Data Analysis Plan (DAP)

Proactive clinical **R**eview of patients taking **O**pioid **M**edicines long-term for persistent **P**ain led by clinical **P**harmacists in primary care **T**eams (**PROMPPT**) cluster randomised trial

Version 1.0

Date: 18th May 2025

ISRCTN/Clinical trials.gov Number: ISRCTN 45616481

This document has been written based on version 1.6. of the study protocol, dated the 2nd of April 2024. Our study protocol has also been published in NIHR Open Research (Ashworth et al. 2025)

Data Analysis Plan (DAP) revision history

Protocol version	Updated DAP version	Section number changed	Description of and reason for the change	Date changed

Roles and responsibilities

The undersigned have written the data analysis plan for the PROMPPT trial and agree its content:

Name	Role	Signature	Date
Dr Elaine Nicholls	Statistician	[redacted]	18/05/25
Dr Julie Ashworth	Associate Investigator	[redacted]	21/05/25
Professor Christian Mallen	Chief Investigator	[redacted]	21/05/25

The undersigned have approved the content of the analysis plan:

Name	Role	Signature	Date
Prof. Richard McManus*	TSC Chair (on behalf of the TSC)	[redacted]	
Prof. Debi Bhattacharya*	DMC Chair (on behalf of the DMC)	[redacted]	

*Signatures may be in the form of an e-mail of endorsement from the TSC/DMC chair that will be printed and stored in the study master file

Elaine Nicholls will undertake analysis of the clinical effectiveness data. Analysis of the primary clinical endpoint will be performed independently by a second statistician, who will be employed by Keele Clinical Trials Unit but not part of the study team. This will ensure the accuracy and integrity of the main study findings.

Contents

1 Introduction

1.1 Background and rationale

Persistent pain, defined as long-term pain not caused by cancer, affects almost half the UK adult population, with 10-14% (around 8 million adults) reporting that persistent pain causes moderate or severe interference with life. In the UK, most persistent pain is managed in primary care, often using prescribed analgesics, with the use of opioid (morphine-like) analgesics increasing markedly during the last 20 years. However, it is now recognised that many people living with persistent pain do not obtain useful relief from opioids, and with opioid-related side-effects, including constipation, nausea, dizziness, and sedation, this worsens quality of life.

The aim of the PROMPPT trial is therefore to investigate the clinical and cost-effectiveness of a practice pharmacist-led primary care intervention (PROMPPT intervention), which is a proactive pain review for patients prescribed long-term opioids for persistent pain, which aims to reduce opioid use, where appropriate, and to support patients to live well with persistent pain.

1.2 Objectives

Primary objective

To determine, in patients prescribed opioids long-term (≥6months) for persistent non-cancer pain, whether providing the PROMPPT intervention (practice pharmacist-led primary care pain review) is more likely to reduce opioid use, without increasing pain/pain-related interference, at 12-month follow-up compared with usual primary care review of patients who are prescribed opioids long-term for persistent non-cancer pain.

Secondary objectives

1) To determine, in patients prescribed opioids long-term (≥6months) for persistent pain, the differences, between treatment arms, in secondary clinical outcomes including pain, pain-related interference, use of opioid and non-opioid pain medicines, confidence to cope with pain, symptoms of depression and anxiety, presence and severity of opioid-related side-effects, and health-related quality of life at 3, 6 and 12-month follow-up.

2) To determine the differences between treatment arms in GP practice-level prescribing of opioids, non-opioid analgesics and other potentially sedating medicines commonly prescribed for patients with persistent pain at 12-month follow-up.

3) To conduct a health economic evaluation to estimate the cost-effectiveness of providing the PROMPPT intervention versus usual primary care for patients prescribed opioids long-term (≥6months) for persistent pain.

4) To conduct a process evaluation to explore potential factors influencing trial results and understand how the PROMPPT intervention was used and perceived by patients and clinicians.

This analysis plan aims to cover the clinical effectiveness analysis only (i.e. primary objective and secondary objectives 1 and 2 excluding analysis of health-related quality of life. The cost-effectiveness

analysis, which will include the evaluation of health-related quality of life, and the process evaluation will be reported separately.

1.3 Estimands for the co-primary outcomes at the primary end-point (12-month follow-up)

Table 1.3.1: Estimands for the co-primary outcome at the primary endpoint based on the ICH E9 statistical principles for clinical trials.

Attribute	
Treatment	**PROMPPT intervention** compared to **Usual primary care for patients prescribed opioids long-term for persistent pain**, in the context of treatment delivery in the UK health service. Further details of the interventions are described in the study protocol (Ashworth et al. 2025; NIHR Open Research).
Population	Patients who are registered at participating UK GP practices that meet the following criteria: 1) List size >= 5000$^{\alpha}$ 2) Has a practice pharmacist working within the practice for at least one session per week or pro rata if working at Primary Care Network (PCN) level 3) The practice pharmacist sees patients for face-to-face and/or remote consultations in the practice 4) The practice pharmacist is an independent prescriber 5) The practice pharmacist consents to participate in the process evaluation including observation/audio-recording of a sample of PROMPPT consultations and an interview 6) One GP from the practice consents to participate in an interview 7) The general practice System of Choice (GPSoC) is either SystmOne or EMIS and who are aged 18 years and over and prescribed any opioid analgesic (defined as any opioid or opioid/paracetamol combination analgesic from sections 4.7.2 and 4.7.1 British National Formulary (BNF) 76th Edition 2018) for chronic non-cancer pain continuously for ≥6 months, with a prescription issued within the previous 2 months. Participants not included in the trial are those who are being treated for acute pain (self-limiting pain, for example after injury or surgery), pain associated with cancer and patients with terminal illness (life expectancy <6m); vulnerable patients (e.g. severe mental illness, learning difficulties, dementia) or patients currently receiving treatment for substance misuse

Outcome	Co-primary outcome 1: reduction in opioid use at 12-month follow-up Co-primary outcome 2: total Brief Pain Inventory score (BPI) at 12-month follow-up See table Table 4.1.1 for outcome definitions
Population-level summary	Co-primary outcome 1: reduction in opioid use: Odds ratio (covariate adjusted) Co-primary outcome 2: total Brief Pain Inventory score (BPI): Mean difference (covariate adjusted) Both co-primary outcomes will be summarised using participant-average treatment effects given we are interested in how effective the PROMPPT intervention is for the average patient (rather than how effective the intervention is for the average cluster) (Kahan et al. 2022). We hypothesise, however, that cluster size will be non-informative i.e. that the effect of treatment will be same irrespective of GP practice size, so it is likely that a participant-average and cluster-average treatment effect would be similar (Kahan et al. 2022)
Intercurrent events	Analysis population 1, predominantly focussed on a Treatment Policy approach, except for those participants who were identified as not being eligible for the trial after randomisation and the event of death, which are treated using a "Principal stratum" and "while alive" strategy respectively (see section Table 2.3.2 for further details)

α Note that there were a small number of practices that had list sizes <5000 to ensure that study recruitment targets were met.

1.4 Trial design

The PROMPPT trial is a pragmatic, multi-centre, 2-arm cluster randomised control trial with a co-primary outcome, an internal pilot study, a linked health economic evaluation and a mixed methods process evaluation. The two co-primary outcomes in the trial are measured at the 12-month follow-up: (1) Binary outcome to indicate if the participant has achieved at least a 25% reduction in opioid use from their baseline level (yes/no); (2) Brief Pain Inventory (BPI) total score (0 – 10). Primary outcome (1) will be tested using a superiority hypothesis and primary outcome (2) with a non-inferiority hypothesis to test the overall hypothesis that, compared to usual care, providing the PROMPPT intervention will be more likely to reduce opioid use, but that this will not lead to an (unintended) increase in pain/pain-related interference.

1.5 Randomisation

The randomisation process is described in the study protocol, but briefly, general practices are randomised (1:1 allocation to the PROMPPT intervention or usual care) in balanced blocks of size 2, stratified by regional centre, by an independent statistician affiliated with Keele CTU. Although a block size of 2 is small, we considered this necessary to increase the likelihood of achieving an equal number

of general practices in each arm of the trial; a consideration that was particularly important as we planned to include only a relatively small number of clusters in the study from each regional site.

We mitigated against the risk of the next treatment allocation being predicted in the study by ensuring that the general practices were randomised in batches and were not randomised directly in the order they were given to the external statistician by the trial manager. This was to ensure that the trial manager, who supplied the details on the general practices requiring randomisation, would not be able to predict the next treatment allocation in the list when details of the given allocations were returned to them e.g. if three practices were randomised in a batch, the trial manager would not know which practice was allocated last, so could then not (confidently) predict the next treatment arm allocation in the study.

The randomisation schedule is determined prior to the trial commencing and according to Keele University's standard operating procedures (SOPs).

1.6 Sample size

As defined in section 1.4, our first co-primary outcome is opioid reduction between baseline and 12 months (yes/no). We judge from our clinical experience that a 20% difference in the proportion of patients reporting at least a 25% reduction in opioid use between the intervention and control arms represents a meaningful difference. Therefore, if we assume that 40% of the control arm and 60% of the intervention arm will reduce opioid use over 12 months, we estimate (using the "power twoproportions" command in STATA) that a total sample size of 260 would be needed for a non-cluster RCT (power 90%, two-tailed statistical significance 5%). Note also that the estimated percentage difference is centred around 50%, which gives the largest sample size needed when considering differences in proportions.

Our assumption that 40% in the control group will reduce opioids is based on our previous (unpublished) analysis of opioid prescribing data from the Clinical Practice Research Datalink (CPRD) indicating that approximately 20% of patients who have been prescribed opioids for 6 months or longer will stop taking opioids completely without any intervention and our assumption, based on clinical experience, that a further 20% of the control group, whilst not stopping completely, will reduce their opioid use according to our definition.

For the second co-primary outcome (total BPI), we assume a 0.6-point non-inferiority margin (for justification see the study protocol) and an SD of 2. We have estimated the standard deviation from published literature (Krebs et al. 2010; Kean et al. 2016), and from our feasibility study. With power 90%, one-tailed statistical significance 2.5%, a sample size of 468 would be needed for a non-cluster RCT (estimated using the "ssi" command in STATA).

Given the co-primary study hypotheses, we have considered that power for our study may not be persevered at 90%. However, given that sample size differs considerably between the two hypotheses the impact on study power is lessened. If we conservatively assume that the two co-primary outcomes are not correlated, using the intersection-union test (Offen et al. 2007; Gillespie et al. 2018) the overall power for the study is 0.89 (i.e. the power for the first co-primary outcome*the power for the second co-primary outcome when power is calculated on the hypothesis with the largest sample size N = 468 i.e. 0.99 * 0.9). As 0.89 is close to 0.9, the sample size was not adjusted to account for the loss of power from testing a co-primary hypothesis.

Given the cluster design, our sample size is inflated to account for clustering. Assuming (i) an intra-cluster correlation coefficient 0.01 (estimated as an average ICC from our POST trial and other similar trials in primary care (Adams et al. 2004; Mallen et al. 2017; Stuart et al. 2020), (ii) a coefficient of variation 0.40 (estimated to be slightly lower than other estimates of CV in primary care trials (Eldridge et al. 2006) to reflect that our study design will cap the number of patients invited per practice so that large general practices are not over-represented in the data), and (iii) an average of 30 patients recruited per practice (estimated from our feasibility study), this translates to a design effect of 1.34 (Rutterford et al. 2015). Given a 20% loss to follow-up at 3 months in our feasibility study, we have conservatively assumed a 30% loss to questionnaire follow-up at the later 12-month follow-up and therefore the sample size for the individually randomised RCT needs to be inflated by 1.91. We therefore plan to recruit 896 patients (448 per arm) from 30 general practices and review the assumptions made in the sample size calculation in the internal pilot study and, if required, we have the potential to increase the number of GP practices in the study to a maximum of 40 practices to achieve the sample size requirements of the study.

1.7 Framework

The framework for the primary analysis includes both a superiority and non-inferiority hypothesis (Schumi et. al. 2011) as specified below.

Superiority hypothesis

Null hypothesis: *there is no difference in the proportion of participants reporting a reduction in their opioid use (defined as a reduction in opioid use of >= 25% (*Table 4.1.1)) at *12-months between the treatment arms of the trial.*

Alternative hypothesis: *there is a significant difference in the proportion of participants reporting a reduction in their opioid use (defined as a reduction in opioid use of >= 25% (*Table 4.1.1)) *at 12-months between the treatment arms of the trial.*

The superiority hypothesis will be tested using a 2-sided 5% significance level and results presented using 95% confidence intervals.

Non-inferiority hypothesis

A non-inferiority margin of 0.6 for the BPI total score is used in the trial, which has been defined and justified in our study protocol. Hence, our hypothesis for the non-inferiority outcome is:

Null hypothesis: *the mean BPI total score is higher for the PROMPPT intervention than usual primary care (i.e. the PROMPPT intervention is worse than usual primary care) and the mean difference is greater than 0.6 at the 12-month follow-up.*

Alternative hypothesis: *the difference in the mean BPI total score between the PROMPPT intervention and usual primary care is less than 0.6 at the 12-month follow-up*

The non-inferiority hypothesis will be tested using a 2-sided 5% significance level (which is equivalent to a one-sided statistical significance of 2.5%) and results presented using 95% confidence intervals.

1.8 Interim analyses and stopping rules

No interim analysis of treatment effectiveness is planned before the end of the trial.

1.9 Timing of analysis

Treatment effectiveness analyses will only be conducted after data from the last-person's 12-month questionnaire has been entered onto the study database, and after all data queries relating to the effectiveness analysis have been resolved. After verification of the primary analysis of the primary outcome by an external statistician, the data will be unblinded.

1.10 Timing of outcome assessments

Primary and secondary outcomes for the clinical effectiveness analysis are in Table 1.10.1 alongside their time-points of data collection. References for each of the outcome measures are given in the protocol and further details on how each outcome measure is measured/scored for use in the trial is given in section 4.

Table 1.10.1: Outcome measures to assess clinical effectiveness.

	Baseline	3-months	6-months	12-months
Co-primary outcomes				
Reduction in opioid use since baseline (yes/no)				x
BPI: Total score				x
Key secondary outcomes				
Opioid use: Daily MED	x	x	x	x
BPI: Total score	x	x	x	x
BPI: Pain severity	x	x	x	x
BPI: Pain interference	x	x	x	x
Non-opioid pain medicines use (Table 12.2.4)	x	x	x	x
Opioid-related side-effects (Table 12.2.5)	x	x	x	x
Additional secondary outcomes				
Depression (PHQ-8)	x	x	x	x
Anxiety (GAD-7)	x	x	x	x
Pain Self-Efficacy (PSEQ)	x	x	x	x

BPI Brief Pain Inventory; MED Morphine equivalent dose; PHQ Patient Health Questionnaire; GAD Generalised Anxiety Disorder Assessment; PSEQ Pain self-efficacy questionnaire

In addition, for each general practice in the study, we will collect outcome data from the electronic health records on the percentage of adult patients, excluding those with both a coded cancer diagnosis and a code for palliative care, who are prescribed opioids, non-opioid analgesics (specifically paracetamol, topical pain treatments, non-steroidal anti-inflammatory drugs (NSAIDs), nefopam, gabapentinoids, antidepressants, benzodiazepines and Z-drug hypnotics (zopiclone, zolpidem) in two time-periods: (1) in the 90 day period up to the date the first participant was recruited at that practice

(2) in the 90 day period up to the date 12-months after the first participant was recruited at that practice.

2 Statistical Principles

2.1 Confidence intervals and p-values

All statistical tests will be 2-sided and tested with 5% significance i.e. presented with 95% confidence intervals. We do not plan to adjust the significance level of our pre-planned analyses to account for multiple testing as we have stated our outcomes and research hypotheses *a priori*.

2.2 Protocol deviations

Treatment will be deemed to have been delivered according to protocol in the PROMPPT arm of the trial if the participant attends the initial pain review (the follow-up pain review was optional for the patients to attend). Any other protocol deviations related to trial procedures will be reported in text or tabular format as appropriate.

It is anticipated that there may be participants who report they are not taking any opioids on the baseline questionnaire, despite this being recorded in their medical record. Such participants will be classified as being "subsequently ineligible" and will not be included in the trial analysis.

2.3 Analysis populations

Our analysis populations are defined by how intercurrent events are handled in the analysis and described using three out of the four strategies from the ICH E9 (R1) addendum on estimands (Treatment policy, While-on-treatment, and Principal stratum) (Clark et al. 2022). Details of how each strategy will be applied to our data are given in Table 2.3.1. Our analysis populations are described in Table 2.3.2.

Table 2.3.1: Strategies for handling intercurrent events and how they will be implemented in the trial data.

Strategy	Implementation in the data
Treatment policy	The value for the variable of interest will be used in the analysis regardless of whether the intercurrent event occurs
While-on-treatment/ While-alive	Any data collected after the intercurrent event will be deleted in the data. However, the data will remain as missing in the analysis and not imputed
Principal stratum	Participants meeting the definition for the "principal stratum" will be analysed

Table 2.3.2: Analysis populations defined by strategy to handle intercurrent events.

Post-randomisation intercurrent events	Analysis population 1	Analysis population 2
Participants identified as not being eligible for the study after randomisation e.g. they were not actually taking opioids for their pain therefore ineligible for the study	Principal stratum	Principal stratum
Protocol deviations that impact clinical outcome data collection	Treatment policy	Treatment policy
Participants that do not attend the pain review if they are randomised to the PROMPPT intervention	Treatment policy	Principal stratum
Adverse events	Treatment policy	Treatment policy
Death[α]	While alive	While alive

[α] We chose a 'while alive' strategy for death to avoid applying the unrealistic assumption of an immortal cohort (Wen L et al. 2017)

3 Trial Population

CONSORT flow diagrams (Schulz et al. 2010 and Campbell et al. 2012) will document the recruitment of clusters (GP practices) to the study (Figure 12.1.1) and the flow of participants through the study (Figure 12.1.2). The participant flow diagram will include both the recruitment and follow-up of patients through the study, along with reasons for ineligibility or withdrawal (if given). It will also specify the timing of the withdrawal and whether the withdrawal was from treatment only, or from the trial overall.

Baseline characteristics of the clusters, and participants, will be described overall (using numbers and percentages for categorical data, means and standard deviations for normally distributed continuous data and median and inter-quartile range for skewed continuous data) and by treatment arm (Table 12.2.1). This analysis will be based on analysis population 1. We will also report (in the text of the results paper) the mean (standard deviation) age, opioid use and sex profiles of the patients who were invited to take part in the study to see how they compare to those recruited in the trial.

The patient-level component of Table 12.2.1 will also be stratified by whether participants have returned a 12-month follow-up questionnaire and will be used to assess the impact of loss to follow-up on the data. For all analyses in section 3, no statistical tests will be performed to compare participant characteristics by group. Instead, the magnitude of any differences between groups will be considered and evaluated for clinical importance.

4 Outcome definitions

4.1 Derivation rules

Derivation rules used to generate the study variables are shown in Table 4.1.1. Prior to implementation of the scoring procedures, the data will be processed using the data coding rules described in our internal Standard Operating Procedure (SOP) 16 – Data Analysis – Version 5.0, which provides guidance on how to process multiple responses to a single questionnaire item and what to do if multiple questionnaires are returned for a single participant – a situation that could arise because of our reminder mailing process. We will follow this guidance and document the decision-making process for any participant as it is required using Table 13.1.1. We will store the computer syntax used to derive each of the variables in the Statistical Analysis folder of the Trial Master file to ensure we can clearly trace how the raw data are converted into the data that are analysed.

Table 4.1.1: Description of the derivation of study outcome measures and other derived measures used in the trial analysis.

Outcome measure	Scoring rule	Missing data considerations	Score interpretation	Scoring reference website (if applicable)
Clinical Effectiveness measures				
Opioid use (Morphine Equivalent Dose)	At baseline and each follow-up time point, a pain medicines use questionnaire will collect self-reported data on opioid use (prescribed and non-prescribed) during the previous 7-days. Opioid medicines use will be standardised to a daily oral morphine equivalent dose (MED) using a Microsoft Excel calculator which was developed by the research team based on published conversion factors.	Information regarding drug name, strength/dose and frequency of use are required to calculated daily MED. Where there are ambiguous or incomplete questionnaire entries that interfere with calculation of daily MED, a member of Keele CTU staff will try to contact participants by telephone or in writing to obtain the missing data. Incomplete or ambiguous questionnaire entries regarding prescribed pain medicines can also be checked with the electronic prescribing record, and / or Practice Pharmacists or GP to obtain or estimate the missing information, if necessary.	The change in daily MED (in mg) between baseline and 12-month follow-up will be calculated. The **percentage** change in daily MED (in mg) between baseline and 12-month follow-up will be calculated by dividing the change in daily MED by the baseline value. The primary outcome for the trial will be define as a binary outcome: 0 = < 25% reduction in daily MED from baseline 1 = >=25% reduction in daily MED from baseline	Opioid Equivalence Conversion Factors Von Korff M et al. Clin J Pain. 2008;24(6):521-7. Faculty of Pain Medicine. Opioids Aware. Dose equivalents and changing opioids | Faculty of Pain Medicine (fpm.ac.uk)Updated August 2020 British National Formulary (BNF) Prescribing in palliative care | Medicines guidance | BNF | NICE

BPI: Pain severity	Scored using the instructions in the scoring reference.	All items need to be present for a score to be calculated	Range 0 – 10 Lower score: lesser problems	The Brief Pain Inventory User Guide: 2009 https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf
BPI: Pain interference	Scored using the instructions in the scoring reference.	A score will be calculated for participants if they have completed 4 or more items in the pain interference scale	Range 0 – 10 Lower score: lesser problems	The Brief Pain Inventory User Guide: 2009 https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf
BPI: total score	Average of BPI Pain Severity and BPI Pain interference	Both subscales need to be present for a score to be calculated	Range 0 – 10 Lower score: lesser problems	Not applicable
Depression (PHQ-8)	Scored using the instructions in the scoring reference	No guide on how to handle missing data is provided in the tool. We will	Range 0 – 24	Kroenke et al. 2009

		therefore use guidance in SOP16 version 5.0 that for scales with between 5 and 10 items a scale score is calculated if >= 80% of the items are present. We will calculate a PHQ-8 score if 7 out of the 8 items are present, by replacing the missing value with the mean of the items that have been completed.	Lower score: lesser problems Cut-offs also used as per the scoring instructions: *No significant depression: 0-4* *Mild depression: 5-9* *Moderate depression: 10-14* *Moderately severe depression: 15-19* *Severe depression: 20-24*	
Anxiety (GAD-7)	Scored using the instructions in the scoring reference	No guide on how to handle missing data is provided in the tool. We will therefore use guidance in SOP16 version 5.0 that for scales with between 5 and 10 items a scale score is calculated if >= 80% of the items are present. We will calculate a GAD-7 score if 6 out of the 7 items are present, by replacing the missing value with the mean of the items that have been completed.	Range 0 – 21 Lower score: lesser problems Cut-offs also used as per the scoring instructions: *No significant anxiety: 0-4* *Mild anxiety: 5-9* *Moderate anxiety: 10-14* *Severe anxiety: 15-21*	Spitzer et al. 2006
Pain Self-Efficacy (PSEQ)	Scored using the instructions in the scoring reference.	No guide on how to handle missing data is provided in the tool. We will therefore use guidance in SOP16 version 5.0 that for scales with between 5 and 10 items a scale score is calculated if >= 80% of the items are present. We will calculate a PSEQ score	Range 0 – 60 Lower score: lower level of confidence in dealing with pain	Nicholas et al. 2007

		if 8 out of the 10 items are present, by replacing the missing values with the mean of the items that have been completed.		
Health Economics Outcome				
EQ-5D-5L	Scored using the Cross-walk value set (van Hout B et al 2012).	All five EQ-5D items are required to be present for an EQ-5D score to be calculated	Range -0.594 – 1 Lower score: worse quality of life	EQ-5D-5L User Guide. Version 3.0. September 2019 https://euroqol.org
Descriptive variables				
Age	Our aim, as far as this is possible, is to calculate "Age" for all participants who have completed the baseline questionnaire. We will calculate "Age" as the number of years between "Date of Birth" and "Date of Baseline questionnaire". We note that when preparing the data for external data release, we will only include "Age" in the dataset, rather than "Date of Birth", to preserve participant anonymity	"Date of birth" will be taken from the date of birth on the baseline questionnaire. However, if this information is missing, we will use other sources of data in the trial (e.g. the follow-up questionnaires), where date of birth is recorded, to determine, where possible, what the missing "Date of Birth" should be. We will also use this process to check consistency of the recording of "Date of birth" across the different data sources. Where a lack of consistency is apparent, we will use the date of birth	Range 18 years and over	Not applicable

		that is most likely to be the true date of birth across the full range of date of birth responses we have received. We anticipate that there will be no missing data for the date of when the baseline questionnaire was completed. This date is the same as the date of consent, which needs to be completed for the patient to be part of the study.		
Sex	"Sex" will be taken from the baseline questionnaire	If information is missing for sex, we will use other sources of data in the trial, where sex is recorded, to determine, where possible, what the missing "Sex" should be (e.g. the follow-up questionnaires). We also use this process to check consistency of the recording of "Sex" across the different data sources. Where a lack of consistency is apparent, we will use the sex that is most likely to be the true sex across the full range of sex responses we have received.	0 = Male 1 = Female	Not applicable
Index of multiple deprivation (IMD) 2019.	Derived from postcode data	All responses used for analysis	Range 1 – 32844 Lower score most deprived	**Research report for 2019 coding:** https://assets.publishing.service.gov.uk/government/uploads/

			Also categorised into quintiles of deprivation 1 = IMD 1 to 6568 2 = IMD 6569 to 13137 3 = IMD 13138 to 19706 4 = IMD 19707 to 26275 5 = IMD 26276 to 32844 This corresponds with combining deciles of IMD into pairs i.e. deciles (1,2); (3,4); (5,6); (7,8); (9,10).	system/uploads/attachment_data/file/833947/IoD2019_Research_Report.pdf **Scoring calculator (2019 version)** https://imd-by-postcode.opendatacommunities.org/imd/2019
Population density score (persons per hectare)	Derived from postcode data	All responses used for analysis	1 = Most urban 2 = Very urban 3 = Urban 4 = Rural 5 = Most rural	Norman, Paul (2019), "UK small area characteristics 2011", Mendeley Data, V1, doi: 10.17632/yn47f2yrt2

Footnote: Where guidance has been given in the published tool as to how missing data should be handled this has been followed, otherwise guidance given in SOP 16 version 5.0 (Data analysis) has been used to determine the maximum number of missing items to allow in the score calculation. BPI Brief Pain Inventory; MED Morphine equivalent dose; PHQ Patient Health Questionnaire; GAD Generalised Anxiety Disorder Assessment; PSEQ Pain self-efficacy questionnaire

5 Analysis methods

5.1 Primary analysis

5.1.1 Statistical model

The statistical models for the primary analysis will be based on analysis population 1, after multiple imputation of missing data has been performed on the data (see section 5.1.4).

5.1.1.1 Co-primary outcome 1 – reduction in opioid use (as defined in Table 4.1.1)

Superiority hypothesis

The first co-primary outcome, opioid reduction (yes/no), will be modelled using a mixed logistic regression model (using the command melogit in STATA), with the mixed component needed to account for the clustering in the data due to randomisation at the GP practice level. The outcome for the model will be measured at the 12-month follow-up time-point. Predictor variables in the model will include fixed effects for baseline MED opioid dose, the adjusting covariates listed in section 5.1.2, and randomised treatment arm (usual care vs PROMPPT), along with a random effect (random intercept) for GP practice. We will also adjust the model for the cluster-level mean MED score at baseline to improve the precision of the model – an approach recommended by Hooper et al. 2018.

We will explore whether using different integration methods to fit the model to the data change the model findings and will report results from the model that is the best fit to the data (as measured by the model that gives the lowest Akaike information criterion (AIC) and Bayesian Information Criteria (BIC)). The treatment effect estimate from the model will be presented as an odds ratio, along with associated two-sided 95% confidence intervals (derived using robust standard errors), to explore whether there is an increased odds of reducing opioids in the PROMPPT group compared to usual care.

To give context to the odds ratio, the number and proportion of patients reducing opioids will be reported at the 12-month follow-up by treatment arm. We will also present the data as absolute risk differences as recommended by Turner et al. 2021, which will be calculated by refitted the primary analysis model with an identity link function (rather than a logit link function) in the mixed logit model Pedroza et al. 2016. The model intra-cluster correlation (ICC) will also be reported (Monsalves et al.2020) and stratified by treatment arm as recommended by Billot et al. 2024 (Table 12.2.2).

5.1.1.2 Co-primary outcome 2 – Brief Pain Inventory (BPI) total score

Non-inferiority hypothesis

The second co-primary outcome, the BPI total score (continuous measure), will be modelled using a mixed linear regression model (using the mixed command in STATA). The outcome for the model will be measured at the 12-month follow-up time-point. Predictor variables in the model will include fixed effects for baseline BPI total score, the adjusting covariates listed in section 5.1.2, and randomised treatment arm (usual care vs PROMPPT), along with a random effect (random intercept) for GP practice. We will also adjust the model for the cluster-level mean BPI total score at baseline to improve the precision of the model.

The model will be fitted using maximum likelihood estimation and robust standard errors if the total BPI score is not normally distributed (as evaluated by visual inspection of the data using a histogram).

We will explore whether a model that estimates separate independent residual terms for each GP practice is a better fit to the data than when a common independent residual is assumed across all GP practices, alongside varying the covariance structure for the residuals e.g. unstructured and exchangeable. We will present findings from the model that converges and gives the best fit to the data (as measured by the model that gives the lowest Akaike information criterion (AIC) and Bayesian Information Criteria (BIC)).

The treatment effect estimate from the model will be presented as a mean difference in the BPI total score comparing PROMPPT to usual care, along with an associated two-sided 95% confidence interval. To give context to the model, the mean and standard deviation of the BPI total score will be reported at the 12-month follow-up by treatment arm. The model intra-class correlation (ICC) will also be reported overall and by treatment arm (Table 12.2.2).

5.1.1.3 Interpretation of trial results from the statistical model

The superiority and non-inferiority hypotheses will be assessed simultaneously from the models described in 5.1.1.1 and 5.1.1.2.

Superiority will be concluded if the 95% two-sided confidence interval for the opioid reduction treatment effect does not contain one

Non-inferiority will be concluded if the upper limit of the 95% two-sided confidence interval for the mean difference in BPI score is <0.6.

For the trial to conclude that the PROMPPT intervention is effective we need to show that the intervention is superior for co-primary outcome 1 (reduction in opioid use) AND non-inferior for co-primary outcome 2 (BPI total score)

5.1.2 Adjusting covariates

Adjusting covariates will be included in the model as fixed effects. They will include the stratification variable used in the generation of the randomisation schedule i.e. region (West Midlands, East Midlands, Wessex, and Thames Valley/South Midlands), along with age (years), sex (male, female) and the patient-level index of multiple deprivation. The model that includes the adjusting covariates will be considered the primary analysis as recommended by Morris et al. 2022.

5.1.3 Checking model assumptions

The linear mixed model will be fitted using full information maximum likelihood (FIML) estimation, however, if the BPI total score follows a non-normal distribution, FIML with robust standard errors will be used to address that the assumption of normality is not met (as evaluated by visual inspection of a histogram of the total BPI score at the 12-month time-point). Assumptions for the linear mixed model will be explored as below, and if not met, this will be reported (Singer 2003):

1. A histogram of model residuals and random intercepts (estimated using empirical Bayes estimation/best unbiased linear predictors (BLUPs)) will be produced to ensure they are normally distributed.
2. Plots of the model residuals and random intercepts against study identification number will be generated to ensure no relationship exists and to identify any specific participants with large residuals or random intercepts (i.e., to check for outliers) (the model will not be re-run excluding outliers as this is a pragmatic trial, but if large, the number of outliers will be reported).

3. Plots of the random intercepts against the fixed effect predictors in the model and the residuals by GP practice. No relationship should exist in these plots; they will also be used to check whether the assumption of homogeneity of variance holds for each variable in the model.

The covariance between the residuals and the random intercepts will be inspected to ensure it is close to 0.

For the mixed logistic regression model a plot of the residuals will also be used to identify any outliers in the data, but such outliers will remain included in the analysis to reflect the pragmatic nature of the trial.

If the model assumptions are not met, or the model above does not converge in the data, we will modify the analysis plan to model the data using generalised estimating equations (GEE) or a cluster-level analysis (Billot et al. 2024)

5.1.4 Missing data

5.1.4.1 Descriptive statistics

The percentage of missing data will be calculated for each co-primary outcome and the secondary outcomes in the trial, and at each time-point, for analysis population 1. Missing data rates will be reported for the co-primary outcomes (Table 12.3.3). Missing data rates for the secondary outcomes will be summarised by inclusion of an overarching sentence in the paper e.g. “Missing data rates for the secondary outcomes was less than x% at all follow-up time-points”. Baseline characteristics of participants lost to follow-up at 12-months (i.e. not returning a questionnaire at the 12-month follow-up time-point) will be described, as defined in section 3.

5.1.4.2 Multiple imputation

Multiple imputation will be used to impute missing data in the primary analysis data set.

The imputation model will include: opioid use (daily MED), BPI pain severity, BPI pain interference, BPI total score, depression, anxiety, pain self-efficacy, non-opioid pain medicines use, opioid-related side effects, at all time-points where data are collected, the adjusting variables in the regression model (Section 5.1.2), GP practice, attendance at the pain review (included as a key predictor of missing data) and the EQ-5D at all time-points where data are collected (for the health economics analysis). We will impute the data for morphine equivalent dose as a continuous variable and then use “mi passive” in STATA to compute the first co-primary outcome i.e. defining participants that have reduced their opioid use by 25% or more from baseline to the 12-month follow-up. We aim to ensure that all variables in our analysis models on imputed data are included in the imputation model (Austin et al. 2021).

To account for the clustered nature of the data, we will initially fit GP practice as a categorical variable in the imputation model, however, given we will have between 35 to 40 clusters in the analysis, this may be too many categories for model convergence to be achieved. If this arises, we will seek to fit a separate imputation model for each GP practice as an alternative model option or we will consider using a multivariate normal model to impute all clusters simultaneously using the “mvn” imputation option in STATA (STATA n.d.).

As the data for pain review attendance is only measured for participants in the intervention arm of the trial, we will assume this variable is coded as “not attended” for all participants in the control practices. To enable treatment interactions to be included in our analysis we will (potentially) fit the imputation

model separately for each arm of the trial (White et al. 2011). This model may not be required, however, if a separate imputation model is fitted for each GP practice (treatment arm will naturally be defined in this model due to the clustered nature of study design).

The imputation model will be fitted using Multiple Imputation by chained equations (MICE), assume the data are missing at random, and will include X imputed datasets. The value of X will be defined initially to equal the percentage of participants with missing data on at least one variable in the primary regression model of interest. The resulting models for the primary and secondary outcomes at the primary endpoint will then be checked to ensure that the Monte Carlo error (MCE) estimates for all parameter estimates are <= 10% of their respective standard errors, that the MCEs for the test statistics are <=0.1 and that the MCEs for the p-values are <= 0.01. If this is not satisfied, then the number of imputations will be increased until this is achieved, and a satisfactory level of reproducibility shown (White et al. 2011). We chose to use MICE as our initial imputation method, rather than Multi-Variate Normal Imputation (MVNI), as MICE offers greater flexibility to form imputation models outside any known standard multivariate density function (van Buuren et al. 2007)

The imputation model will include continuous outcome measures, modelled using predictive mean matching (nearest neighbours = 10 (Morris et al. 2014)); binary outcomes, modelled using logistic regression; and ordinal outcomes, modelled using ordinal regression. Predictive mean matching will be used for continuous measures as this method is suitable for the imputation of both normally distributed and skewed outcomes and produces imputed values restricted to the range of values that the measure requiring imputation can take (Morris et al. 2014).

The imputation model will be fitted to the data, however, given the complexity of the model, it may arise that the imputation model will breakdown, so it may not be possible for it to be fitted to the data (this is a real possibility given the large number of categorical variables in the model, whose data format are known to make model convergence challenging). If this occurs, then the techniques described in section 13.3 will be explored to see how the imputation model can be adapted to ensure it can be fitted to the data. If adaptations need to be made to the imputation model, this will be explained in the results publication for the trial. If a successful imputation model can be developed, analysis models will then be fitted, and Rubin's rules (Rubin and Schenker, 1991, Austin et al. 2021) used to combine the treatment effects and their associated standard errors across the imputed data sets. This will provide a single estimate of the treatment effect for each analysis outcome.

5.1.4.3 Checking the imputation model

Descriptive graphs (histograms, box plots) and statistics (means, standard deviations, ranges) will be used to check that the imputed data for each variable appear theoretically plausible from what is known about the (clinical) range of the scales in the observed data. We will also check that the distribution of the co-primary outcomes in the imputed datasets are similar to the observed data (we plan to do this to increase our understanding of the impact that multiple imputation has on our dataset as if the data are missing not at random then it may not be of concern if the imputed data differ from the observed data).

5.2 Sensitivity analyses for the primary analysis

Sensitivity analyses will be conducted for the co-primary outcomes and the results compared to the primary analysis. Results of the sensitivity analyses will be presented using outline Table 12.2.2.

5.2.1 Accounting for a small number of clusters in the trial analysis

We anticipate that we will have between 35 and 40 clusters (GP practices in this study), which is borderline as to whether using a small-sample correction is required for the analysis (normally required when the number of clusters is less than 40 (Leyrat et al. 2017)). To ensure that we have considered this issue, we will run a sensitivity analysis for the continuous co-primary outcome measure (the total BPI score) whereby the degrees of freedom are corrected using the Satterthwaite method to account for the small number of clusters in the trial (Leyrat et al. 2017). We will report the results of this revised model only if it changes the conclusion from the primary analysis as we are anticipating that the results of this model will be very similar to the primary analysis, given that the number of clusters in the trial is not excessively small.

5.2.2 Primary analysis estimated when data are assumed to be missing not at random.

Our primary analysis assumes that data are "missing at random" (MAR), however, it may be that this assumption does not hold in our data set, particularly as we anticipate that our follow-up response rates in the trial are lower than expected in our sample size calculation.

We will therefore test how sensitive our primary analysis is to this assumption using controlled imputation (Hayati et al. 2018, Cro et al. 2020). We will use the delta method of controlled imputation applied to the primary analysis in section 5.1, with the values of delta calculated separately for the daily MED and BPI total score outcomes. It is unlikely that we will have rich information on the reason for withdrawal, so we will use the same value of delta irrespective of the reason for withdrawal.

We plan to use trial data to define a range of delta values to test in the data. We will calculate the mean change in each outcome (daily MED and BPI total score) between baseline and the 12-month follow-up and will define a range of delta values as: 25%, 50% ,75% and 100% of the mean change as calculated (this will be a cluster-level calculation i.e. the mean change will be calculated for each GP practice separately and the average taken, and calculated without any knowledge of cluster treatment allocation). We will then review these values against our knowledge of the clinical area and our outcome of interest to see whether they represent a plausible change that could occur in a real-life setting. If they do not, e.g., if 100% of the mean change is unlikely to happen, then we will highlight this as a limitation of the analysis. We will consider both scenarios, that participants who withdraw from the trial could have better, or worse outcomes than predicted under a MAR assumption, by changing the sign of the delta coefficient in each analysis from positive to negative.

We will impute the data as described in section 5.1.4.2, but will add on values of delta to the imputed scores for daily MED and the BPI total score at 12-months for those participants with a missing score on each outcome respectively. We will then use "mi passive" to generate the co-primary outcome that is measured on a binary scale (i.e. opioid reduction (yes/no)) and will re-run the primary analysis on this revised dataset to explore how extreme the missing data assumptions would need to be in the data before an alternative analysis conclusion would be drawn.

5.3 Secondary analysis

5.3.1 Patient-level clinical outcomes

All secondary outcome models will be run on a dataset with no imputation of missing data, as missing data will be handled via the use of longitudinal modelling and will assume that the data are missing at random.

For the continuous co-primary (total BPI score) and the secondary outcomes (BPI pain, BPI pain interference, daily MED, Depression, Anxiety, and Self-efficacy) we will use descriptive plots (spaghetti plots) to initially understand how the outcome data change over time (stratified by treatment arm and GP practice). Longitudinal multilevel linear mixed models will be used to explore whether change in the outcomes over time differ by treatment arm (Twisk J et al. 2018; Bell et. al. 2020; Billot et al. 2024). The models will be fitted to the outcomes, as measured at 3-, 6- and 12-month follow-up and will include random effect terms to account for clustering by GP practice, and that data points are not independent given that each patient has responded at multiple time-points. Model fixed effects will include time (coded as a categorical variable at 3-, 6- or 12-months), treatment arm, the interaction between time and treatment arm, the baseline in the outcome of interest (e.g. baseline total BPI score when the total BPI score is being modelled in the data), the GP cluster-level mean of the outcome of interest at baseline, and the adjusting covariates listed in 5.1.2.

The models will be fitted using full information maximum likelihood (FIML) estimation if the outcome follows a normal distribution, or FIML with robust standard errors if this assumption is not met (as evaluated by visual inspection of the data using a histogram). We will fit the model using an unstructured variance-covariance structure for the random effects and an independence structure for the residuals, but we will explore how sensitive our results are to varying the variance-covariance structure for the random effects (e.g. by trying other structures such as exchangeable or independent) and residuals and whether model fit can be improved by changing these assumptions (taking the best fitting model to be the one with the lowest AIC and BIC values). We will use the pairwise comparison command in STATA to estimate treatment effects at each time-point. The beta coefficients and 95% confidence intervals for the interaction term will also be reported along with the ICCs at the level of the GP practice and for participants within GP practice (Monsalves et al.2020) as shown in outline Table 12.2.3.

For the remaining continuous secondary outcome, percentage change in daily MEDs (baseline to the 12-month follow-up), multilevel linear mixed models will also be used to model this outcome using the method described above, but as the data will be a single observation per participant, this model will not include a term for “time”, the interaction between “time and treatment”, nor a random effect term to account for the lack of independence between data collected on the same person at multiple time-points. The model results will therefore be presented as an adjusted mean difference and 95% confidence interval between the two treatment arms in the trial. Only a single ICC value will be reported to reflect that this model now only includes one random effect of interest (Table 12.2.3).

For the secondary outcomes measured on a non-continuous scale (non-opioid pain medicine use and treatment side effects) logistic (using melogit in STATA), and ordinal (using meologit in STATA), mixed effect models will be used respectively. The model structure will be like that used in the linear mixed models for outcomes collected at more than one time-point, however, we will not adjust for the cluster-mean in the outcome of interest, as the outcome is not measured on a continuous scale. Also, given the observation in Twisk et al. 2018 around non-collapsibility that can occur in logistic regression models, our primary model will not adjust for the baseline proportion in the outcome of interest. We will however, complete a sensitivity analysis to explore the impact that adjusting for baseline has on the findings. We will present model results from both the logistic and ordinal models as odds ratios and 95% confidence intervals (Table 12.2.4 and Table 12.2.5).

Assumptions for the longitudinal mixed models will be explored, using the same strategy as for the primary analysis. In addition, we will explore the suitability of assuming a proportional odds model when analysing the ordinal side-effects data. We will fit a non-proportional odds model to the data

and will consider whether the magnitude of the odds ratios is similar for each cut-point of the ordinal outcome. At the time of writing the analysis plan, it is not clear that STATA has an option to do this model using a single command (we are aware of a user written command "regoprob2" that can fit such models, but note that this model can only include a single random effect, which is not ideal, given that our data is clustered at the GP practice level and by patients providing multiple points of data collection). We will explore other software programs to fit this model, but as an alternative solution we will reduce the data down to multiple binary mixed logistic regression models to emulate what would be achieved by fitting a partial proportional odds model using a single analysis command. If the odds ratios for the non-partial/partial odds model differ to a degree that they would change the overall conclusion from the model, this will be reported in any publication arising from the trial.

5.3.2 GP Practice-level secondary outcomes

Practice-level prescribing of opioids, non-opioid analgesics, commonly prescribed for patients with persistent pain will be assessed from electronic prescribing records at baseline (defined as the period 90 days up to the date the first participant was recruited at that practice) and 12-months (defined as the period 90 days up to the date 12 months after the first participant was recruited at that practice) to determine:

1) The proportion of adult patients registered with the practice, aged ≥18 years, excluding those with both a coded cancer diagnosis and a code for palliative care, who have been prescribed a weak, intermediate or strong opioid analgesic, based on a published categorisation for prescribed analgesics in primary care.[57]
2) The proportion of adult patients registered with the practice, aged ≥18 years, excluding those with both a coded cancer diagnosis and a code for palliative care, who have been prescribed each of the following classes of medicines: paracetamol, topical pain treatments, non-steroidal anti-inflammatory drugs (NSAIDs), nefopam, gabapentinoids, antidepressants, benzodiazepines and Z-drug hypnotics (zopiclone, zolpidem).

Descriptive statistics and 95% confidence intervals will be used to describe the aggregate data from the pseudonymised electronic health records at baseline and 12-months, to include a comparison of the average proportion of opioid and non-opioid treatment use by treatment arm at 12-months, and to explore within-practice changes in these proportions between baseline and 12-month follow-up. Cluster-level analyses will also be considered that adjust the findings for practice-level variables such as practice size and practice index of multiple deprivation e.g. by fitting a linear regression model to the summary data: outcome proportion at 12-months; adjusted for baseline proportion, treatment arm and practice-level characteristics

5.4 Supplementary/exploratory analysis

5.4.1 Reduction in opioid use and the BPI total score at 12-month follow-up for participants who attended the pain review.

A principal stratum approach will be used for this analysis based on analysis population 2.

Complier average causal effect (CACE) models will be fitted to the data to estimate the difference in the proportion of patients reducing opioids at 12-months, and the mean difference in the total BPI score between participants who attended the pain review, and those participants who would have attended the pain review if they had been randomised to the PROMPPT intervention.

The CACE models will be fitted using the gsem procedure in STATA (Troncoso et al. 2022). The model for reduction in opioid use will be fitted using a logit distribution (family(bernoulli) link(logit)) and results will be presented as an odds ratio and 95% confidence intervals. The model for mean difference in total BPI score will be fitted using a normal distribution (family(gaussian) link(identity)) and results will be presented as mean differences and 95% confidence intervals. Robust standard errors clustered at the GP level will be used to account for the clustered nature of the data in the trial. Initially, the model will be fitted with no predictors of the outcome of interest, and no predictors of pain review attendance. We will then explore whether model fit (as measured using Akaike's information criterion (AIC) and the Bayesian information criterion (BIC)) improves when such predictors are added to the model:

Candidate predictors of reduction in opioids at 12-month follow-up: baseline morphine equivalent dose, average baseline morphine equivalent dose at the GP practice that the participant belongs, and the adjusting covariates listed in section 5.1.2.

Candidate predictors of the mean difference in total BPI score at 12-month follow-up: baseline total BPI score, average baseline total BPI score at the GP practice that the participant belongs, and the adjusting covariates listed in section 5.1.2.

Candidate predictors of attendance at the pain review: baseline morphine equivalent dose, average baseline morphine equivalent dose at the GP practice that the participant belongs, age and pain self-efficacy

Currently, we have found little guidance in the literature on how to fit gsem models to data after multiple imputation has been applied, hence, our analysis will be applied to data prior to multiple imputation being performed to impute the missing data.

We will report the results from the models with the lowest AIC and BIC values to represent the models that are the best fit to the data (Table 12.2.2). We will consider these findings noting that for the non-inferiority hypothesis the CACE estimate is the most conservative approach for analysis i.e. the analysis that is less likely to reject the null hypothesis.

5.4.2 Reduction in opioid use at 12-month follow-up: relationship with key participant characteristics of interest

We will explore whether the findings from the primary analysis of reduction in opioid use at 12-month follow-up depends on baseline levels of pain self-efficacy (as measure by the PSEQ) and baseline levels of opioid use (MED). This will be achieved by including the covariate of interest into the primary model for the opioid reduction outcome at 12-month follow-up (if it is not already in the model), along with an interaction term between treatment and the covariate of interest. Two revised models will therefore be produced, one for each covariate. The beta coefficients and 95% confidence intervals for the interaction terms will be reported Table 12.3.4. Graphical methods will be used to display the nature of any significant interaction found in the data ($p<0.05$), using the “marginsplot” command in STATA.

Furthermore, although not in our original protocol, we also plan to explore whether the reduction in opioid use at 12-months depends on whether the participant is taking a gabapentinoid alongside their opioids. The rationale for this additional analysis to inform a future study that may explore whether the PROMPPT intervention could be adapted to be suitable for patients on gabapentinoids. We will use the same method as described above to explore this in a subgroup analysis.

6 Safety

The process for reporting adverse events is described in the study protocol. The number and percentage of participants experiencing a serious and unexpected adverse event (SUAE) that was related to the PROMPPT intervention will be reported (Table 12.3.6). Percentages will be calculated from the number of participants randomised into the trial who attended the pain review i.e., defining a "safety population". Details of each event will be described in text or table as appropriate. We do not anticipate that many participants will experience multiple SUAE, but if they do, then the number of SUAEs that each person experienced will be reported as a percentage of participants experiencing one or more SUAEs.

The percentage of adverse events will be evaluated descriptively and assessed for clinical significance.

7 Additional analysis plans

Separate analysis plans will be written to describe the health economics analysis and the process evaluation. The process evaluation will use a mixed methods approach to describe (i) how PROMPPT was delivered and received, the quantity and quality of what was delivered (fidelity and dose) (ii) mechanisms of impact (How did PROMPPT produce change in opioid use?) (iii) contextual factors (How does context affect implementation of PROMPPT and outcomes?). The quantitative aspects of the process evaluation will use descriptive statistics only to describe the data (numbers and percentages, means and standard deviations, medians and interquartile ranges as appropriate) (outlined in Table 12.3.7, Table 12.3.8 and Table 12.3.9) and will be triangulated with the findings from the qualitative study to draw overall conclusions and recommendations for the potential implementation of the PROMPPT intervention into clinical practice.

8 Software

Analysis in this analysis plan will be generated using STATA software and will use the most up-to-date version of the software available for analysis. The software version number will be reported in any published papers arising from the trial.

9 Data management plan

Trial data collection followed Data Management Plan (DMP) version 0.1 – 03/04/24.

10 References

Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol*. 2004;57(8):785-794. doi: 10.1016/j.jclinepi.2003.12.013.

Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: A tutorial on multiple imputation. Can J Cardiol. 2021;37(9):1322-1331. doi: S0828-282X(20)31111-9 [pii].

Austin PC, van Buuren S. Logistic regression vs. predictive mean matching for imputing binary covariates. Stat Methods Med Res. 2023 Nov;32(11):2172-2183. doi: 10.1177/09622802231198795. Epub 2023 Sep 26. PMID: 37750213; PMCID: PMC10683343.

Bell, M.L., Rabe, B.A. The mixed model for repeated measures for cluster randomized trials: a simulation study investigating bias and type I error with missing continuous data. *Trials* **21**, 148 (2020). https://doi.org/10.1186/s13063-020-4114-9

Billot L, Copas A, Leyrat C, Forbes A, Turner EL. How should a cluster randomized trial be analyzed? J Epidemiol Popul Health. 2024 Feb;72(1):202196. doi: 10.1016/j.jeph.2024.202196. Epub 2024 Feb 10. PMID: 38477477; PMCID: PMC7616648.

Clark TP, Kahan BC, Phillips A, White I, Carpenter JR. Estimands: Bringing clarity and focus to research questions in clinical trials. BMJ Open. 2022;12(1):e052953. http://bmjopen.bmj.com/content/12/1/e052953.abstract. doi: 10.1136/bmjopen-2021-052953.

Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: Effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol*. 2006;35(5):1292-1300. doi: 10.1093/ije/dyl129.

Gillespie D, Francis NA, Carrol ED, Thomas-Jones E, Butler CC, Hood K. Use of co-primary outcomes for trials of antimicrobial stewardship interventions. *Lancet Infect Dis*. 2018;18(6):595-597. doi: 10.1016/S1473-3099(18)30289-5.

Hayati P, Lee K, Simpson J. Sensitivity analysis within multiple imputation framework using delta-adjustment: Application to longitudinal study of australian children. *Longitudinal and Life Course Studies*. 2018;9:259-278. doi: 10.14301/llcs.v9i3.503.

Hooper R, Forbes A, Hemming K, Takeda A, Beresford L. Analysis of cluster randomised trials with an assessment of outcome at baseline. *Bmj-British Medical Journal*. 2018;360:k1121. doi: 10.1136/bmj.k1121.

Huque MH, Carlin JB, Simpson JA, Lee KJ. A comparison of multiple imputation methods for missing data in longitudinal studies. BMC Medical Research Methodology. 2018;18(1):168. https://doi.org/10.1186/s12874-018-0615-6. doi: 10.1186/s12874-018-0615-6.

Joint Formulary Committee. British National Formulary [76] ed. London: BMJ Group and Pharmaceutical Press; 2018.

Kahan BC, Li F, Copas AJ, Harhay MO. Estimands in cluster-randomized trials: Choosing analyses that answer the right question. *Int J Epidemiol*. 2022;52(1):107-118. https://doi.org/10.1093/ije/dyac131. Accessed 2/3/2024. doi: 10.1093/ije/dyac131.

Kean J, Monahan PO, Kroenke K, et al. Comparative responsiveness of the PROMIS pain interference short forms, brief pain inventory, PEG, and SF-36 bodily pain subscale. *Med Care*. 2016;54(4):414-421. doi: 10.1097/MLR.0000000000000497.

Kleinke K. Multiple imputation by predictive mean matching when sample size is small. Methodology. 2018;14(1):3-15. https://doi.org/10.1027/1614-2241/a000141. doi: 10.1027/1614-2241/a000141.

Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Med Care*. 2010;48(11):1007-1014. doi: 10.1097/MLR.0b013e3181eaf835.

Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*. 2009;114(1-3):163-173. doi: 10.1016/j.jad.2008.06.026.

Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: Which analyses should be used? *Int J Epidemiol*. 2017;47(1):321-331. https://doi.org/10.1093/ije/dyx169. Accessed 1/24/2024. doi: 10.1093/ije/dyx169.

Mallen CD, Nicholl BI, Lewis M, et al. The effects of implementing a point-of-care electronic template to prompt routine anxiety and depression screening in patients consulting for osteoarthritis (the primary care osteoarthritis trial): A cluster randomised trial in primary care. *PLoS Med*. 2017;14(4):e1002273. doi: 10.1371/journal.pmed.1002273.

Marion K Campbell, Gilda Piaggio, Diana R Elbourne, Douglas G Altman. Consort 2010 statement: Extension to cluster randomised trials. *BMJ*. 2012;345:e5661. http://www.bmj.com/content/345/bmj.e5661.abstract. doi: 10.1136/bmj.e5661.

Monsalves, M.J., Bangdiwala, A.S., Thabane, A. *et al.* LEVEL (Logical Explanations & Visualizations of Estimates in Linear mixed models): recommendations for reporting multilevel data and analyses. *BMC Med Res Methodol* **20**, 3 (2020). https://doi.org/10.1186/s12874-019-0876-8

Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Medical Research Methodology. 2014;14(1):75. https://doi.org/10.1186/1471-2288-14-75. doi: 10.1186/1471-2288-14-75.

Morris TP, Walker AS, Williamson EJ, White IR. Planning a method for covariate adjustment in individually randomised trials: A practical guide. *Trials*. 2022;23(1):328. https://doi.org/10.1186/s13063-022-06097-z. doi: 10.1186/s13063-022-06097-z.

Nguyen CD, Carlin JB, Lee KJ. Practical strategies for handling breakdown of multiple imputation procedures. Emerging Themes in Epidemiology. 2021;18(1):5. https://doi.org/10.1186/s12982-021-00095-3. doi: 10.1186/s12982-021-00095-3.

Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. *Eur J Pain*. 2007;11(2):153-163. doi: 10.1016/j.ejpain.2005.12.008.

Offen W, Chuang-Stein C, Dmitrienko A, et al. Multiple co-primary endpoints: Medical and statistical solutions: A report from the multiple endpoints expert team of the pharmaceutical research and manufacturers of america. *Drug information journal : DIJ / Drug Information Association*. 2007;41(1):31-46. https://doi.org/10.1177/009286150704100105. doi: 10.1177/009286150704100105.

Pedroza, C., Truong, V.T. Performance of models for estimating absolute risk difference in multicenter trials with binary outcome. *BMC Med Res Methodol* **16**, 113 (2016). https://doi.org/10.1186/s12874-016-0217-0

Rombach I, Gray AM, Jenkinson C, Murray DW, Rivero-Arias O. Multiple imputation for patient reported outcome measures in randomised controlled trials: Advantages and disadvantages of imputing at the item, subscale or composite score level. *BMC Med Res Methodol*. 2018;18(1):87-6. doi: 10.1186/s12874-018-0542-6.

Rubin D, Schenker N. Multiple imputation in health-are databases: An overview and some applications. Stat Med 1991; 10:585–98.

Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *Int J Epidemiol*. 2015;44(3):1051-1067. doi: 10.1093/ije/dyv113.

Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732. doi: 10.7326/0003-4819-152-11-201006010-00232.

Schumi J, Wittes JT. Through the looking glass: Understanding non-inferiority. *Trials*. 2011;12(1):106. https://doi.org/10.1186/1745-6215-12-106. doi: 10.1186/1745-6215-12-106.

Singer, J., & Willet, J. (2003). Applied longitudinal data analysis (First ed.). USA: Oxford University Press.

Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med*. 2006;166(10):1092-1097. doi: 10.1001/archinte.166.10.1092.

StataCorp. 2022. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.

STATA: https://www.stata.com/support/faqs/statistics/clustering-and-mi-impute/. Last accessed 22/1/2024

Stuart B, Becque T, Moore M, Little P. Clustering of continuous and binary outcomes at the general practice level in individually randomised studies in primary care - a review of 10 years of primary care trials. *BMC Med Res Methodol*. 2020;20(1):83-7. doi: 10.1186/s12874-020-00971-7.

Turner E, et al. Completeness of reporting and risks of overstating impact in cluster randomised trials: a systematic review. The Lancet Global Health (2021), Volume 9, Issue 8, e1163 - e1168

Twisk J; Bosman L; Hoekstra T; Rijnhart J; Welten M and Heymans M. Different ways to estimate treatment effects in randomised controlled trials. Contemporary Clinical Trials Communications. 2018;10:80-85. https://www.sciencedirect.com/science/article/pii/S2451865417301849. doi: 10.1016/j.conctc.2018.03.008.

van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007;16(3):219-242. https://doi.org/10.1177/0962280206074463. doi: 10.1177/0962280206074463.

Von Korff M, Saunders K, Thomas Ray G, et al. De facto long-term opioid therapy for noncancer pain. *Clin J Pain*. 2008;24(6):521-527. doi: 10.1097/AJP.0b013e318169d03b.

Wen L, Terrera GM, Seaman SR. Methods for handling longitudinal outcome processes truncated by dropout and death. *Biostatistics*. 2017;19(4):407-425. https://doi.org/10.1093/biostatistics/kxx045. Accessed 4/22/2023. doi: 10.1093/biostatistics/kxx045.

White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-399. doi: 10.1002/sim.4067 [doi].

11 Acknowledgments

We acknowledge that parts of this analysis plan have been adapted from the analysis plan that was written for the PROP-OA trial (https://www.isrctn.com/ISRCTN28555470).

12 Outline tables and figures

12.1 Clinical effectiveness paper – primary figures

Figure 12.1.1: Cluster-level recruitment

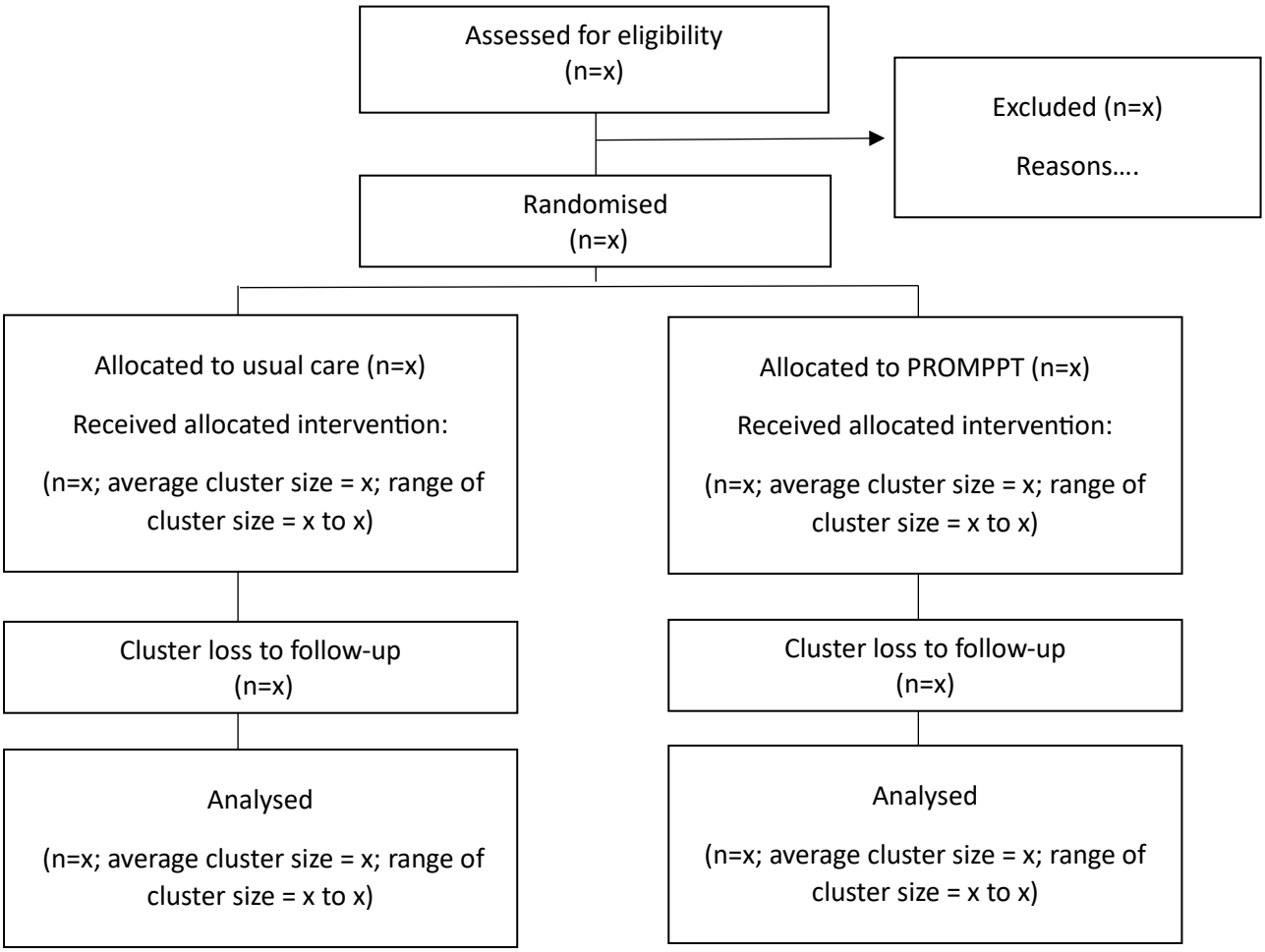
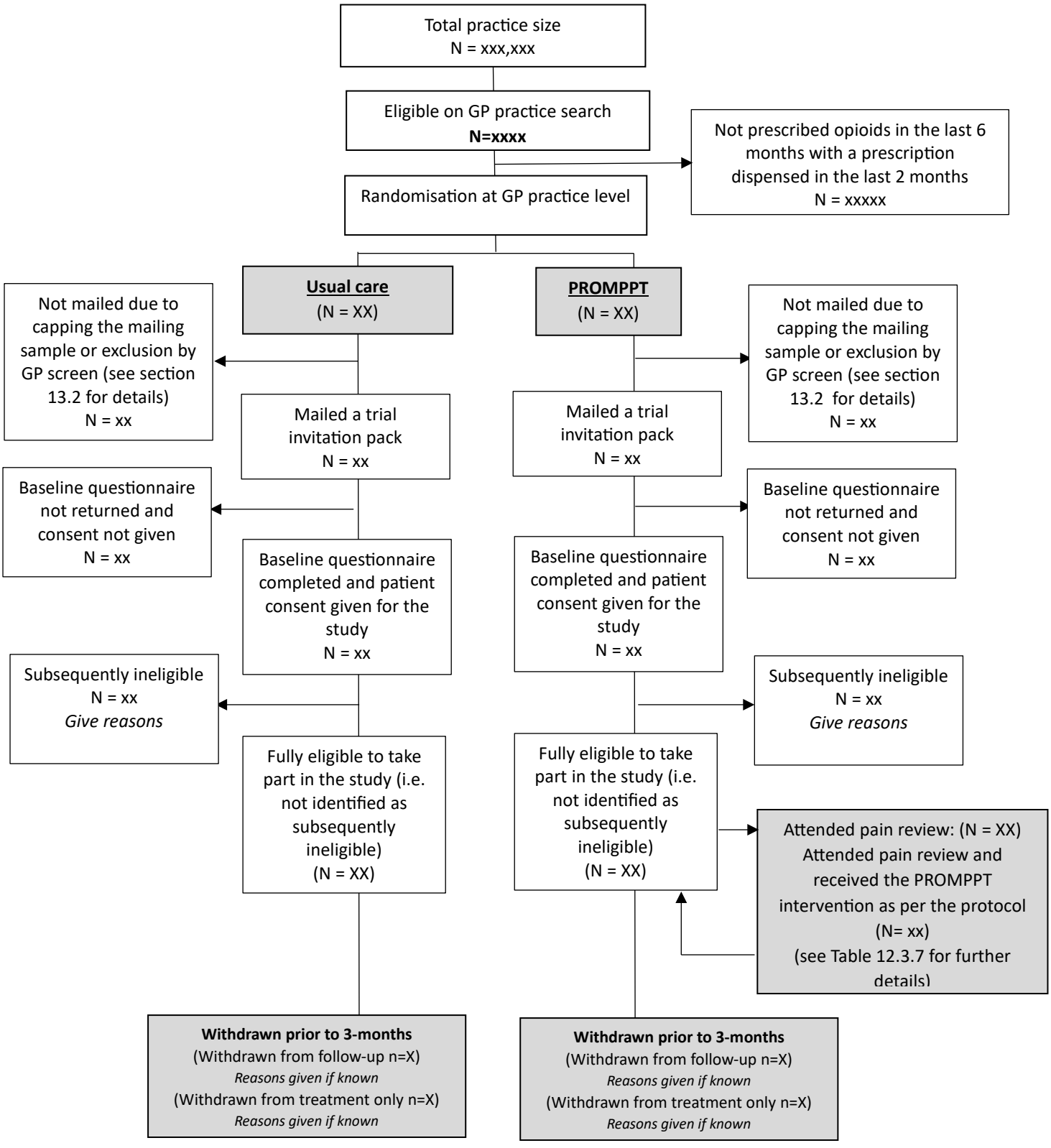
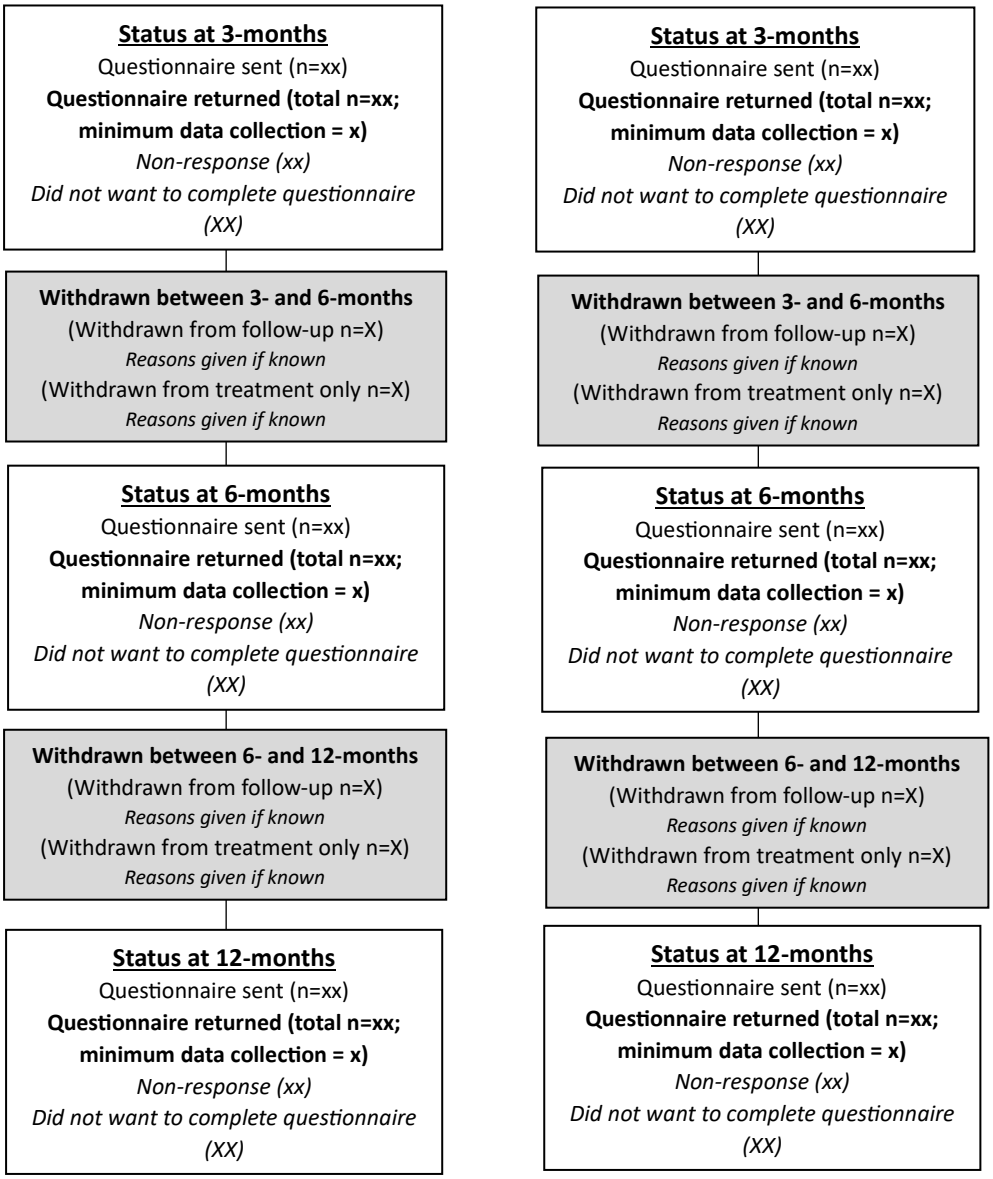


Figure 12.1.2: Participant recruitment and follow-up




Status at 3-months
Questionnaire sent (n=xx)
Questionnaire returned (total n=xx; minimum data collection = x)
Non-response (xx)
Did not want to complete questionnaire (XX)
Withdrawn between 3- and 6-months
(Withdrawn from follow-up n=X)
Reasons given if known
(Withdrawn from treatment only n=X)
Reasons given if known
Status at 6-months
Questionnaire sent (n=xx)
Questionnaire returned (total n=xx; minimum data collection = x)
Non-response (xx)
Did not want to complete questionnaire (XX)
Withdrawn between 6- and 12-months
(Withdrawn from follow-up n=X)
Reasons given if known
(Withdrawn from treatment only n=X)
Reasons given if known
Status at 12-months
Questionnaire sent (n=xx)
Questionnaire returned (total n=xx; minimum data collection = x)
Non-response (xx)
Did not want to complete questionnaire (XX)
Status at 3-months
Questionnaire sent (n=xx)
Questionnaire returned (total n=xx; minimum data collection = x)
Non-response (xx)
Did not want to complete questionnaire (XX)
Withdrawn between 3- and 6-months
(Withdrawn from follow-up n=X)
Reasons given if known
(Withdrawn from treatment only n=X)
Reasons given if known
Status at 6-months
Questionnaire sent (n=xx)
Questionnaire returned (total n=xx; minimum data collection = x)
Non-response (xx)
Did not want to complete questionnaire (XX)
Withdrawn between 6- and 12-months
(Withdrawn from follow-up n=X)
Reasons given if known
(Withdrawn from treatment only n=X)
Reasons given if known
Status at 12-months
Questionnaire sent (n=xx)
Questionnaire returned (total n=xx; minimum data collection = x)
Non-response (xx)
Did not want to complete questionnaire (XX)

12.2 Clinical effectiveness paper – primary tables

Table 12.2.1: Key baseline characteristics at the cluster and patient level

Characteristic	Usual Care N=XXX	PROMPPT N=XXX	All participants N = xxx
Cluster-level characteristics			
Practice size (number of registered patients): Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Practice index of multiple deprivation: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Practice rurality			
Most urban	xx (xx)	xx (xx)	xx (xx)
Very urban	xx (xx)	xx (xx)	xx (xx)
Urban	xx (xx)	xx (xx)	xx (xx)
Rural	xx (xx)	xx (xx)	xx (xx)
Most rural	xx (xx)	xx (xx)	xx (xx)
Patient-level characteristics			
Age (years): mean (SD)	xx (xx)	xx (xx)	xx (xx)
Opioid group			
Weak	xx (xx)	xx (xx)	xx (xx)
Intermediate	xx (xx)	xx (xx)	xx (xx)
Strong	xx (xx)	xx (xx)	xx (xx)
Female sex	xx (xx)	xx (xx)	xx (xx)
White ethnicity	xx (xx)	xx (xx)	xx (xx)
Currently in a paid job	xx (xx)	xx (xx)	xx (xx)
Index of multiple deprivation (IMD) (1 - 32,844): mean (SD)	xx (xx)	xx (xx)	xx (xx)
Length of time with persistent pain			
< 1 year	xx (xx)	xx (xx)	xx (xx)
1-2 years	xx (xx)	xx (xx)	xx (xx)
3-5 years	xx (xx)	xx (xx)	xx (xx)
6-10 years	xx (xx)	xx (xx)	xx (xx)
> 10 years	xx (xx)	xx (xx)	xx (xx)

Morphine Equivalent Dose: mean (SD)	xx (xx)	xx (xx)	xx (xx)
Morphine Equivalent Dose: median (IQR)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Brief Pain Inventory: pain severity (0-10): mean (SD)	xx (xx)	xx (xx)	xx (xx)
Brief Pain Inventory: pain interference (0-10): mean (SD)	xx (xx)	xx (xx)	xx (xx)
Pain Self-Efficacy Questionnaire (0-60): mean (SD)	xx (xx)	xx (xx)	xx (xx)
Depression: PHQ-8: (0-24): mean (SD)	xx (xx)	xx (xx)	xx (xx)
Anxiety: GAD-7 (0-21): median (interquartile range)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)

Figures are numbers and percentages, unless otherwise stated. SD = standard deviation. Index of multiple deprivation: higher score = less deprived, Brief pain inventory: higher score = worse pain severity, greater pain interference; Pain self-efficacy: higher score = greater self-efficacy for pain; PHQ-8: higher score, more depressed; GAD-7: higher score, more anxious.

Table 12.2.2: Treatment effect estimates for the primary analysis at the 12-month follow-up

Primary analysis	12-months
Co-primary outcome 1 – reduction in morphine equivalent dose (yes/no)	
Descriptive statistics	
Usual care: Number (%)	xx(xx)
PROMPPT: Number (%)	xx(xx)
Treatment effect: Usual care vs PROMPPT: Adjusted$^{\alpha}$ odds ratio (95% CI)	xx (xx, xx)
Treatment effect: Usual care vs PROMPPT: Adjusted$^{\alpha}$ absolute risk difference (95% CI)	xx (xx, xx)
ICC (95% CI)	xx (xx, xx)
ICC - Usual care (95% CI)	xx (xx, xx)
ICC – PROMPPT (95% CI)	xx (xx, xx)
AIC	xx
BIC	xx
Co-primary outcome 2 – BPI total score (continuous measure)	
Descriptive statistics	
Usual care: Mean (SD)	xx(xx)
PROMPPT: Mean (SD)	xx(xx)
Treatment effect: Usual care vs PROMPPT: Adjusted$^{\alpha}$ mean difference (95% CI)	xx (xx, xx)
ICC (95% CI)	xx (xx, xx)
ICC - Usual care (95% CI)	xx (xx, xx)
ICC – PROMPPT (95% CI)	xx (xx, xx)
AIC	xx
BIC	xx
Sensitivity analysis – Exploring the impact of the data not being missing at random	
Co-primary outcome 1 – reduction in morphine equivalent dose (yes/no)	
Treatment effect: Usual care vs PROMPPT: Adjusted$^{\alpha}$ odds ratio (95% CI)	
Delta = x	xx (xx, xx)
Delta = x	xx (xx, xx)
Delta = x	xx (xx, xx)
Delta = x	xx (xx, xx)
Co-primary outcome 2 – BPI total score (continuous measure)	

Treatment effect: Usual care vs PROMPPT: Adjusted[α] mean difference (95% CI)	
Delta = x	xx (xx, xx)
Delta = x	xx (xx, xx)
Delta = x	xx (xx, xx)
Delta = x	xx (xx, xx)
Supplementary analyses: complier average causal effect (CACE) analysis for the co-primary outcomes and for the principal stratum of participants who attend the pain review	
Treatment effect: Usual care vs PROMPPT: Adjusted[α] odds ratio: reduction in opioid use (95% CI)	xx (xx, xx)
Treatment effect: Usual care vs PROMPPT: Adjusted[α] mean difference: BPI total score (95% CI)	xx (xx, xx)

[α] Adjusted for region, age, sex and the patient-level index of multiple deprivation, the baseline individual measure and baseline cluster mean in the relevant outcome of interest. AIC = Akaike information criteria, BIC = Bayesian information criteria, CI = confidence interval, ICC = intra-cluster correlation coefficient, BPI = Brief Pain Inventory, SD = Standard deviation

Table 12.2.3: Treatment effect estimates for the secondary outcomes measured on a continuous scale

Outcome	3-months	6-months	12-months
Morphine Equivalent Dose			
Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Percentage change in Morphine Equivalent Dose (baseline to 12-month follow-up)			
Usual primary care: Mean (SD)	N/A	N/A	xx (xx)
PROMPPT intervention: Mean (SD)	N/A	N/A	xx (xx)
Adjusted$^{\beta}$ mean difference (estimate and 95% CI)	N/A	N/A	xx (xx, xx)
ICC$^{\gamma}$: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
BPI total score			
Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
BPI pain severity			
Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
BPI pain interference			

Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Pain Self-Efficacy Questionnaire: PSEQ			
Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Depression: PHQ-8			
Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Anxiety: GAD-7			
Usual primary care: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: Mean (SD)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] mean difference (estimate and 95% CI)	xx (xx)	xx (xx)	xx (xx)
Interaction term between treatment and time (estimate and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)

[α] Adjusted for region, age, sex and the patient-level index of multiple deprivation, the baseline individual measure and baseline cluster mean in the relevant outcome of interest [β] Adjusted for region, age, sex and the patient-level index of multiple deprivation. [γ] correlation among all the values between and within participants nested in GP practices. [δ] correlation among the repeated measures within participants nested within GP practices. BPI = Brief pain inventory; PHQ Patient Health Questionnaire; GAD Generalised Anxiety Disorder Assessment; PSEQ Pain self-efficacy questionnaire

Table 12.2.4: Treatment effect estimates for the secondary outcomes measured on a binary scale (i.e. non-opioid pain medicine use)

Outcome	3-months	6-months	12-months
Non-opioid pain medicines used for pain			
Paracetamol			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Topical treatments			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Nefopam			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Systematic NSAIDs			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)

Gabapentinoids			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Antidepressants			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Benzodiazepines			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Z-drugs			
Usual primary care: n (%)	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)

[α] Adjusted for region, age, sex and the patient-level index of multiple deprivation. [γ] correlation among all the values between and within participants nested in GP practices. [δ] correlation among the repeated measures within participants nested within GP practices

Table 12.2.5: Treatment effect estimates for the secondary outcomes measured on an ordinal scale (i.e. medication side effects)

Outcome	3-months	6-months	12-months
Medication side effects			
Constipation			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Itching			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)

ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Daytime sleepiness			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Dry mouth			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)

Nausea (feeling sick)			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Vomiting			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Dizziness			
Usual primary care: n (%)			

None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Headache			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted$^{\alpha}$ odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC$^{\gamma}$: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC$^{\delta}$: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Confusion			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)

Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Difficulty concentrating			
Usual primary care: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
PROMPPT intervention: n (%)			
None	xx (xx)	xx (xx)	xx (xx)
Slight (mild)	xx (xx)	xx (xx)	xx (xx)
Moderate	xx (xx)	xx (xx)	xx (xx)
Severe	xx (xx)	xx (xx)	xx (xx)
Very Severe	xx (xx)	xx (xx)	xx (xx)
Adjusted[α] odds ratio (estimate and 95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
Interaction term between treatment and time (odds ratio and 95% CI)	0 (ref)	xx (xx, xx)	xx (xx, xx)
ICC[γ]: GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
ICC[δ]: participants within GP practice (95% CI)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)

[α] Adjusted for region, age, sex and the patient-level index of multiple deprivation. [γ] correlation among all the values between and within participants nested in GP practices. [δ] correlation among the repeated measures within participants nested within GP practices

12.3 Clinical effectiveness paper – supplementary tables

Table 12.3.1: Additional baseline characteristics

Outcome		Usual care	PROMPPT	All participants
Non-opioid pain medicines used for pain				
	Paracetamol	xx (xx)	xx (xx)	xx (xx)
	Topical treatments	xx (xx)	xx (xx)	xx (xx)
	Nefopam	xx (xx)	xx (xx)	xx (xx)
	Systematic NSAIDs	xx (xx)	xx (xx)	xx (xx)
	Gabapentinoids	xx (xx)	xx (xx)	xx (xx)
	Antidepressants	xx (xx)	xx (xx)	xx (xx)
	Benzodiazepines	xx (xx)	xx (xx)	xx (xx)
	Z-drugs	xx (xx)	xx (xx)	xx (xx)
Medication side effects				
Constipation				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Itching				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Daytime sleepiness				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Dry mouth				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Nausea (feeling sick)				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Vomiting				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)

	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Dizziness				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Headache				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Confusion				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)
Difficulty concentrating				
	None	xx (xx)	xx (xx)	xx (xx)
	Slight (mild)	xx (xx)	xx (xx)	xx (xx)
	Moderate	xx (xx)	xx (xx)	xx (xx)
	Severe	xx (xx)	xx (xx)	xx (xx)
	Very Severe	xx (xx)	xx (xx)	xx (xx)

Figures are numbers (percentages)

Table 12.3.2: Key baseline characteristics by loss to follow-up

<<The patient-level data in Table 12.2.1 will be copied, but stratified by whether the participant has returned a 12-month follow-up questionnaire, rather than by treatment arm>>

Table 12.3.3: Missing data rates for the co-primary outcomes

	N(%) of missing data			
	Baseline N=XX	3-months N=XX	6-months N=XX	12-months N=XX
Morphine equivalent dose (MED) reduction (yes/no)				xx (xx)
Morphine equivalent dose (MED)	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Brief pain inventory (BPI) total score	xx (xx)	xx (xx)	xx (xx)	xx (xx)

Missing data rates for the remaining secondary outcome measures will not be presented in tabular format but will be reported in the text of the paper with wording such as: “Missing data rates for the secondary outcomes was less than x% at all follow-up time-points”.

Table 12.3.4: Exploratory subgroup analyses for the co-primary outcome of opioid reduction at 12-month follow-up

	Interaction (95% CI)
Baseline morphine equivalent dose	xx (xx, xx)
Pain Self-Efficacy Questionnaire: PSEQ	xx (xx, xx)
Taking gabapentinoids at baseline	xx (xx, xx)

CI = confidence interval

Table 12.3.5: Protocol deviations

Deviation Number	Deviation	How many participants affected	Treatment Arm
1	xxxx	xx	xx
2	xxxx	xx	xx
etc	xxxx	xx	xx

If this table is very long, we will add a footnote to say that only deviations that impact of primary and secondary data collection are listed.

Table 12.3.6: Serious and unexpected adverse events related to the PROMPPT intervention.

Date of adverse event onset (if known)	Date of report	Description
xx/xx/xx	xx/xx/xx	xx
xx/xx/xx	xx/xx/xx	xx
etc		

Table 12.3.7: Intervention fidelity

Mandatory pain review component	Number of times component recorded as completed (N = xxx CRFs available)
1. Information gathered about the patient's experience of living with persistent pain	xx (xx)
2. Information gathered about how the patient takes their opioids and other pain medicines	xx (xx)
3. Information gathered about the patient's personal upsides and downsides of taking and reducing opioids	xx (xx)
4. Assessment made on the patient's perspective on change	xx (xx)
5. Pain action plan created with the patient	xx (xx)
6. 'Your Pain Action Plan' completed and given or sent to the patient	xx (xx)
7. 'My Pain Review: A Positive Change' leaflet given or sent to the patient	xx (xx)
8. Further contact arrangements discussed and agreed	xx (xx)
All mandatory pain review components delivered	**xx (xx)**

Footnote: CRF = case report form

Table 12.3.8: Acceptability of the PROMPPT intervention

Theoretical Framework of Acceptability Questionnaire	N (%)
How acceptable was the pain management review	
Completely unacceptable	xx (xx)
Unacceptable	xx (xx)
No opinion	xx (xx)
Acceptable	xx (xx)
Completely acceptable	xx (xx)
Did you like or dislike the pain management review	
Strongly dislike	xx (xx)
Dislike	xx (xx)
No opinion	xx (xx)
Like	xx (xx)
Strongly like	xx (xx)
How much effort did it take to participate in the pain management review	
No effort at all	xx (xx)
A little effort	xx (xx)
No opinion	xx (xx)
A lot of effort	xx (xx)
Huge effort	xx (xx)
How fair (to all patients) is a system where patients with long-term pain are invited for a routine pain management review	
Very unfair	xx (xx)
Unfair	xx (xx)
No opinion	xx (xx)
Fair	xx (xx)
Very fair	xx (xx)
The pain management review is likely to change how I manage my pain	
Strongly disagree	xx (xx)
Disagree	xx (xx)
No opinion	xx (xx)
Agree	xx (xx)
Strongly agree	xx (xx)
It is clear to me how the pain management review I attended with the clinical pharmacist will help me manage my pain	
Strongly disagree	xx (xx)
Disagree	xx (xx)
No opinion	xx (xx)
Agree	xx (xx)
Strongly agree	xx (xx)
How confident would you feel about making changes to how you manage your pain	
Very unconfident	xx (xx)
Unconfident	xx (xx)
No opinion	xx (xx)
Confident	xx (xx)
Very confident	xx (xx)

Making changes to how I manage my pain will interfere with my other priorities	
Strongly disagree	xx (xx)
Disagree	xx (xx)
No opinion	xx (xx)
Agree	xx (xx)
Strongly agree	xx (xx)

Table 12.3.9: Treatment acceptability and credibility measure

	Median (interquartile range)
How logical does this type of review seem to you? (0-10)	x (x, x)
How successful do you think the review will be in changing how you manage your pain? (0-10)	x (x, x)
How confident would you be in recommending this review to a friend (0-10)	x (x, x)
How much improvement in your ability to manage pain do you think will occur (0-10)	x (x, x)
Median (interquartile range) across the four acceptability/credibility questions	x (x, x)

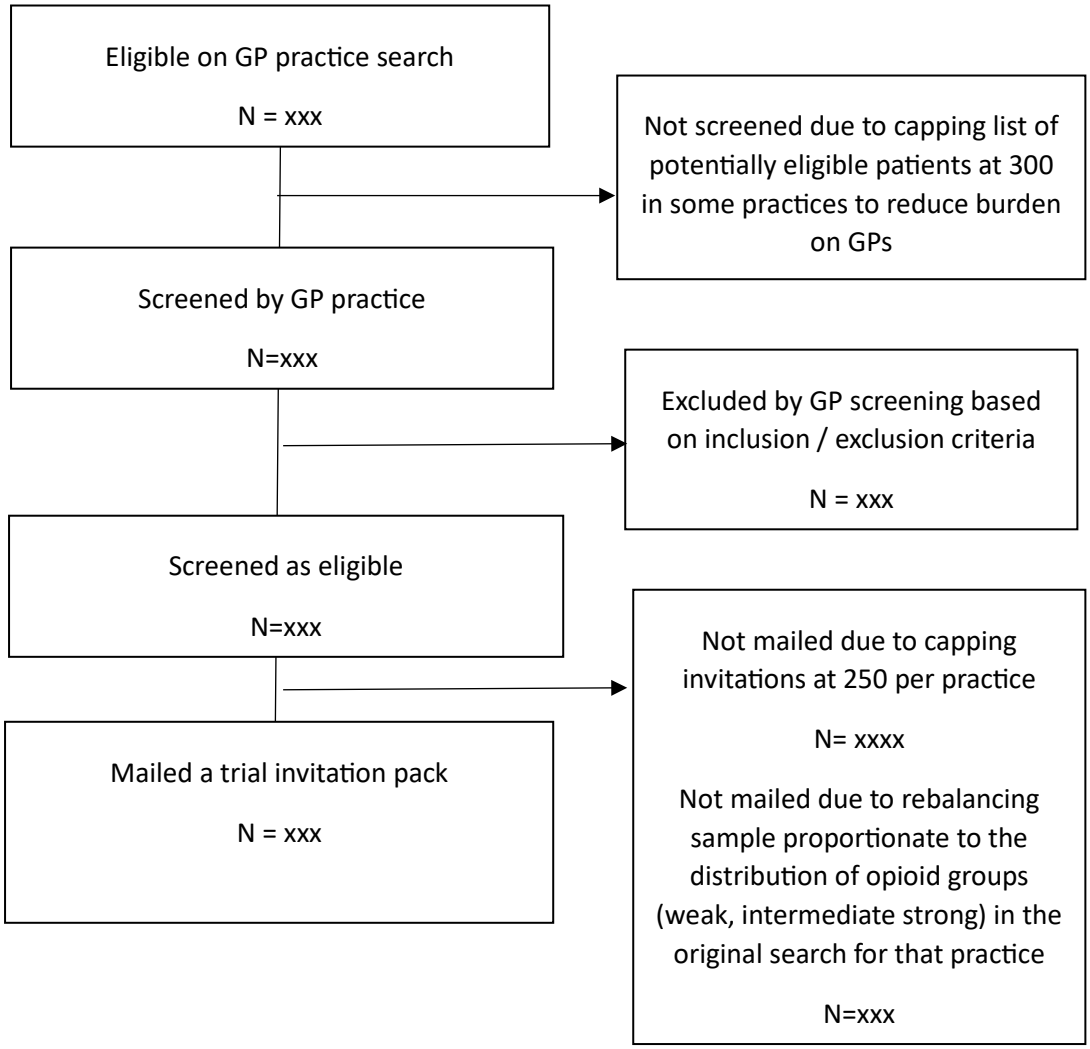
13 Appendices

13.1 Data coding rules applied prior to analyses

Table 13.1.1: Data coding rules applied prior to analyses

Participant Study Identification Number	Data source and question number	Issue	Principle Applied to the data	Action Required
xx	xx	xx	xx	xx
xx	xx	xx	xx	xx
xx	xx	xx	xx	xx
xx	xx	xx	xx	xx
xx	xx	xx	xx	xx
xx	xx	xx	xx	xx

13.2 Additional flow chart to show the impact of capping the GP patient list sizes, and GP screening, on patient recruitment



13.3 Pre-planned adaptations to the imputation strategy

Numerical issues, failure, and breakdown of the multiple imputation algorithm can arise, particularly when there are many variables to include in the imputation model (Nguyen et al. 2021). If this does arise, we plan to use the strategy below (sequentially) to explore how the imputation model can be adapted to ensure that it can be applied to the data.

13.3.1 Perfect prediction

Perfect prediction can arise when multiple categorical variables are included in the imputation model. This would be addressed by adding the STATA "augment" option to the imputation model – a procedure that works by adding in additional "pseudo-observations" to prevent the outcome being perfectly predicted (Nguyen et al. 2021).

13.3.2 Number of nearest neighbours (k) in the predictive mean matching (PMM) models

Kleinke 2018, highlight that there is a trade-off when considering the number of nearest neighbours (k) to include in the PMM model: if k is too small a single participant's data could be repeatedly chosen as a donor in the imputation model, which would underestimate model standard errors, whereas if k is too large, this might result in inadequate donors and implausible imputations, hence biased inferences.

We have used the recommendation by Morris et al. 2014 to set the value of k in the imputation model to be 10. If this decision means that the imputation model breaks down when we fit it to our data, we will re-run the imputation model, firstly with k=5 and then secondly with k = 15 to see if these changes enable the imputation model to run in our data. We will try K=5, before K=15, as the former is preferred default value for K used in the alternative statistical software packages of SAS and R (Kleinke 2018).

13.3.3 Collinearity

We have chosen to include all three BPI scores in our imputation model to reduce bias (BPI pain, BPI pain interference, BPI total score). However, this could lead to a breakdown of the model as the BPI total score is derived as a direct transformation of the BPI pain, and BPI pain interference, subscales. If the model does not run for this reason, we would exclude the BPI total score from the imputation model and use the "mi passive" procedure in STATA to derive these measures after the imputation had been performed instead. The advantage of this approach is that the derived variables will always be consistent with the subscale scores (e.g. if a participant scored 0 for all BPI subscales it would guarantee that the imputed BPI total score would be 0), which can't always be assumed under the primary approach. However, this was not chosen as the primary approach as reducing bias in treatment effect estimates was considered a greater priority.

13.3.4 Ordinal variables

Ordinal variables can be challenging to include in an imputation model due to the number of categories they contain. If, after inspection of the imputation model, it appears that the reason why the imputation model will not run is due to the inclusion of too many ordinal variables, we will use the STATA "ascontinuous" option for the ordinal variables. This imputes the ordinal outcomes using ordinal regression, but, when these outcomes are included as predictor variables in the imputation model for

other outcomes, they are assumed to be continuous variables, rather than categorical, to reduce the number of degrees of freedom in the imputation model (StataCorp. 2022).

13.3.5 Dropping the "Opioid-related side-effects" questions from the imputation model

The opioid-related side effects questions are measure on a 5-point ordinal scale, with response options ranging from "None" to "very severe", which we know, given their measurement scale could be challenging to include in the imputation model with difficulties experienced through lack of model convergence. If model convergence issues remain after the considerations in 13.3.4 have been explored, we will consider dropping these variables from the imputation model, particularly as the main analysis of these variables will not be on imputed data (they will be analysed using longitudinal mixed ordinal models instead).

13.3.6 Binary variables

If the model fails to converge due to the presence of binary variables in the model, we will consider using predictive mean matching to impute the binary variable. This approach has been shown to give similar results to using logistic regression, but can be fitted with greater computational ease (Austin et al. 2023). If this approach is unsuccessful, we will consider dropping the binary variables from the model (i.e. the questions on non-opioid medication use). These variables are not included in the primary analysis model, so it is not essential to include them in the imputation model.

13.3.7 Adapting the imputation model

We have chosen to use MICE as our imputation method, however, other imputation models exist, such as Multi-Variate Normal Imputation (MVNI), which could be used as an alternative approach (Nguyen et al. 2021). Therefore, if our MICE imputation model is unsuccessful, we will explore changing the imputation method to MVNI, to see if we can successfully impute the data using this method. In addition, as our data are in a repeated measures format (i.e. we have data collected at multiple follow-up time-points), we would also consider whether a fully conditional specification (FCS) two-fold imputation model would be appropriate for our data. This procedure offers greater flexibility to the full imputation approach, as it includes only a subset of data collected within a pre-specified time-window in the imputation model, reducing the number of variables in the imputation model, and making it less likely to breakdown in the data (e.g. a time window of one would include in the imputation model only data collected at time t, t+1 and t-1, where applicable) (Huque et al. 2018).