average Causal Effect analysis to examine the effects of a family intervention in reducing illicit drug use among high-risk Hispanic adolescents. Fam Process. 2014;53(2):336-47.
12. McConnachie A, Haig C, Sinclair L, Bauld L, Tappin DM. Birth weight differences between those offered financial voucher incentives for verified smoking cessation and control participants enrolled in the Cessation in Pregnancy Incentives Trial (CPIT), employing an intuitive approach and a Complier Average Causal Effects (CACE) analysis. Trials [Electronic Resource]. 2017;18(1):337.
13. Stanger C, Ryan SR, Fu H, Budney AJ. Parent training plus contingency management for substance abusing families: a Complier Average Causal Effects (CACE) analysis. Drug Alcohol Depend. 2011;118(2-3):119-26.
14. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. Journal of the American Statistical Association. 1996;91:444-55.
15. Stuart EA, Jo B. Assessing the sensitivity of methods for estimating principal causal effects. Stat Methods Med Res. 2015;24(6):657-74.
16. Fairhall N, Sherrington C, Cameron ID, Kurrle SE, Lord SR, Lockwood K, et al. A multifactorial intervention for frail older people is more than twice as effective among those who are compliant: complier average causal effect analysis of a randomised trial. J Physiother. 2017;63(1):40-4.

17. Knox CR, Lall R, Hansen Z, Lamb SE. Treatment compliance and effectiveness of a cognitive behavioural intervention for low back pain: a complier average causal effect approach to the BeST data set. BMC Musculoskelet Disord. 2014;15:17.

18. 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.

19. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum2012.

20. Jo B. Statistical power in randomized intervention studies with noncompliance. Psychol Methods. 2002;7(2):178-93.

21. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully Conditional Specification in Multivariate Imputation. Journal of Statistical Computation and Simulation. 2006;76(12):1049-64.

22. Muthén L.K., Muthén B.O. Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén; 1998-2012.

23. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. Journal of Clinical Epidemiology. 2006;59(1):45-52.

24. Saragiotto BT, Machado GC, Ferreira ML, Pinheiro MB, Abdel Shaheed C, Maher CG. Paracetamol for low back pain. Cochrane Database Syst Rev. 2016(6):CD012230.

25. Van Wambeke P, Desomer A, Ailliet L, ABerquin A, Demoulin C, Depreitere B, et al. Low Back Pain And Radicular Pain: Assessment And Management. Belgian Health Care Knowledge Centre. 2017.

26. 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.

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

28. 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. Annals of Internal Medicine. 2017;166(7):514-30.

29. NSW Agency for Clinical Innovation. Management of people with acute low back pain: model of care. Chatswood; NSW Health. 2016:39 p.

30. Institute of Health Economics Alberta Canada. Evidence-Informed Primary Care Management of Low Back Pain. wwwiheca or wwwtopalbertadoctorsorg. 2017.

31. Stochkendahl MJ, Kjaer P, Hartvigsen J, Kongsted A, Aaboe J, Andersen M, et al. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. Eur Spine J. 2018;27(1):60-75.

32. Bons S.C.S., Borg M.A.J.P., Van den Donk M., Koes B.W., Kuijpers T., Ostelo R.W.J.G., et al. The revised Dutch College of General Practitioners (NHG) practice guideline on 'Non-specific Low Back Pain'(in Dutch). Huisarts Wet. 2017;60(2).

APPENDIX 1: SUPPLEMENTARY INFORMATION

Background to Complier Average Causal Effect (CACE) analysis

Conventional Intention-to-Treat (ITT) analysis produces the overall mean effect for all participants randomized, regardless of compliance status. While an ITT analysis provides an unbiased estimate of the effect of treatment allocation, it does not estimate the efficacy of treatment in compliers. Common analysis strategies used to estimate treatment effect in compliers are Per Protocol (PP) analysis, where effects for compliers in the intervention group are compared to effects for participants in the control group, and As-treated (AT) analysis, where effects for those who received the intervention are compared to effects for those who did not receive the intervention. However, the use of PP and AT analyses to account for compliance may lead to biased results (1). The reason for this is that likelihood of compliance to placebo is fundamentally different from likelihood of compliance to an active treatment; this could actually work in two directions: participants receiving an active drug may be more likely to comply to treatment if they experience a drug effect that participants receiving placebo cannot experience. On the other hand, participants receiving an active drug may be less likely to comply to treatment if they experience adverse effects that participants receiving placebo don't experience. Compliers to an active treatment in one group and compliers to placebo in another group may thus be incomparable, while the overall treatment groups would of course be comparable due to randomization.

To obtain unbiased estimates for complying participants in the intervention group of a trial, the effect estimates for compliers in the intervention group could be compared to the effects for those in the control group, who would have complied to the treatment had they been randomized to the intervention group (so-called 'would-be-compliers'). This comparison can be made using an average causal effects (CACE) analysis (2, 3). Of course, the active treatment was never offered to the control group participants in real life; therefore we have no observed compliance data in this group. Therefore, the CACE analysis is essentially a missing data problem in the control group; once 'would-be-compliers' in this group have been identified, a normal ITT analysis can be performed.

Underlying assumptions in CACE analysis (3) and their translation to the PACE trial

1. **Ignorable Treatment Assignment:** Treatment assignment is independent of the potential outcomes, conditional on the observed baseline covariates (3).
   PACE: This assumption is automatically satisfied in randomized experiments.
2. **Stable Unit Treatment Value Assumption (SUTVA):**

   a. **The potential outcomes for each person are unaffected by the treatment assignment of other individuals:** this means there is no interference between patients in different groups (3).

      PACE: Although interference between individuals in different treatment groups could not be ruled out in PACE, it is unlikely that this has played a significant role in the outcomes as all participants were blinded to the treatment they observed and everyone received medication through the double-dummy design; this means that unblinding because of dosage scheme is highly unlikely. SUTVA is therefore considered to be fulfilled.

   b. **There is only one version of each treatment:** the treatments given to each individual within each treatment condition do not vary across individuals (3).

      PACE: Both active and control groups had one fixed treatment protocol in PACE. SUTVA is therefore considered to be fulfilled.

3. **Monotonicity:** This assumption states there are no Defiers (i.e. participants who do the exact opposite of the instructions given in the trial; they take the intervention if they're in the control group and they don't take the intervention when they are in the active group) (3).

   PACE: Because of the blinded double-dummy design used in PACE, patients in all groups received the same instructions and were blinded for the intervention they received (this means in this case, active paracetamol group and placebo group do not get contradicting instructions). It seems unlikely that there were Defiers in PACE.

4. **Exclusion Restriction:** Treatment assignment does not affect the outcome if it does not affect the treatment actually received (no direct effect of treatment assignment on outcome). This means there is no effect of the treatment assignment for always takers or never takers; therefore Always-Taker Average Causal Effect (AACE) = 0 and Never-Taker Average Causal Effect (NACE) = 0 (3).

   PACE: Translation of the assumption to terms used in the PACE trial: patients who never take or always take paracetamol for low back pain won't experience an effect of being randomized in the PACE trial. This assumption is may not be fulfilled if trial randomization (i.e. telling people to take the given medication) has an effect on outcome through other behavior than taking pills; for example if always-takers or never-takers start exercising more for their back pain because of the attention given to the back pain during the trial.

5. **Principal Ignorability:** Potential outcomes are the same across compliance strata, conditional on covariates. In other words, principal stratum membership is independent of the potential outcomes given the observed covariates. This assumption implies that we can identify principal stratum membership using only the observed covariates. This is what enables us to find the "likely compliers" in the control group, using the model of compliance behavior as a function of covariates fit among treat-

ment group members; i.e., it implies that the outcomes of the control group members identified as "likely compliers" actually reflect well what the potential outcomes under control would have been had the treatment group compliers been in the control group instead (3).

PACE: This assumption is very likely to hold as the intervention and control are so similar in PACE due to the blinded double dummy design.

6. **Missing Values Assumptions:**
   a. **Missing At Random (MAR) Assumption:** Non-response is associated with non-compliance only among individuals with observed compliancy information. The probability of the outcome being recorded is not associated with the outcome conditional on treatment assignment, observed treatment receipt status and pre-treatment covariates. Under this assumption, missingness is not attributable to unobserved data, including unobserved compliance status (3).

      PACE: This assumption is unlikely to hold in PACE, as nonresponse in the placebo group is very likely to be associated to the trial outcomes (i.e. participants who have recovered are less likely to continue filling out the trial questionnaires). However, the vast majority of patients provided data for the primary outcome in PACE (97%). For this reason, missing data were not considered influential in the CACE analyses presented in this article.

   b. **Response Exclusion Restriction (RER) Assumption:** For never-takers, the probability of outcomes being recorded is not affected by treatment assignment status. This assumption will be violated if response probability is affected by treatment assignment. This means it is violated if never-takers provide outcome data more when assigned to the intervention condition then when assigned to the control condition (3).

      PACE: This assumption may hypothetically be violated in 2 ways:
      - Poorly complying participants may have felt some benefit from active paracetamol and might have felt more obliged to provide outcome information when assigned to the active regular paracetamol group than when assigned to the placebo group.
      - Poorly complying participants might have been demoralized when assigned to the intervention condition, by failing to comply with the intervention. This might not have happened if they had been assigned to the placebo group.

   c. **Stable Complier Response (SCR) Assumption:** For compliers, the probability of outcomes being recorded is not affected by treatment assignment status. This assumption will be violated if response probability is affected by treatment assignment. This means it is violated if compliers provide outcome data more when assigned to the intervention condition then when assigned to the control condition (3).

PACE: This assumption may hypothetically be violated in 2 ways:

- Compliers may have felt some benefit from active paracetamol and might have felt more obliged to provide outcome information when assigned to the active regular paracetamol group than when assigned to the placebo group.
- Compliers might have been demoralized when assigned to the placebo condition because they did not feel any effect despite taking the prescribed medication. This might not have happened if they had been assigned to the intervention group.

Supplementary Results

Baseline characteristics of the complete regular paracetamol and placebo groups (as shown in the original analysis) are presented with corresponding standardized differences (St.Diffs) in Supplementary Table 1; due to chance, significant differences were found for employment status and household income. Because no significant differences could be found in all other baseline characteristics, we still assume correct randomization.

Estimates of the CACE models compared to Per Protocol analysis

Mean differences and corresponding p-values were very similar between Per Protocol analysis and joint modeling CACE for all outcomes that were assessed (Supplemental Table 2). A reason for this may be that in this trial, inferred compliance behavior to regular paracetamol in the placebo group was similar to observed compliance to placebo in this group. However, it should be noted that Per Protocol analysis should not be routinely used to account for non-compliance, as using this statistical technique may lead to biased effect estimates (3, 4).

Technical considerations about this CACE analysis

Three assumptions are used in CACE analysis (3). The first assumption is ignorable treatment assignment, which states that treatment assignment is independent of the potential outcomes, conditional on the observed baseline covariates. This assumption is automatically satisfied in randomized experiments. The second assumption is stable unit treatment value assumption (SUTVA), which states two things: firstly, the potential outcomes for each person are unaffected by the treatment assignment of other individuals and secondly, there is only one version of each treatment, meaning that treatments given to each individual within each treatment condition do not vary across individuals. It is unlikely that SUTVA has been violated in PACE due to the fixed treatment protocols and double-dummy design, which prevents unblinding and thus interference between participants. The third assumption is monotonicity, which states that there are no defiers (participants who do the exact opposite of the instructions given in the trial. Because all patients received the same instructions and blinding was shown to have been successful, it is unlikely the monotonicity assumption has been violated in PACE (5).

In joint modeling CACE estimation, exclusion restriction is an additional underlying assumption; this assumption states that treatment assignment does not affect the outcome if it does not affect the treatment actually received (i.e., there is no direct effect of treatment assignment of the outcome). In propensity-weighted CACE estimation, exclusion restriction is replaced by principal ignorability; this assumption states that principal stratum membership is independent of the potential outcomes given the observed covariates. In practice, exclusion restriction is unverifiable and can hardly ever be completely dismissed; in PACE, it could very well be that receiving instructions to

take medication may lead to always-takers or never-takers of medication being more active, which in turn could affect the outcomes of their back pain. However, in settings like the PACE trial where there is no effect of the intervention at all, the exclusion restriction assumption is automatically satisfied and joint modeling CACE estimation is likely to perform very well (given no severe deviations from normality of outcome variables); although the assumption is not necessarily violated, principal ignorability propensity-weighted CACE estimation does not necessarily perform well in this scenario (3). For this reason, we expect the joint modeling approach will have resulted in the more reliable CACE estimates of the two methods used.

REFERENCES:

1. Jo B. Statistical power in randomized intervention studies with noncompliance. Psychol Methods. 2002;7(2):178-93.
2. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. Journal of the American Statistical Association. 1996;91:444-55.
3. Stuart EA, Jo B. Assessing the sensitivity of methods for estimating principal causal effects. Stat Methods Med Res. 2015;24(6):657-74.
4. Little RJ, Yau LHY. Statistical Techniques for Analyzing Data From Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model. Psychological Methods. 1998;3(2):147-59.
5. 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(9954):1586-96.

APPENDIX 2: SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1: Baseline characteristics for complete patient groups (ITT).

Patient characteristics	All patients (ITT)		
	Paracetamol (N = 550)	Placebo (N = 547)	Standardized difference
Age (years)	44.1 (14.8) N = 550	45.4 (15.9) N = 546	0.085
Women	263/547 (48%)	245/544 (45%)	0.061
Private health insurance	275/550 (50%)	248/544 (46%)	0.088
Currently employed	424/550 (77%)	389/542 (72%)	0.122*
Household income per week (per year)			0.179*
Negative or no income	19/540 (4%)	22/531 (4%)	
AUD 1-649 (1-33799)	133/540 (25%)	168/531 (32%)	
AUD 650-1699 (33800-88399)	243/540 (45%)	226/531 (43%)	
AUD 1700-3999 (88400-207999)	119/540 (22%)	97/531 (18%)	
≥AUD 4000 (≥208000)	26/540 (5%)	18/531 (3%)	
Use of drugs for another disorder	201/550 (37%)	202/544 (37%)	0.012
Episode characteristics			
Days since onset of pain	10.1 (10.1) N = 550	9.7 (9.8) N = 546	0.039
Number of previous episodes	6.3 (13.7) N = 547	7.2 (16.8) N = 544	0.058
Presence of pain extending beyond the knee	108/547 (20%)	99/544 (18%)	0.039
Number of days reduced usual activity	3.7 (6.3) N = 548	3.4 (5.3) N = 545	0.075
Disability (RMDQ)	12.8 (5.6) N = 543	13.3 (5.5) N = 531	0.081
Feelings of depression in last week	3.2 (2.9) N = 547	3.1 (2.9) N = 546	0.048
Perceived risk of persistent pain	4.5 (2.8) N = 548	4.4 (2.8) N = 545	0.050
Back pain episode compensable	31/546 (6%)	43/546 (8%)	0.088
Pain intensity	6.3 (1.9) N = 550	6.2 (1.8) N = 546	0.054
Global rating of symptom change	0.0 (2.1) N = 548	-0.1 (2.1) N = 546	0.046
Poor sleep quality	273/549 (50%)	272/546 (50%)	0.002
Function (Nominated Activity)	3.5 (1.7) N = 547	3.7 (1.9) N = 545	0.069
Quality of life – physical (SF-12)	42.7 (9.1) N = 537	42.1 (9.2) N = 538	0.065
Quality of life – mental (SF-12)	44.1 (7.7) N = 537	44.4 (7.9) N = 538	0.040
Credibility score (CEQ)	19.0 (4.9) N = 544	19.4 (4.9) N = 540	0.078
Expectation score (CEQ)	19.7 (5.3) N = 544	20.2 (5.1) N = 542	0.093

St.Diffs: Standardized Differences; Data are mean (SD) or n/N (%);* under standardized differences indicate St.Diffs > 0.1. AUD: Australian Dollar; LBP: Low Back Pain; NRS: Numerical Rating Scale; RMDQ: Roland Morris Disability Questionnaire; SF-12; 12-item Short Form Survey; CEQ: Credibility/Expectancy Questionnaire.

Supplementary Table 2: Outcomes of PACE trial (Pain Intensity, Disability, Global Rating of Symptom Change and Function) at week 2 with compliance defined as an average intake of ≥4 tablets per day for regular paracetamol group vs placebo group.

Outcome	Per Protocol	Joint Modeling CACE
Pain Intensity (NRS) (scale range 0-10)	0.29 (-0.074, 0.65) p = 0.12	0.23 (-0.16, 0.62) p = 0.24
Disability (RMDQ) (scale range 0-24)	0.41 (-0.40, 1.22) p = 0.32	0.37 (-0.55, 1.30) p = 0.43
Global Rating of Symptom Change (scale range-5 to +5)	-0.13 (-0.44, 0.18) p = 0.41	-0.083 (-0.42, 0.25) p = 0.62
Function (Patient Specific Function Scale) (scale range 0-10)	-0.28 (-0.65, 0.089) p = 0.14	-0.28 (-0.67, 0.11) p = 0.16

All values represent mean difference (lower limit of 95% CI, upper limit of 95% CI) p value; mean differences calculated by subtracting placebo group mean from regular paracetamol group mean. Values rounded to 2 significant figures. NRS: Numerical Rating Scale; RMDQ: Roland Morris Disability Questionnaire.

Supplementary Table 3: Assessment for residual imbalances between the regular paracetamol group and weighted placebo group in the distribution of categorical variables after the weighting procedure in propensity-weighted CACE analysis at week 2 with compliance defined as an average intake of ≥4 tablets per day.

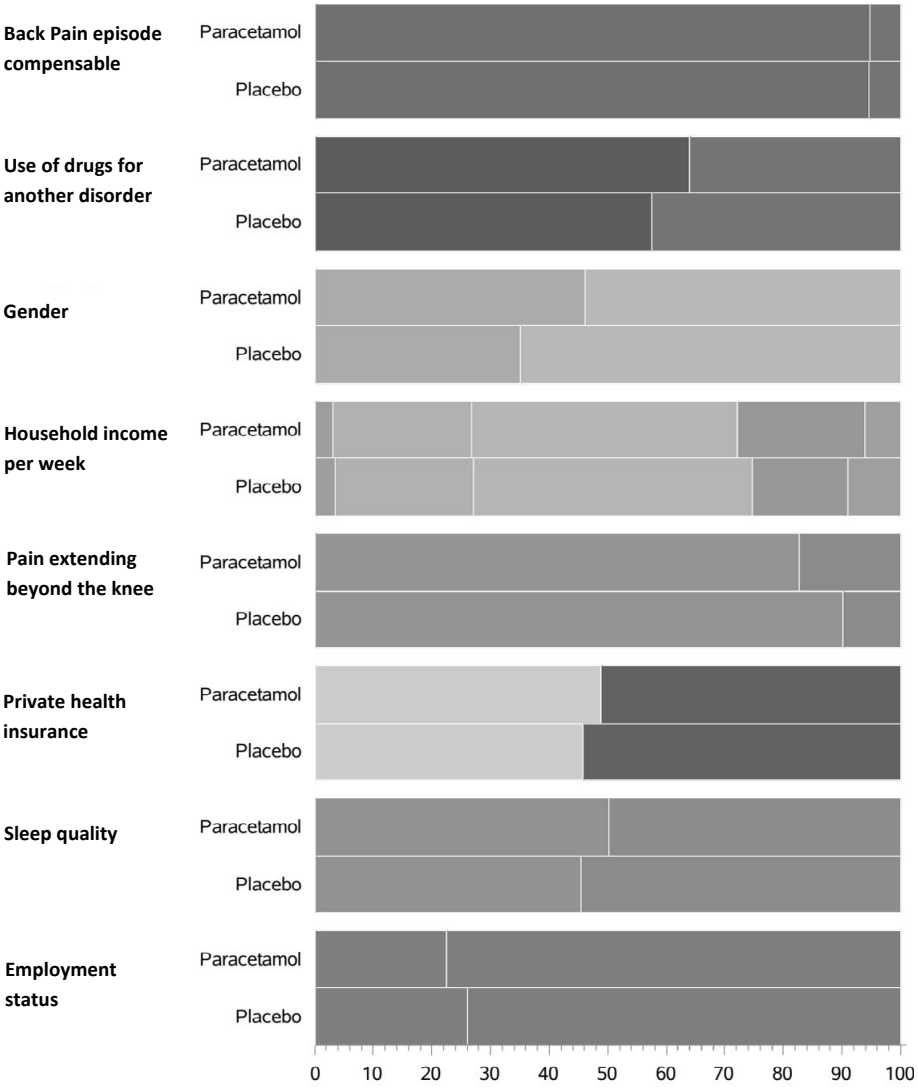
Variable	Placebo (proportion)	Paracetamol (proportion)	Standardized difference
Back Pain episode compensable	0.054	0.053	0.00314
Use of drugs for another disorder	0.42	0.36	0.134*
Women	0.35	0.46	0.226*
Household income per week (per year):	0.035	0.031	0.0228
Negative or no income	0.24	0.24	0.00158
AUD 1-649 (1-33799)	0.48	0.46	0.0440
AUD 650-1699 (33800-88399)	0.16	0.22	0.137*
AUD 1700-3999 (88400-207999)	0.090	0.062	0.106*
≥AUD 4000 (≥208000)	0.098	0.17	0.223*
Private health insurance	0.54	0.51	0.0618
Poor sleep quality	0.55	0.50	0.0962
Currently employed	0.74	0.78	0.0838

St.Diffs: Standardized Differences; * under standardized differences indicate St.Diffs > 0.1; AUD: Australian Dollar.

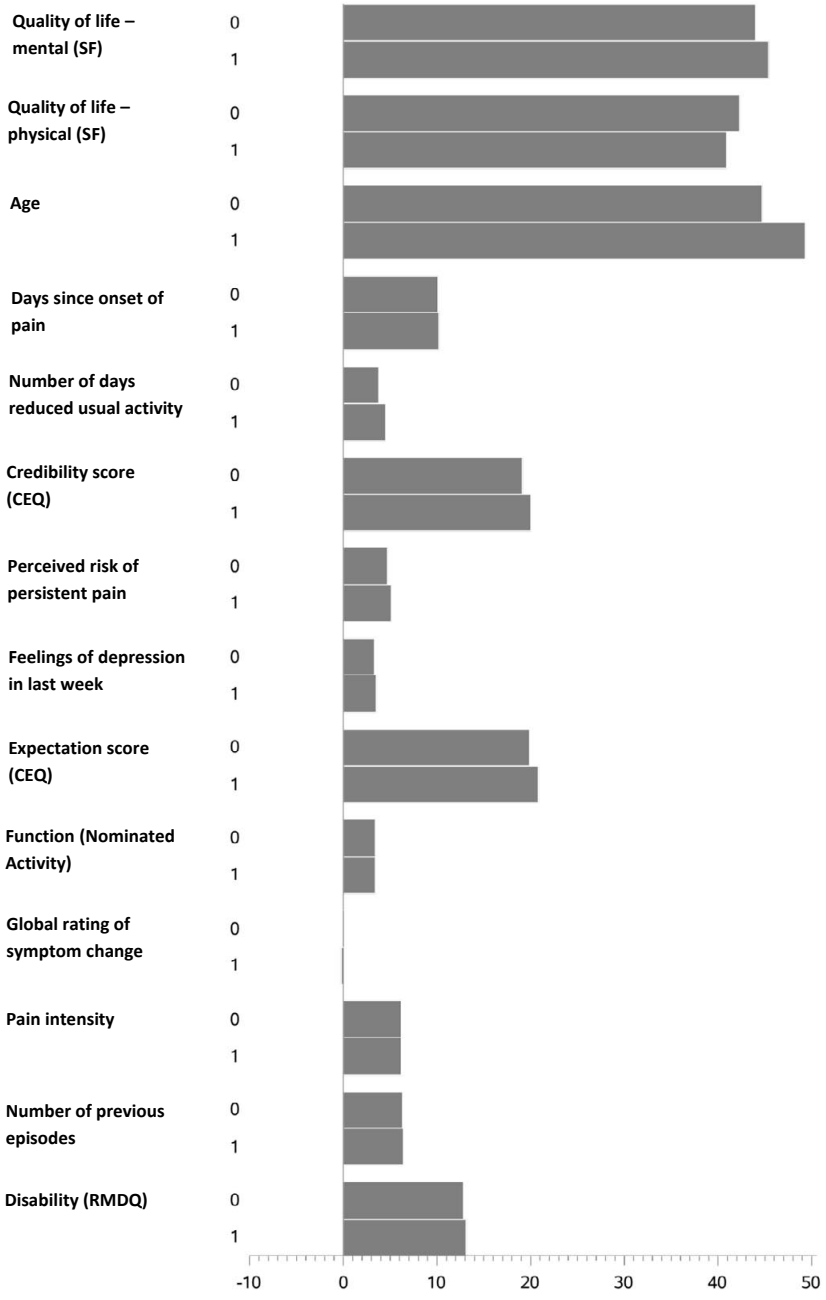
Supplementary Table 4: Assessment for residual imbalances between the regular paracetamol group and weighted placebo group in the means of continuous variables after the weighting procedure in propensity-weighted CACE analysis at week 2 with compliance defined as an average intake of ≥4 tablets per day.

Variable	Placebo mean (SD)	Paracetamol mean (SD)	Standardized difference
Quality of life – mental (SF-12)	45.46 (15.19)	44.08 (6.72)	0.0908
Quality of life – physical (SF-12)	40.91 (17.26)	42.34 (7.82)	0.0831
Age	49.33 (31.27)	44.77 (12.86)	0.146*
Days since onset of pain	10.28 (19.31)	10.12 (8.63)	0.00862
Number of days reduced usual activity	4.57 (12.95)	3.83 (5.53)	0.0571
Credibility score (CEQ)	20.04 (9.02)	19.13 (4.19)	0.100*
Perceived risk of persistent pain	5.14 (5.32)	4.77 (2.35)	0.0700
Feelings of depression in last week	3.57 (6.04)	3.34 (2.49)	0.0381
Expectation score (CEQ)	20.81 (9.00)	19.94 (4.64)	0.0967
Function (Nominated Activity)	3.40 (3.63)	3.43 (1.41)	0.00664
Global rating of symptom change	-0.12 (4.11)	-0.01 (1.81)	0.0262
Pain intensity	6.20 (3.53)	6.29 (1.59)	0.0251
Number of previous episodes	6.39 (27.04)	6.31 (11.33)	0.00326
Disability (RMDQ)	13.11 (10.65)	12.87 (4.67)	0.0233

St.Diffs: Standardized Differences; * under standardized differences indicate St.Diffs > 0.1; RMDQ: Roland Morris Disability Questionnaire; SF-12; 12-item Short Form Survey; CEQ: Credibility/Expectancy Questionnaire.



Supplementary Figure 1: Graphical assessment for residual imbalances between the regular paracetamol group and weighted placebo group in the distribution of categorical variables after the weighting procedure in propensity-weighted CACE analysis at week 2 with compliance defined as an average intake of ≥4 tablets per day.



Supplementary Figure 2: Graphical assessment for residual imbalances between the regular paracetamol group and weighted placebo group in the means of continuous variables after the weighting procedure in propensity-weighted CACE analysis at week 2 with compliance defined as an average intake of ≥4 tablets per day.
0: Paracetamol; 1: Placebo.