Supplement 2

A group-based educational intervention for reducing opioid use for chronic pain: a randomized clinical trial

List of Investigators

Sandhu H, Booth K, Furlan A, Shaw J, Carnes D, Taylor SJC, Abraham C, Alleyne S, Balasubramanian S, Betteley L Haywood K, Iglesias Urrutia CP, Krishnan, S, Lall R, Manca A, Mistry D, Newton S, Noyes J, Nichols V, Padfield E, Rahman A, Seers K, Tang NKY, Tysall C, Eldabe S/Underwood M

Table of Contents

List of Tables

List of Figures

Section 1: Study Information

eTable 1: Participant inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
• Provision of written informed consent • Aged 18 years old or above • Using opioids for chronic non-malignant pain • Using strong opioids for at least 3 months • Using strong opioids on most days in the preceding month • Fluent in written and spoken English • Able to attend group sessions • Willingness for GP to be informed of participation	• Regular use of injected opioid drugs • Chronic headache as the dominant painful disorder • Serious mental health problems that preclude participation in a group intervention • Previous entry or randomisation in the present trial • Participation in a clinical trial of an investigational medicinal product in the last 90 days • Pregnant at time of eligibility assessment, or actively trying to become pregnant • People receiving strong opioid for the management of pain due to active malignant disease

Letter of Support, Independent Trial Steering Committee

HTA programme
NETSCC
Alpha House
Chilworth
Southampton

Primary Care Services
Royal South Hants Hospital
Brinton's Terrace
Southampton
SO14 0YG

Tel: 02380 716 528
www.solent.nhs.uk

12/10/2018

Dear Team,

Following on from your last monitoring visit to Warwick we held an interim mini TSC meeting to review plans for recruitment. The outcomes from this were:

- Recruitment has really taken off with the team projected to hit target in November
- They are obliged to continue until December due to the way mailing packs have been sent Out to practices

The team have previously discussed with you the importance of looking at two primary measures - improvement in quality of life and adding in reduction in opioids. This was because reduction in opioids is still an important outcome measure for the NHS even if quality of life does not improve. I understand you were happy for the team to consider this an additional study. However there is a likelihood that, if recruitment goes well, as it has been doing, the team may be able to power a joint primary outcome analysis. Please note that this is all based on predictive figures, although the chances are very good that there will be more patients in the study than required, and enough to power such an analysis. For the I-WOTCH team to do this, there will be a short low cost extension.

Having reviewed the information, the TSC congratulated the team on their recruitment, agreed that the two end points were equally as important and the necessity of over recruitment presents a unique opportunity to do this.

The TSC is therefore in support of the extension request.

We hope that you are minded to support this in what is a trial of international importance and interest.

With kind regards
Dr Cathy Price

Opioid equivalence calculations for primary analysis

Different authorities recommend different values for opioid equivalence. During the study we reviewed these to decide on morphine equivalences for use in the final analysis. These may vary from those set at the start for the study for the purposes of stratification, which were primarily based on recommendations from the Faculty of Pain Medicine. In March 2020 we searched for opioid equivalence tables provided by national bodies, published in English. We included tables focussing on both malignant and non-malignant pain. From these we extracted the conversion factor needed to convert each drug into the equivalent of 1mg of morphine. For transdermal preparations we converted the micrograms of index preparation per hour to milligrams per day to generate our conversion factor. We have assumed that patches are worn for multiples of one day. We used a stepwise process to agree the conversion factor

1. Where possible we selected the modal value of MME
2. Where a range is provided, we used the midpoint of the range when deciding on the modal value the modal value
3. Where two mode values exist, we used the midpoint of the two modes
4. Where no tables included a value for oral equivalence, but data on equivalence of injected medication is provided that would allow calculation of oral equivalence these data were used
5. Where no conversion factors are available for buccal absorption then values for transdermal absorption were used
6. Conversion value agreed column list the mathematical conversion factor to obtain the MME in mg.
7. We also extracted the lowest and highest conversion factor for each drug for planned sensitivity analyses

We identified six eligible sources

- Faculty of Pain Medicine, Royal College of Anaesthetists. https://fpm.ac.uk/opioids-aware-structured-approach-opioid-prescribing/dose-equivalents-and-changing-opioids
- British National Formulary https://bnf.nice.org.uk/guidance/prescribing-in-palliative-care.html
- Faculty of Pain Medicine Australia and New Zealand College of Anaesthetists. https://www.anzca.edu.au/getattachment/6892fb13-47fc-446b-a7a2-11cdfe1c9902/PS01(PM)-(Appendix)-Opioid-Dose-Equivalence-Calculation-Table
- USA Centre for Disease Control https://www.cdc.gov/drugoverdose/prescribing/guideline.html#tabs-2-3
- USA Centres for Medicare and Medicaid Services https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Oral-MME-CFs-vFeb-2018.pdf
- Canada Macmaster National Pain Centre https://www.cmaj.ca/content/cmaj/189/18/E659.full.pdf

All sites last accessed 31 March 2022

No one source included all of the opioid used by trial participants. For no single drug, for which more than one estimate was available did all sources agree on opioid equivalence. For oral pethidine there was no published equivalence. However, the Australian and New Zealand College of Anaesthetists provided a figure for the equivalence of parenteral pethidine compared to parenteral Morphine (0.4mg = 3mg) and also a figure for the equivalence of oral and parenteral morphine (1mg=3mg). If conversion factor between oral and parenteral pethidine is the same as that for oral and parenteral morphine, then the conversion factor is 0.12.

Since we completed this exercise the Faculty of Pain Medicine have revised their recommendations to be consistent with the British National Formulary. If Faculty of Pain Medicine Recommendations are removed, as these are no longer independent guidance, then the only change in values would be that the conversion factor for Tramadol would be 0.1 rather than 0.125. We have not updated the previously finalised opioid equivalence values used in our final analysis in light of this change that we became aware of after the main analyses were complete. At the time we developed these tables, we were expecting our primary analysis to be the mean difference in opioid use. Since our primary analysis is on proportion not using opioids at 12 months any differences in opioid equivalences used will not affect our conclusions. For Final Values see eTable2.

eTable 2: Opioid equivalences for final analyses

	UK RCOA FPM	UK BNF	Australia ANZCA	US CDC	US CMS	Canada IMPS/ National Pain Center	Mid-point Canada range	Mode	Conversion Value Agreed	Minimum	Maximum
Tablets											
Buprenorphine Sub Lingual			40		30			35	35	30	40
Codeine	0.1	0.1	0.13	0.15	0.15	0.1-0.2	0.15	0.15	0.15	0.1	0.2
Dihydrocodeine	0.1	0.1			0.25			0.1	0.1	0.1	0.25
Hydromorphone	7.5	5	5	4	4	5		5	5	4	7.5
Morphine	1	1	1	1	1	1		1	1	1	1
Oxycodone	2	1.5	1.5	1.5	1.5	1.5		1.5	1.5	1.5	2
Pethidine			0.12					0.12	0.12	0.12	0.12
Tapentalol	0.4		0.3	0.4	0.4	0.3-0.4		0.4	0.4	0.3	0.4
Tramadol	0.15	0.1	0.2		0.1	0.1-0.2	0.15	0.1,0.15	0.125	0.1	0.2
Liquids											
Morphine	1	1	1		1			1	1	1	1
Oxycodone	2	1.5	1.5	1.5	1.5	1.5		0.5	1.5	1.5	2
Patches											
Buprenorphine 5	12	12	10		9			12			
Buprenorphine 10	24	24	20		18						
Buprenorphine 20	48	48	40		36						
Buprenorphine mcg/hr	2.4	2.4	2		1.8			2.4	1mcg/hr=2.4mg/day	1.8	2.4
Fentanyl 12	45	30	37.5	30	30			30			
Fentanyl 25	90	60	75	60	60	60-134	97	60			
Fentanyl 37	135	90	112.5	90	90	135-179	157	90			
Fentanyl 50	180	120	150	120	120	180-224	202	120			
Fentanyl 62	225	150	187.5	150	150	225-269	247	150			
Fentanyl 75	270	180	225	180	180	270-314	292	180			
Fentanyl 87	315	210	262.5	210	210	315-359	337	210			
Fentanyl 100	360	240	300	240	240	360-404	382	240			
Fentanyl 300	1120	720	900	720	720			720			
Fentanyl mcg/hr	3.6	2.4	3	2.4	2.4	3.43	3.43	2.4	1mcg/hr=2.4mg/day	2.4	3.6

1 mcg/hr buprenorphine or fentanyl equivalent to 2.4 mg morphine=2.4 mg/day

Range of midpoints of Canadian conversion for buprenorphine according to dose 2.82 to 4.04. Midpoint of midpoints =3.43

Formulations for Tapering App

We developed an App to support the nurses to produce a tapering plan for participants in the intervention arm. Typically, a change of medication or route was not advised. On occasion a change of drug or route was necessary. To support this, we identified morphine equivalent dose of each preparation based on the recommendations of the Faculty of Pain Medicine at the Royal College of Anaesthetists London, current at that time, supplemented data from other sources including summaries of product characteristics (eTable3). The Faculty of Pain Medicine has since updated its recommendations on morphine equivalence. Our full tapering schedule is available on request: harbinder.k.sandhu@warwick.ac.uk and will be made freely available. All participants were encouraged to taper on their own preparation and only those on fentanyl patches had to switch to Morphine Sulphate (MST) when lowest patch dose was reached.

eTable 3: Formulations for opioid tapering App based on evidence available at time of development

Opioid Drug	Dosage Form	Release Type	Dose	Unit	Dose	Unit	Conversion	Morphine
Buprenorphine	Sublingual tablet	Immediate release	200	mcg	0.2	mg	40	8mg
Buprenorphine	Sublingual tablet	Immediate release	400	mcg	0.4	mg	40	16mg
Buprenorphine	Sublingual tablet	Immediate release	1	mg	1	mg	40	40mg
Buprenorphine	Sublingual tablet	Immediate release	2	mg	2	mg	40	80mg
Buprenorphine	Sublingual tablet	Immediate release	4	mg	4	mg	40	160mg
Buprenorphine	Sublingual tablet	Immediate release	6	mg	6	mg	40	240
Buprenorphine	Sublingual tablet	Immediate release	8	mg	8	mg	40	320
Fentanyl	Nasal spray	Immediate release	50	mcg	0.05	mg	167	8.35
Fentanyl	Nasal spray	Immediate release	100	mcg	0.1	mg	167	16.7
Fentanyl	Buccal tablet	Immediate release	100	mcg	0.1	mg	167	16.7
Fentanyl	Sublingual tablet	Immediate release	100	mcg	0.1	mg	167	16.7
Fentanyl	Sublingual tablet	Immediate release	200	mcg	0.2	mg	167	33.4
Fentanyl	Oromucosla lozenges	Immediate release	200	mcg	0.2	mg	167	33.4
Fentanyl	Buccal tablet	Immediate release	200	mcg	0.2	mg	167	33.4
Fentanyl	Nasal spray	Immediate release	200	mcg	0.2	mg	167	33.4
Fentanyl	Sublingual tablet	Immediate release	300	mcg	0.3	mg	167	50.1
Fentanyl	Sublingual tablet	Immediate release	400	mcg	0.4	mg	167	66.8
Fentanyl	Oromucosla lozenges	Immediate release	400	mcg	0.4	mg	167	66.8
Fentanyl	Buccal tablet	Immediate release	400	mcg	0.4	mg	167	66.8

Opioid Drug	Dosage Form	Release Type	Dose	Unit	Dose	Unit	Conversion	Morphine
Fentanyl	Buccal tablet	Immediate release	600	mcg	0.6	mg	167	100.2
Fentanyl	Oromucosla lozenges	Immediate release	600	mcg	0.6	mg	167	100.2
Fentanyl	Sublingual tablet	Immediate release	600	mcg	0.6	mg	167	100.2
Fentanyl	Sublingual tablet	Immediate release	800	mcg	0.8	mg	167	133.6
Fentanyl	Oromucosla lozenges	Immediate release	800	mcg	0.8	mg	167	133.6
Fentanyl	Buccal tablet	Immediate release	800	mcg	0.8	mg	167	133.6
Hydromorphone	Capsule	Immediate release	1.3	mg	1.3	mg	7.69	9.997
Hydromorphone	Capsule	Modified release	2	mg	2	mg	7.69	15.38
Hydromorphone	Capsule	Immediate release	2.6	mg	2.6	mg	7.69	19.994
Hydromorphone	Capsule	Modified release	4	mg	4	mg	7.69	30.76
Hydromorphone	Capsule	Modified release	8	mg	8	mg	7.69	61.52
Hydromorphone	Capsule	Modified release	16	mg	16	mg	7.69	123.04
Hydromorphone	Capsule	Modified release	24	mg	24	mg	7.69	184.56
Meptazinol	Tablet	Immediate release	200	mg	200	mg	33.3333333	6.000000001
Morphine	Tablet/capsule	Modified release	5	mg	5	mg	1	5
Morphine	Tablet/capsule	Modified release	10	mg	10	mg	1	10
Morphine	Tablet	Immediate release	10	mg	10	mg	1	10
Morphine	Tablet/capsule	Modified release	15	mg	15	mg	1	15
Morphine	Tablet	Immediate release	20	mg	20	mg	1	20
Morphine	Tablet/capsule	Modified release	30	mg	30	mg	1	30
Morphine	Tablet	Immediate release	50	mg	50	mg	1	50
Morphine	Tablet/capsule	Modified release	60	mg	60	mg	1	60
Morphine	Tablet/capsule	Modified release	90	mg	90	mg	1	90
Morphine	Tablet/capsule	Modified release	100	mg	100	mg	1	100
Morphine	Tablet/capsule	Modified release	120	mg	120	mg	1	120
Morphine	Tablet/capsule	Modified release	200	mg	200	mg	1	200
Oxycodone	Tablet	Immediate release	5	mg	5	mg	2	10

Opioid Drug	Dosage Form	Release Type	Dose	Unit	Dose	Unit	Conversion	Morphine
Oxycodone	Targinact (oxycodone and naloxone)	Modified release	5	mg	5	mg	2	10
Oxycodone	Tablet	Modified release	5	mg	5	mg	2	10
Oxycodone	Tablet	Modified release	10	mg	10	mg	2	20
Oxycodone	Targinact (oxycodone and naloxone)	Modified release	10	mg	10	mg	2	20
Oxycodone	Tablet	Immediate release	10	mg	10	mg	2	20
Oxycodone	Tablet	Modified release	15	mg	15	mg	2	30
Oxycodone	Tablet	Modified release	20	mg	20	mg	2	40
Oxycodone	Tablet	Immediate release	20	mg	20	mg	2	40
Oxycodone	Targinact (oxycodone and naloxone)	Modified release	20	mg	20	mg	2	40
Oxycodone	Tablet	Modified release	30	mg	30	mg	2	60
Oxycodone	Tablet	Modified release	40	mg	40	mg	2	80
Oxycodone	Targinact (oxycodone and naloxone)	Modified release	40	mg	40	mg	2	80
Oxycodone	Tablet	Modified release	60	mg	60	mg	2	120
Oxycodone	Tablet	Modified release	80	mg	80	mg	2	160
Oxycodone	Tablet	Modified release	120	mg	120	mg	2	240
Pethidine	Tablet	Immediate release	50	mg	50	mg	0.1125	5.625
Tapentadol	Tablet	Immediate release	50	mg	50	mg	0.25	12.5
Tapentadol	Tablet	Slow release	50	mg	50	mg	0.25	12.5
Tapentadol	Tablet	Immediate release	75	mg	75	mg	0.25	18.75
Tapentadol	Tablet	Slow release	100	mg	100	mg	0.25	25
Tapentadol	Tablet	Slow release	150	mg	150	mg	0.25	37.5
Tapentadol	Tablet	Slow release	200	mg	200	mg	0.25	50
Tapentadol	Tablet	Slow release	250	mg	250	mg	0.25	62.5

Opioid Drug	Dosage Form	Release Type	Dose	Unit	Dose	Unit	Conversion	Morphine
Tramadol	Tablet/capsule	Immediate release	50	mg	50	mg	0.067	3.35
Tramadol	Tablet/capsule	Modified release	50	mg	50	mg	0.067	3.35
Tramadol	Tablet/capsule	Modified release	100	mg	100	mg	0.067	6.7
Tramadol	Tablet/capsule	Immediate release	100	mg	100	mg	0.067	6.7
Tramadol	Tablet/capsule	Modified release	150	mg	150	mg	0.067	10.05
Tramadol	Tablet/capsule	Modified release	200	mg	200	mg	0.067	13.4
Tramadol	Tablet/capsule	Modified release	300	mg	300	mg	0.067	20.1
Tramadol	Tablet/capsule	Modified release	400	mg	400	mg	0.067	26.8
Buprenorphine	Transdermal patch	Immediate release	5	mcg/hr		mg		12
Buprenorphine	Transdermal patch	Immediate release	10	mcg/hr		mg		24
Buprenorphine	Transdermal patch	Immediate release	15	mcg/hr		mg		36
Buprenorphine	Transdermal patch	Immediate release	20	mcg/hr		mg		48
Buprenorphine	Transdermal patch	Immediate release	35	mcg/hr		mg		84
Buprenorphine	Transdermal patch	Immediate release	52.5	mcg/hr		mg		126
Buprenorphine	Transdermal patch	Immediate release	70	mcg/hr		mg		168
Fentanyl	Transdermal patch	Immediate release	12	mcg/hr		mg		45
Fentanyl	Transdermal patch	Immediate release	25	mcg/hr		mg		90
Fentanyl	Transdermal patch	Immediate release	50	mcg/hr		mg		180
Fentanyl	Transdermal patch	Immediate release	75	mcg/hr		mg		270
Fentanyl	Transdermal patch	Immediate release	100	mcg/hr		mg		360
Morphine	Oral solution	Immediate release	2	mg/ml	1	ml	2	2

Opioid Drug	Dosage Form	Release Type	Dose	Unit	Dose	Unit	Conversion	Morphine
Morphine	Oral solution	Immediate release	20	mg/ml	1	ml	20	20
Oxycodone	Oral solution	Immediate release	1	mg/ml	1	ml	2	2
Oxycodone	Oral solution	Immediate release	10	mg/ml	1	ml	20	20

Based on data available at the time

Faculty of Pain Medicine – https://www.fpm.ac.uk/node/21126 (historical)

Current table available at https://www.fpm.ac.uk/opioids-aware-structured-approach-opioid-prescribing/dose-equivalents-and-changing-opioids

http://fpm.anzca.edu.au/documents/opioid-dose-equivalence.pdf

http://www.wales.nhs.uk/sites3/Documents/814/OpiateConversionDoses%5BFinal%5DNov2010.pdf

http://www.rcoa.ac.uk/faculty-of-pain-medicine/opioids-aware/structured-approach-to-prescribing/dose-equivalents-and-changing-opioids

Section 2: Missing Data

Management of partial missing data

Incomplete opioid use data

To calculate total opioid use we required self-report of drug use. For tablets we required data on tablet strength, number of tablets on one dose, number of doses taken on a typical day, and number of days taken over the previous 28 days. For patches we needed patch strength and frequency, and for liquids we needed strength, volume and how long a bottle lasts.

In some cases, one or more parameters were missing. To address this, before any analysis took place, we developed a set of decision rules for the interpretation of missing data.

1. If dose not stated use commonest dose used for same drug. i.e., Co-codamol = 30mg Codeine. For Codeine use 30mg, for dihydrocodeine use 30mg, for morphine use 10mg, for tramadol use 50mg
2. If number of tablets not stated use commonest value for same preparation; for co-codamol =2, for codeine 30mg = 2, for tramadol =2
3. If number of times per day not stated use commonest for same preparation. i.e., codeine=four times per day, Morphine = twice per day, tramadol four times per day
4. Tramadol dose of 37.5mg should be listed as tramacet – but dose is correct
5. For patches if dose not stated use commonest strength patch dose fentanyl 50mcg/hr, Buprenorphine 20mcg/hr
6. For liquid morphine
    - If no strength use commonest reported, i.e. 10mg,
    - If strength is 10mg and volume missing use 5ml
    - If strength is 10mg/5ml and bottle size is missing use 100ml
    - IF 10mg/5ml & 100 ml bottle if frequency missing use 28 days
    - If strength is 20mg and volume missing use 1ml
    - If strength is 20mg/ml and bottle size is missing use 120ml (commonest stated and is the standard pack size.
7. For oxycodone consider change 200ml pack sizes to 250ml.
8. If oxycodone frequency missing use most frequent value – 14 days
9. Where doses not available, or unclear, most common dose from dataset used

These data were used for the health economic analyses and for estimating numbers achieving a greater than 50% reduction in opioid use. The primary outcome analysis of proportion stopping opioids any record of any opioid use at follow up was interpreted as not having stopped opioids.

Imputation for loss to follow-up

We specified in our statistical analysis plan that, if appropriate, we would do a sensitivity analysis using an imputed dataset. We have considered an imputed sensitivity analyses for the primary outcome that was statistically significant. It is here that it is most important to check for potential bias. Because we were unable to analyse opioid use as a continuous outcome the use of conventional multiple imputation techniques was not possible. Analytical approaches to impute categorical outcomes are limited. We have therefore used an inverse probability of missingness weighting analysis to minimise any potential bias in the missing data. The variables used to create the probability weights were the variables listed used in the primary analysis logistic regression model; baseline age, gender, baseline pain intensity (PROMIS 3A) score, baseline opioid usage (MED), trial arm allocation, region (South Tees/West Midlands). The inverse probability weights were then used on the complete case primary analysis for opioid usage at 12 months.

eTable 4: Summary of Opioid use using inverse probability weights

	Education and support Intervention	**Control**	**TOTAL**	**Adjusted estimate (95% CI)*; p-value**
Opioid use: Total MED of all opioid pain killers taken over the last 4 weeks (daily MED) IF taking opioids at 12 months				4.61 (2.09 to 10.13); **p<0.001**
N	225	208	433	
Taking non-opioids only (MED=0)	65 (29%)	15 (7%)	80 (18%)	
Taking opioids (MED>0)	160 (71%)	193 (93%)	353 (82%)	

** Based on partially nested mixed-effect logistic model adjusted for age, gender, baseline pain intensity, geographical location and baseline opioid band, with inverse probability weights. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care. Odds ratio and 95% confidence interval reported.*

Section 3: Ethnicity and Deprivation Data

Ethnicity and deprivation data

The participants in the I-WOTCH trial are overwhelming White British. This might be interpreted as limiting the generalisability of our findings. However, rates for opioid prescribing vary considerably across England. Opioid prescribing is substantially higher in the North-East of England than in London and the South-East. For this reason, we specifically targeted recruitment in the North-East of England. There is much less ethnic diversity in the North-East than the rest of England meaning that targeting the area of greatest need reduced ethnic diversity.

In the original design we had planned to also recruit in London, the most ethnically diverse part of the UK. However, we were unable to secure funding from local health care purchasers to deliver the intervention in London. This might reflect that opioid use was not seen as a major health care problem when compared to the North-East where health care providers were keen to support intervention delivery.

The North-East of England is also typically less affluent than the West Midlands and is typically underserved in terms of recruitment to randomised controlled trials. To explore how these factors might have affected recruitment we extracted data from the National General Practices Profiles database (last accessed 2nd June 2022 https://fingertips.phe.org.uk/profile/general-practice) which records the deprivation decile in which the practice is sited, and the ethnicity of the practice population as recorded on the general practice record

We recruited from 120 sites in the West Midlands and 71 sites in South Tees (not including separate sites from self-referrals). We obtained data from 113 practices from the West Midlands and 70 from the North-East, which covered 598/608 (98.4%) of recruited participants. We excluded a University Health Centre in the West Midlands with a very large, recorded list size (>46,000) principally of young people who are likely to be highly mobile many of whom may have left the locality or the country, and very ethnically mixed. Overall, this practice's patients are not representative our population of interest. We recruited just 2 participants from this practice. This left us with a 1,891,992 registered patients. In the West Midlands 11% were recorded as other than white and in the North-East was just 3% making the overall total just 8% (eTable 5). In the West Midlands most patients were registered with practices based in less deprived locations, whilst in the North-East most came from more deprived locations. Overall, 46% of our pool of potential participants came from practices based in more deprived localities (eTable 6). The deprivation decile of where a practice's patients live may be different from that of where the practice is located, so some caution is needed in interpreting these data.

Overall, we randomised 596/1,878,976 of the practice populations; 368/1,196,264 in West Midlands (excluding 2 randomisations from the University health centre, 1 self-referral, and 9 randomisations from sites with no ethnicity data available) and 228/682,712 in the North-East.

These data give some insights into the challenges of recruiting an ethnically, and socio-economically diverse group of participants to the trial.

eTable 5: Ethnicity by region

	West Midlands		North-East		Total	
Mixed	23875	2%	3226	<1%	27101	1%
Asian	83180	7%	9534	1.4%	92714	5%
Black	21689	2%	811	<1%	22500	1%
Non-white	8642	1%	7231	1%	15873	1%
'White'[a]	1,058,879	89%	661,910	97%	1720788	92%
Total	1,196,264	100%	682,712	100%	1,878,976	100.0%

[a] Calculated by subtracting sum of Mixed, Asian, Black & Non-White from total list size. This may overestimate White population if ethnicity has not been recorded.

eTable 6: Deprivation by region

Deprivation decile [a]	West Midlands		North-East		Total	
1	82742	7%	51416	8%	134158	7%
2	49955	4%	141366	21%	191321	10%
3	57784	5%	94600	14%	152384	8%
4	73501	6%	76837	11%	150338	8%
5	167146	14%	75033	11%	242179	13%
6	92802	8%	110883	16%	203685	11%
7	139751	12%	38867	6%	178618	10%
8	164694	14%	38166	6%	202860	11%
9	139930	12%	38516	6%	178446	9%
10	227959	19%	17028	2%	244987	13%
Total	1,196,264	100%	682712	100%	1,878,976	100%

[a] 1=Most deprived, 10 = least deprived

Section 4: Results - Supplementary Figures and Tables

eTable 7: Randomised participants by treatment and randomisation stratum Opioid band[a]

	Opioid band (0-29 MED per day)	
Region	**Education and support Intervention**	**Usual care**
South Tees	69	66
West Midlands	96	98
	Opioid band (30-59 MED per day)	
Region	**Education and support Intervention**	**Usual care**
South Tees	10	15
West Midlands	34	30
	Opioid band (60-89 MED per day)	
Region	**Education and support Intervention**	**Usual care**
South Tees	8	14
West Midlands	18	13
	Opioid band (90-119 MED per day)	
Region	**Education and support Intervention**	**Usual care**
South Tees	7	5
West Midlands	12	12
	Opioid band (120-149 MED per day)	
Region	**Education and support Intervention**	**Usual care**
South Tees	5	3
West Midlands	10	12
	Opioid band (≥150 MED per day)	
Region	**Education and support Intervention**	**Usual care**
South Tees	15	11
West Midlands	21	24

[a] Calculated using initial conversion values used for the tapering App

eTable 8: Randomised participants by treatment and randomisation stratum pain intensity

Region	Low Intensity Pain		High Intensity Pain	
	Education and support intervention	Usual care	Education and support intervention	Usual care
South Tees	11	9	103	105
West Midlands	11	12	180	177

eTable 9: Baseline confidence outcomes of all randomised participants by treatment group

	Education and support intervention N=305	Usual care N=303
I want to reduce my opioid use		
N	299	299
Not at all	21 (7%)	25 (8%)
By a little	37 (12%)	45 (15%)
By Half	44 (15%)	36 (12%)
So I only use a little	95 (32%)	60 (20%)
So I use no opioids	102 (34%)	133 (44%)
I expect in 4 months' time, I will have reduced my opioid use		
N	300	296
Not at all	43 (14%)	45 (15%)
By a little	82 (27%)	78 (26%)
By Half	56 (19%)	56 (19%)
So I only use a little	82 (27%)	67 (23%)
So I use no opioids	37 (12%)	50 (17%)
I am confident I could reduce my opioid use a lot over 4 months		
N	301	296
Not at all confident	90 (30%)	90 (30%)
Somewhat confident	77 (26%)	70 (24%)
Fairly confident	79 (26%)	79 (27%)
Strongly confident	40 (13%)	35 (12%)
Completely confident	15 (5%)	22 (7%)
I feel that involvement in this study can help me to reduce my opioid use		
N	296	293
Not at all	22 (7%)	25 (9%)

	Education and support intervention N=305	Usual care N=303
By a little	73 (25%)	76 (26%)
By Half	46 (16%)	38 (13%)
So I only use a little	86 (29%)	68 (23%)
So I use no opioids	69 (23%)	86 (29%)

eTable 10: Overall summary of withdrawals by treatment group*

	Education and support intervention N=305	Usual care N=303	TOTAL N=608
Participant withdrew from intervention package only and will be followed up	8 (3%)	Not Applicable	8 (1%)
Participant withdrew from study completely and will not be followed up	46 (15%)	53 (17%)	99 (16%)
Participant withdrew consent for receiving text messages in relation to study	47 (15%)	52 (17%)	99 (16%)
Participant withdrew consent for taking part in the interview study	45 (15%)	46 (15%)	91 (15%)

** Summary of total number of withdrawal requests i.e. participants may have more than one request*

eTable 11: Follow-up rates throughout the trial, by treatment group[1]

Treatment Group	Time-point	Total to reach time point	Not followed up		Follow-up due	Follow-up outcome		Total completed**
			Deceased	Withdrawn completely		Completed*	Non Responder	
Usual care	**Baseline**	303	0 (0%)	0 (0%)	303	303 (100%)	0 (0%)	303 (100%)
	4 month	303	0 (0%)	5 (2%)	298	202 (68%)	96 (32%)	202 (67%)
	8 month	298	1 (<1%)	32 (11%)	265	166 (63%)	99 (37%)	166 (55%)
	12 month	265	0 (0%)	8 (3%)	257	211 (82%)	46 (18%)	211 (70%)
Education and support intervention	**Baseline**	305	0 (0%)	0 (0%)	305	305 (100%)	0 (0%)	305 (100%)
	4 month	305	2 (<1%)	8 (3%)	295	228 (77%)	67 (23%)	228 (75%)
	8 month	295	0 (0%)	26 (9%)	269	199 (74%)	70 (26%)	199 (65%)
	12 month	269	2 (<1%)	9 (3%)	258	229 (89%)	29 (11%)	229 (75%)

*% out of follow-up due
**% out of total randomised

[1] Empty questionnaires that were returned have been classed as non-responder

eTable 12: Timing of complete withdrawals throughout the trial, by treatment group

	Education and support intervention N=305	Usual care N=303	TOTAL N=608
Post-randomisation to 4-month follow-up	8 (32.6%)	5 (2%)	13 (2%)
4 - 8 months follow-up	26 (98.5%)	32 (11%)	58 (10%)
8 - 12 months follow-up	12 (43.9%)	16 (5%)	28 (5%)
Overall	46 (15.1%)	53 (17%)	99 (16%)

eTable 13: Baseline characteristics of complete withdrawals

	Education and support intervention N=46	Usual care N=53
Age (years)		
Mean (SD)	65.7 (12.4) [n=46]	62.8 (14.1) [n=53]
Gender		
Female	36/46 (78%)	35/53 (66%)
Male	10/46 (22%)	18/53 (34%)
Ethnicity		
White	44/46 (96%)	51/53 (96%)
Black Caribbean	0/46 (0%)	1/53 (2%)
Indian	2/46 (4%)	1/53 (2%)
Employment status		
Employed	6/46 (13%)	7/53 (13%)
Unemployed	1/46 (2%)	2/53 (4%)
Unable to work due to long term sickness	11/46 (24%)	9/53 (17%)
Looking after your family	2/46 (4%)	0 (0%)
Retired from paid work	25/46 (54%)	33/53 (62%)
Other[i]	4/46 (9%)	4/53 (8%)
Age left full time education		

No formal education	0/46 (0%)	1/53 (2%)
Age 16 years or under	27/46 (59%)	32/53 (60%)
Age 17 years or over[ii]	19/46 (41%)	19/53 (36%)
Other	0/46 (0%)	1/53 (2%)
Length of time pain experienced		
5 years or less	7/46 (15%)	10/53 (19%)
More than 5 years	39/46 (85%)	43/53 (81%)
Opioid band[iii]		
0-29.9 MED per day	19/46 (41%)	19/53 (36%)
30-59.9 MED per day	12/46 (26%)	21/53 (40%)
60-89.9 MED per day	5/46 (11%)	5/53 (9%)
90-119.9 MED per day	4/46 (9%)	2/53 (4%)
120-149.9 MED per day	1/46 (2%)	1/53 (2%)
≥150 MED per day	5/46 (11%)	5/53 (9%)
How long opioids taken		
5 years or less	18/46 (39%)	25/53 (47%)
More than 5 years	28/46 (61%)	28/53 (53%)
Type of pain disorder[iv]		
Lower Back Pain	39/45 (85%)	41/53 (77%)
Chronic Widespread Pain	31/45 (67%)	17/53 (32%)
Multi-site pain	43/45 (93%)	45/53 (85%)
Pain interference (PROMIS-8A)[a] Mean (SD)	68.7 (5.6) [n=46]	68.7 (6.2) [n=53]
Pain intensity (PROMIS-3A)[b] Mean (SD)	70.9 (6.2) [n=46]	69.8 (7.4) [n=53]
SF-12 Mental[c] Mean (SD)	41 (11.7) [n=46]	42 (9.5) [n=53]
SF-12 Physical[c] Mean (SD)	32 (8.2) [n=46]	32 (7.5) [n=53]
Pittsburgh SQI[d] Mean (SD)	12 (4.4) [n=41]	11 (4.2) [n=50]

HADS Anxiety[e] Mean (SD)	8 (4.3) [n=46]	9 (5.2) [n=51]
HADS Depression[e] Mean (SD)	9 (3.7) [n=46]	9 (4.2) [n=53]
Pain self-efficacy[f] Mean (SD)	24 (11.3) [n=46]	25 (13.2) [n=53]
EQ-5D-5L utility[g] Mean (SD)	0.3 (0.3) [n=46]	0.3 (0.3) [n=53]
EQ-5D-5L VAS[g] Mean (SD)	46 (19.4) [n=46]	46 (22.0) [n=53]
ShOWS[h] Mean (SD)	10 (5.5) [n=46]	10 (5.0) [n=53]

[i]Other employment status includes participants who are still in education part/full time, look after home/family, unemployed or other [ii]Leaving education at age 17 years or over includes participants who left education between age 17-19 years, age 20 or over, or participants still in education [iii]Opioid band by region, See eTable 2 [iv]Participants were asked to place crosses on a body chart to report where on their body they experienced pain. A grid was then used to split the body into pain areas. Two researchers MU & HS Independently inspected these and conferred on any disagreements. The American College of Rheumatology definition of chronic widespread pain was used. [1] We used Carne's et al's approach to defining multisite pain – i.e., pain marked in more than two different areas in an overlaid scoring grid.[2] Any mark over the lower back, using the same scoring grid as Carnes et al was taken to indicate low back pain. Therefore, multiple sources of pain have been reported for some participants.

a Patient-reported Outcomes Measurement Information System (PROMIS) Pain interference Short Form (8A) standardised T scores reported, calculated using the recommended HealthMeasures Scoring Service.[3] T scores range from 40.7-77 with higher scores indicating a worse outcome (more pain interference).
b Patient-reported Outcomes Measurement Information System (PROMIS) Pain intensity Short Form (3A) standardised T scores reported, calculated using the recommended HealthMeasures Scoring Service.[3] T scores range from 36.3-81.8 with higher scores indicating a worse outcome (more pain intensity).
c 12-Item Short Form Health Survey (SF-12) Mental and Physical scores, calculated using the recommended PRO CoRE software provided by the authors Optum, range from 0-100 with higher scores reflecting better physical and mental functioning.
d Pittsburgh Sleep Quality Index (PSQI) scores range from 0-21, with higher scores indicating worse sleep quality. The 19 self-reported questions are combined to create seven component scores. The score is calculated by summing the seven component scores (range 0-3) to create a global score ranging from 0-21. This global score has been reported.
e Hospital Anxiety and Depression Scale (HADS) anxiety and depression scores range from 0-21, with higher scores indicating worse anxiety/depression. Each of the seven questions measuring anxiety have a score ranging from 0-3. These seven scores are summated to create the reported anxiety score. The same method applies to depression score.
f Pain self-efficacy questionnaire (PSEQ) scores range from 0-60 with higher scores indicating stronger self-efficacy beliefs. The PSEQ consists of 10 questions, each having a score ranging from 0-6. The PSEQ score is calculated by summing these 10 scores to create the reported score.
g EuroQol-5 Dimension (EQ-5D-5L) utility score ranges from <0-1, with higher scores indicating better quality of life. The EQ-5D-5L utility scores were calculated using the EQ-5D-5L Crosswalk Index Value Calculator developed by the EuroQol Group. EQ-5D-5L Visual Analogue Scale (VAS) score ranges from 0-100, with scores of 100 indicating 'best health you can imagine' and 0 indicating 'worst health you can imagine'. These scores ranging from 0-100 were self-reported by participants and that self-reported score is reported.
h Short Opioid Withdrawal Scale (ShOWS) score ranges from 0-30 where a higher score indicates more severe symptoms. The ShOWS consists of 10 questions, each with a score of 0-3, which are summed together to give the reported score.

eTable 14: Time from randomisation to withdrawal and follow-up summarised by treatment group

	Education and support intervention N=305	Usual care N=303	TOTAL N=608
Time from randomisation to withdrawing completely from trial (weeks)			
N	46	53	99
Mean (SD)	26.8 (15.4)	30.0 (16.8)	28.5 (16.2)
Median (IQR)	24 (19-36)	24 (20-39)	24 (20-38)
Time from randomisation to 4 month follow-up (months)			
N	228	201[2]	429
Mean (SD)	4.8 (0.8)	4.8 (2.0)	4.8 (1.5)
Median (IQR)	4.5 (4.2-5.1)	4.6 (4.2-5.5)	4.6 (4.2-5.3)
Time from randomisation to 8 month follow-up (months)			
N	199	165[3]	364
Mean (SD)	8.4 (1.0)	8.4 (1.5)	8.5 (1.2)
Median (IQR)	8.4 (8.1-8.7)	8.3 (8.1-8.7)	8.4 (8.1-8.7)
Time from randomisation to 12 month follow-up (months)			
N	229	211	440
Mean (SD)	12.8 (1.2)	12.7 (1.5)	12.8 (1.4)
Median (IQR)	12.4 (12.1-12.9)	12.5 (12.1-13.1)	12.4 (12.1-13.0)

[2] 1/202 at 4 months had no questionnaire completion date, so time from randomisation cannot be calculated
[3] 1/166 at 8 months had no questionnaire completion date, so time from randomisation cannot be calculated

eTable 15: Summary of intervention data

	Education and support intervention N=305
Time from randomisation to first group session (days)	
Mean (SD)	19.9 (24.8) [n=296]
Median (IQR)	12 (6, 23) [n=296]
Didn't attend first group session	85/296 (29%)
Time from randomisation to first one-to-one consultation (days)	
Mean (SD)	28.0 (20.5) [n=269]
Median (IQR)	22 (16, 33) [n=269]
Didn't attend first one-to-one	98/269 (36%)
Time from first group session to final one-to-one consultation (Course duration) (days)	
Mean (SD)	33.4 (28.7) [n=273]
Median (IQR)	44 (9, 58) [n=273]
Didn't attend final one-to-one	107/273 (39%)
Group session attendance[4,5]	
Number randomised to intervention	305
Attended day 1 only	13/305 (4%)
Attended day 1 and 2 only	17/305 (6%)
Attended day 1 and 3 only	10/305 (3%)
Attended day 1, 2 & 3	166/305 (54%)
Attended no days	90/305 (30%)
No attendance data available	9/305 (3%)
Group size at randomisation (participants) – I-WOTCH Intervention group only	
Number of groups	35
Mean (SD)	8.71 (2.9) [n=35]
Median (IQR)	9 (5, 11) [n=35]
Group size at Session 1 (participants)[6]	
Number of groups	35
Mean (SD)	6.24 (2.82) [n=35]
Median (IQR)	7 (3, 8) [n=35]
Missing groups	2/35 (6%)
Face to Face interviews	
Attended first F2F interview	190/290 (66%)
Attended both F2F interviews	131/290 (45%)
Telephone interviews	
Attended first telephone session	167/271 (62%)
Attended both telephone sessions	152/271 (56%)
Compliance	
Number of participants who had full compliance	144/305 (47.2%)
Number of participants who had minimal compliance	190/305 (62.3%)
Number who had less than minimal compliance	115/305 (37.7%)[7]

NOTE: Minimal compliance is defined as attending at least Day 1 and F2F#1. Full compliance is defined as attending at least Day 1, 2 & 3, F2F#1 & at least one phone call.

[4] 161/305 participants achieved minimal compliance by attending at least Day 1 and the first one-to-one consultation.

[5] 144/305 participants achieved full compliance by attending at least Day 1, 2 & 3, the first one-to-one consultation and 1 telephone call.

[6] 6 participants who attended day 1 attended groups they were not randomised too. This has been summarised by groups they attended, not randomised too.

[7] 9 of these had no attendance data, so were classed as non-complier.

eTable 16: Fidelity scores of group sessions in percentages

Session	Adherence *(italics 10% check)*			Totals	Competence			Totals
	Early	Mid	Late		Early	Mid	Late	
Day 1 session 2	100	81	25 *(25)*		92	64	0 *(0)*	
Day 1 session 3	94 *(94)*	100	44		90 *(87)* agreed 88	86	42	
Day 1 session 4	100	75	88		70	71	67	
Day1 session 7	72	89	83		80	75	90	
Day 1 session 8	100	88	81		100	60	75	
Day 2 session 13	77	88 *(90)* agreed 89	84		57	100 *(100)*	100	
Day 2 session 14	56 *(66)* agreed 61	100	94		58*(43)* agreed 50	92	90	
Day 2 session 16	79	88	82		50	70	100	
Day 3 session 21	88	100	89		80	100	100	
Day 3 session 22pt 1	93	86	64		70	92	83	
Day3 session 23	90	73	91		100	100	100	
Average	**86.72**	**88.09**	**75.00**	**83.27%**	**76.09**	**82.73**	**77**	**78.61%**
Range	**61-100**	**73-100**	**25-94**	**25-100%**	**50-100**	**60-100**	**0-100**	**0-100%**
Median	**90**	**88**	**83**	**88**	**80**	**86**	**90**	**86**

eTable 17: Fidelity percentage scores of one-to-one nurse consultations

Timepoint Early	Group ID	Adherence Score	Competence Score
1st	1	100%	100%
2nd	2	93% *(93%)*	100% (*100%)*
1st	3	61%	67%
2nd	3	86%	83%
1st	4 NE	100%	100%
2nd	4 NE	100%	100%
1st	5 NE	100%	100%
2nd	5 NE	100%	100%
2nd	6	100%	100%
1st	7 NE	67%	100%
1st	8	100%	100%
1st	9 NE	83% *(94%)*94	100% *(83%)*92
2nd	10	88%	83%
2nd	10	100%	100%
Early averages		**92.07%**	**94.64%**
Range		**61 to 100**	**67 to 100**
Timepoint Mid	**Group ID**	**Adherence**	**Competence**

1st	11	100%	100%
1st	12 NE	94%	100%
2nd	12 NE	100%	100%
1st	13	89%	92%
1st	14 NE	61% *(61%)*	50% *(50%)*
1st	15	94%	100%
1st	15	94%	100%
2nd	20	71%	83%
Mid averages		**87.87%**	**90.63%**
Range		**61 to 100**	**50 to 100**
Timepoint Late	**Group ID**	**Adherence Score**	**Competence Score**
1st	21	100%	92%
2nd	21	93%	100%
1st	24	94%	100%
2nd	24	86% *(86%)*	100% *(100%)*
1st	27	89%	83%
Late averages		**92.4%**	**95%**
Range		**86 to 100**	**83 to 100**
Total averages		**90.78%**	**93.42%**
Range		**61 to 100**	**50 to 100**
Legend: NE – North-East region			

eTable 18: Numbers fully tapered off opioids at each time point, out of the baseline opioid usage bands

Baseline opioid daily usage (MED)	Fully tapered at 4 months		Fully tapered at 8 months		Fully tapered at 12 months	
	Education and support intervention	Usual care	Education and support intervention	Usual care	Education and support intervention	Usual care
0-29 (N=201)	33/103 (32%)	3/98 (3%)	31/103 (30%)	6/98 (6%)	36/103 (35%)	6/98 (6%)
30-59 (N=198)	16/95 (17%)	3/103 (3%)	14/95 (15%)	5/103 (5%)	13/95 (14%)	6/103 (6%)
60-89 (N=86)	6/42 (14%)	0/44 (0%)	7/42 (17%)	0/44 (0%)	7/42 (17%)	1/44 (3%)
90-119 (N=34)	1/17 (6%)	1/17 (6%)	1/17 (6%)	0/17 (0%)	4/17 (24%)	1/17 (6%)
120-149 (N=23)	1/11 (9%)	0/12 (0%)	2/11 (18%)	0/12 (0%)	2/11 (18%)	0/12 (0%)
≥150 (N=66)	1/37 (3%)	0/29 (0%)	0/29 (0%)	2/37 (5%)	1/29 (3%)	3/37 (8%)

eTable 19: Pre-specified ITT and instrumental variable analysis to adjusted for non-adherence, at each time point (using full compliance definition of compliance)

	ITT model		IV analysis	
	Adjusted estimate (95% CI)*	p-value	Adjusted estimate (95% CI)†	p-value
PROMIS-8A				
4 months	-0.73 (-1.93 - 0.48)	0.24	-1.27 (-3.62 - 1.07)	0.29
8 months	-0.75 (-2.10 - 0.59)	0.27	-1.24 (-3.93 - 1.45)	0.37
12 months	-0.89 (-2.12 - 0.33)	0.15	-1.47 (-4.04 - 1.09)	0.26

** Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band and baseline PROMIS-8A T-score. The education support group was used as the cluster variable for the education and support intervention arm, with clusters of size 1 used for each participant in usual care.*

†Based on a single equation instrumental variable regression model with outcome adjusted for age, gender, baseline pain intensity and baseline PROMIS-8A T-score, using full compliance as the instrumented variable to estimate treatment effect for the compliers. Full compliance is defined as attending at least Day 1, 2 & 3, F2F#1 & at least one phone call.

eTable 20: Pre-specified ITT and instrumental variable analysis to adjusted for non-adherence, at each time point (using minimal compliance definition of compliance)

	ITT model		IV analysis	
	Adjusted estimate (95% CI)*	p-value	Adjusted estimate (95% CI)†	p-value
PROMIS-8A				
4 months	-0.73 (-1.93 - 0.48)	0.24	-0.88 (-2.63 - 0.88)	0.33
8 months	-0.75 (-2.10 - 0.59)	0.27	-0.85 (-2.89 - 1.19)	0.42

12 months	-0.89 (-2.12 - 0.33)	0.15	-0.99 (-2.93 - 0.96)	0.32

** Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band and baseline PROMIS-8A T-score. The education support group was used as the cluster variable for the education and support intervention arm, with clusters of size 1 used for each participant in usual care.*

†Based on a single equation instrumental variable regression model with outcome adjusted for age, gender, baseline pain intensity and baseline PROMIS-8A T-score, using minimal compliance as the instrumented variable to estimate treatment effect for the compliers. Minimal compliance is defined as attending at least Day 1 and F2F#. Non-compliance is attending any less than this.

eTable 21: Pre-specified Sub-group analyses of the 12-month PROMIS-8A outcome

Subgroups	Education and support intervention N; mean (95% CI)	Usual care N; mean (95% CI)	Unadjusted effect estimate (95% CI)	Interaction effect; p-value*
Anxiety				p=0.15
<9	114; 62.6 (61.1, 64.1)	93; 63.3 (61.8, 64.8)	-0.71 (-2.84, 1.43)	
≥9	113; 65.7 (64.4, 67.0)	115; 65.7 (64.3, 67.0)	0.04 (-1.81, 1.90)	
Depression				p=0.25
<9	109; 61.7 (60.2, 63.1)	100; 62.3 (60.8, 63.7)	-0.60 (-2.66, 1.46)	
≥9	119; 66.4 (65.1, 67.7)	105; 66.7 (65.5, 68.0)	0.35 (-2.14, 1.45)	

** Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band, baseline PROMIS-8A T-score and interaction term. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care.*

eTable 22: Pre-specified Sub-group analyses of the 12-month Opioid use outcome

Subgroups		Usual care N (%)	Education and support intervention N (%)	Unadjusted effect estimate (95% CI)*	Interaction effect; p-value**
Anxiety <9	MED=0	8/92 (9%)	32/110 (29%)	4.31 (1.87 - 9.92)	p=0.31
	MED>0	84/92 (91%)	78/110 (71%)		
≥9	MED=0	6/114 (5%)	33/113 (29%)	7.42 (2.97 - 18.57)	
	MED>0	108/114 (95%)	80/113 (71%)		
Depression <9	MED=0	8/99 (8%)	29/106 (27%)	4.28 (1.85 - 9.92)	P=0.46
	MED>0	91/99 (92%)	77/106 (73%)		
≥9	MED=0	6/104 (6%)	36/118 (30.5%)	7.17 (2.88 - 17.86)	
	MED>0	98/104 (94%)	82/118 (69.5%)		

** Odds ratio (95% CI) reported.*

*** Based on partially nested mixed-effect logistic model adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band and interaction term. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care. Odds ratio and 95% confidence interval reported.*

Section 5: Additional Analysis

eTable 23: Pre-specified Sensitivity analysis - treatment effectiveness estimate based on the primary outcomes having excluded those participants included in the process evaluation interviews

	Education and support intervention N=208	Usual care N=192	Unadjusted estimate (95% CI); p-value	Adjusted estimate (95% CI); p-value
PROMIS-8A			-0.68 (-2.19 to 0.82); p=0.37	-1.04 (-2.33 to 0.26); p=0.12*****
N	208	191		
Mean (SD)	63.9 (7.8) [n=208]	64.6 (7.5) [n=191]		
Opioid use			5.87 (3.10 to 11.10); **p<0.001****	6.12 (2.95 to 12.71); **p<0.001**†
Taking non-opioids only (MED=0)	62/205 (30.2%)	13/189 (7%)		
Taking opioids (MED>0)	143/205 (69.870%)	176/189 (93%)		

** Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band and baseline PROMIS-8A T-score. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care.*

*** Odds ratios (95% CI) and p-value reported.*

† Based on partially nested mixed-effect logistic model adjusted for age, gender, baseline pain intensity, geographical location and baseline opioid band. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care. Odds ratio and 95% confidence interval reported.

eTable 24: Pre-specified Sensitivity analysis - treatment effectiveness estimate based on the primary outcome having adjusted for any imbalance in death rates across both treatment groups

	Education and support intervention N=305	Usual care N=303	Adjusted estimate (95% CI); p-value
PROMIS-8A			-0.90 (-2.09 - 0.30); p=0.14*****
Mean (SD)	64.2 (7.7) [n=229]	64.7 (7.3) [n=210]	
Opioid use			5.55 (2.80 to 10.99); **p<0.001**†
Taking non-opioids only (MED=0)	65/225 (29%)	15/208 (7%)	
Taking opioids (MED>0)	160/225 (71%)	193/208 (93%)	

** Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band, baseline PROMIS-8A T-score and death. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care.*

† Based on partially nested mixed-effect logistic model, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid use (MED) and death. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care. Odds ratio (95% CI) and p-value reported.

eTable 25: Pre-specified Treatment effectiveness estimates based on the primary outcome for people with different pain disorders

Primary outcome	Pain disorder*	Usual care	Education and support intervention	Adjusted estimate (95% CI); p-value
PROMIS-8A	**Lower back pain**			
	Mean (SD)	64.7 (7.1) [n=175]	64.4 (7.7) [n=178]	-0.83 (-2.24 to 0.57); p=0.24**
	Chronic wide spread pain			
	Mean (SD)	66.0 (6.5) [n=103]	66.0 (7.1) [n=103]	-0.52 (-2.33 to 1.30); p=0.57**
	Multi-site pain			
	Mean (SD)	64.7 (7.3) [n=183]	64.7 (7.5) [n=207]	-0.41 (-1.70 to 0.88); p=0.53**
Opioid use	**Back pain**			
	Taking non-opioids only (MED=0)	9/173 (5%)	52/174 (30%)	7.66 (3.38 to 17.35); **p<0.001**†
	Taking opioids (MED>0)	164/173 (95%)	122/174 (70%)	
	Chronic wide spread pain			
	Taking non-opioids only (MED=0)	7/101 (7%)	31/101 (31%)	5.38 (1.73 to 16.71); **p<0.004**†
	Taking opioids (MED>0)	94/101 (93%)	70/101 (69%)	
	Multi-site pain			
	Taking non-opioids only (MED=0)	13/181 (7%)	59/205 (29%)	5.15 (2.52 to 10.52); **p<0.001**†
	Taking opioids (MED>0)	168/181 (93%)	146/205 (71%)	

** 482 participants were classed under multiple pain categories. 34 participants did not fall into any of the pain categories.*
*** Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location, baseline opioid band and baseline PROMIS-8A T-score. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care.*
† Based on partially nested mixed-effect logistic model adjusted for age, gender, baseline pain intensity, geographical location and baseline MED. The education support group was used as the cluster variable for the education and support intervention arm, with individual clusters of size 1 used for each participant in usual care. Odds ratio and 95% confidence interval reported.

eTable 26: Adverse events (AE) and serious adverse events (SAE) summarised by treatment group*

	Education and support intervention N=305	Usual care N=303
AEs		
Number of AEs reported	25	11
Number of participants reporting AE	22/305 (7%)	8/303 (3%)
SAEs		
Number of SAEs reported	32	20
Number of particpants reporting SAE	25/305 (8%)	16/303 (5%)
Reason Serious Adverse Event deemed serious*		
Death[8]	4/305 (1%)	1/303 (<1%)
Life-threatening	1/305 (<1%)	2/303 (1%)
Hospitalisation or prolongation of existing hospitalisation	25/305 (8%)	17/303 (6%)
Persistent or significant disability or incapacity	1/305 (<1%)	0/303 (0%)
Congenital anomaly/birth defect	0/305 (0%)	0/303 (0%)
Other	2/305 (1%)	1/303 (<1%)
SAE severity assessment*		
Mild	0/305 (0%)	1/303 (<1%)
Moderate	9/305 (3%)	5/303 (2%)
Severe	19/305 (6%)	13/303 (4%)
Fatal/life threatening	4/305 (1%)	1/303 (<1%)

*% out of total randomised, however some participants have reported multiple SAEs

eTable 27: Assessment of SAEs summarised by treatment group*

Assessment of SAEs	Education and support intervention N=305	Usual care N=303
SAE related to trial intervention:		
Definitely	0/305 (0%)	0/303 (0%)
Probably	1/305 (<1%)	0/303 (0%)
Possibly	3/305 (1%)	1/303 (<1%)
Unlikely	4/305 (1%)	7/303 (2%)
Unrelated	24/305 (8%)	12/303 (4%)
Expectedness of SAE[9]:		
Expected	2/305 (1%)	1/303 (<1%)
Unexpected	2/305 (1%)	0/303 (0%)

*% out of total randomised

[8] All 5 deaths were classed as unrelated
[9] Only applicable if relation to trial is definitely, probably or possibly

eTable 28: Full details of AEs by treatment group (N=36)

Adverse Events	**Education and support intervention (N=305)**		**Usual care (N=303)**	
	N (%)	**Details**	**N (%)**	**Details**
Cardiovascular	2 (1%)	- Tiredness/Dizzy (n=2)	1 (<1%)	- Coldness in feet (n=1)
Psychological	7 (2%)	- Sleep disturbance (n=2) - Panic attack (n=1) - Excessive sweating and UTI (n=1) - Suicidal thoughts (n=1) - Low mood and suicidal ideation (n=2)	2 (1%)	- Sleep apnoea and suicidal ideation (n=1) - Previous suicide attempt (n=1)
Respiratory	2 (1%)	- Bronchitis chest infection (n=1) - Sleep apnoea (n=1)	4 (1%)	- Cold (n=1) - Throat infection (n=1) - Chest infection (n=2)
Nervous system	6 (2%)	- Headache (n=3) - Headaches and muscle spasms (n=1) - Withdrawal symptoms (n=2)	1 (<1%)	- Vertigo (n=1)
Renal/Urological	1 (<1%)	- Urinary tract infection (n=1)	1 (<1%)	- Urinary tract infection (n=1)
Locomotor/ Musculoskeletal	1 (<1%)	- Fall (n=1)	2 (1%)	- Muscle cramps (n=1) - Fall (n=1)
Gastrointestinal	4 (1%)	- Food poisoning and gastroenteritis (n=1) - Flu and pancreatic pain (n=1) - Constipation (n=1) - Hiatus Hernia (n=1)	0 (0%)	
Other	2 (1%)	- Lymphoma (n=1) - Itching (n=1)	0 (0%)	
Total	**25 (8%)**		**11 (4%)**	

% Out of total randomised by arm

eTable 29: Full details of SAEs by treatment group (N=52)

SAES	Education and support intervention (N=305)		Usual care (N=303)	
	N (%)	Details	N (%)	Details
Cardiovascular	3 (1%)	- Hospital admission for chest pain (n=2) - HA for Myopericarditis (n=1)	2 (1%)	- Hospital admission for chest pain (n=1) - Hospital admission with stroke, aspiration, pneumonia (n=1)
Respiratory	3 (1%)	- Surgery for chronic cough (n=1) - Hospital admission for respiratory issues (asthma) (n=1) - HA for chest infection (n=1)	1 (<1%)	- Hospital admission for back pain, shortness of breath, exacerbation of COPD (n=1)
Nervous system	1 (<1%)	- Hospital admission for left weakness and numbness (n=1)	1 (<1%)	- Hospital admission for severe headache (n=1)
Gastrointestinal	7 (2%)	- Hospital admission for duodenal bleed (n=1) - Hospital admission for cholangitis (n=1) - Hospital admission for acute pancreatis (n=1) - Hospital admission for gastric band surgery (n=1) - Hospital admission for constipation (n=1) - Hospital admission for tension and hot sensations (withdrawal symptoms) (n=1)*	0 (0%)	
Locomotor/ Musculoskeletal	6 (2%)	- Hospital admission for procedure on lower back (n=1) - Hospital admission for spinal surgery (for scoliosis) (n=1) - Hospital admission for joint and back pain and spasms in abdomen (n=1)* - Hospital admission for decompression of spine (n=1) - Hospital admission due to weakness in legs (n=1) - Hospital admission after fall (n=1)	4 (1%)	- Surgery for lower back decompression and nerve clearing (n=1) - HA for arthritis flare up (n=1)* - HA for total knee replacement (n=1) - HA after road traffic accident (n=1)
Renal/Urological	2 (1%)	- Hospital admission for kidney stones (n=1) - Hospital admission for pain (n=1)	1 (<1%)	- HA for kidney removal (n=1)
Accident and Emergency care	3 (1%)	- Chest pain (n=1) - Viral URTI (n=1)	7 (2%)	- Breathlessness (n=2)

SAES	Education and support intervention (N=305)		Usual care (N=303)	
		- Tired and feeling unwell (n=1)		- Muscle cramp (lower leg) (n=1) - Chest pain (n=1) - Back pain and bowel incontinence (n=1) - Suicidal ideation (n=1) - Pleuritic cough and shortness of breath (n=1)
Acute assessments	1 (<1%)	- Chest pain (n=1)	0 (0%)	
Deaths	4 (1%)	- Death (n=1) - Complications due to lymphoma (n=1) - Mastoiditis and subdural empyema secondary to otitis media (n=1) - Unknown reason (prior to intervention) (n=1)	1 (<1%)	- Dense stroke then hospice care (n=1)
Other	2 (1%)	- Suicide attempt (overdose) (n=1)* - Shortness of breath, shooting pains in limbs, hot flushes, high temperature and hospitalised overnight (withdrawal symptoms) (n=1)*	3 (1%)	- Overdose (n=2) - Prostate cancer diagnosis (n=1)
Total	**32 (10%)**		**20 (7%)**	

% Out of total randomised by arm

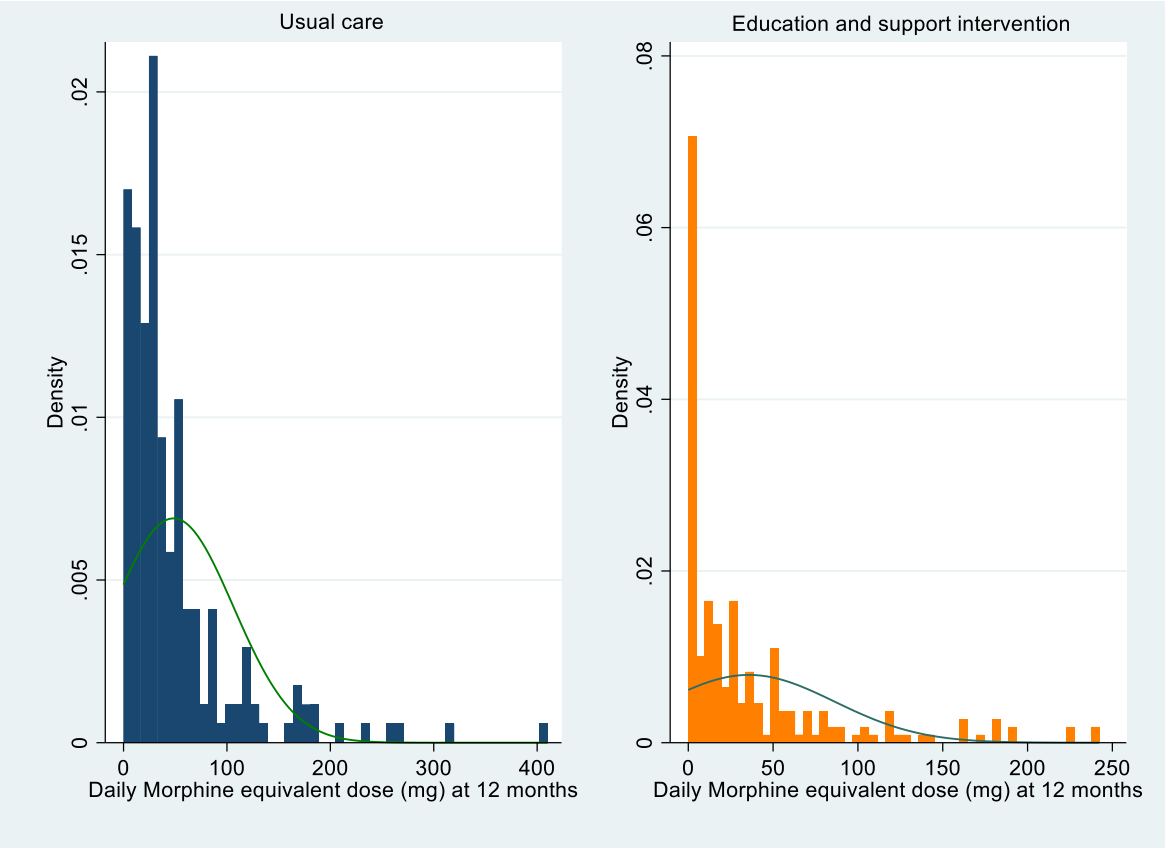
**Possibly or probably related to trial*

*** All deaths were deemed unrelated to the study intervention.*

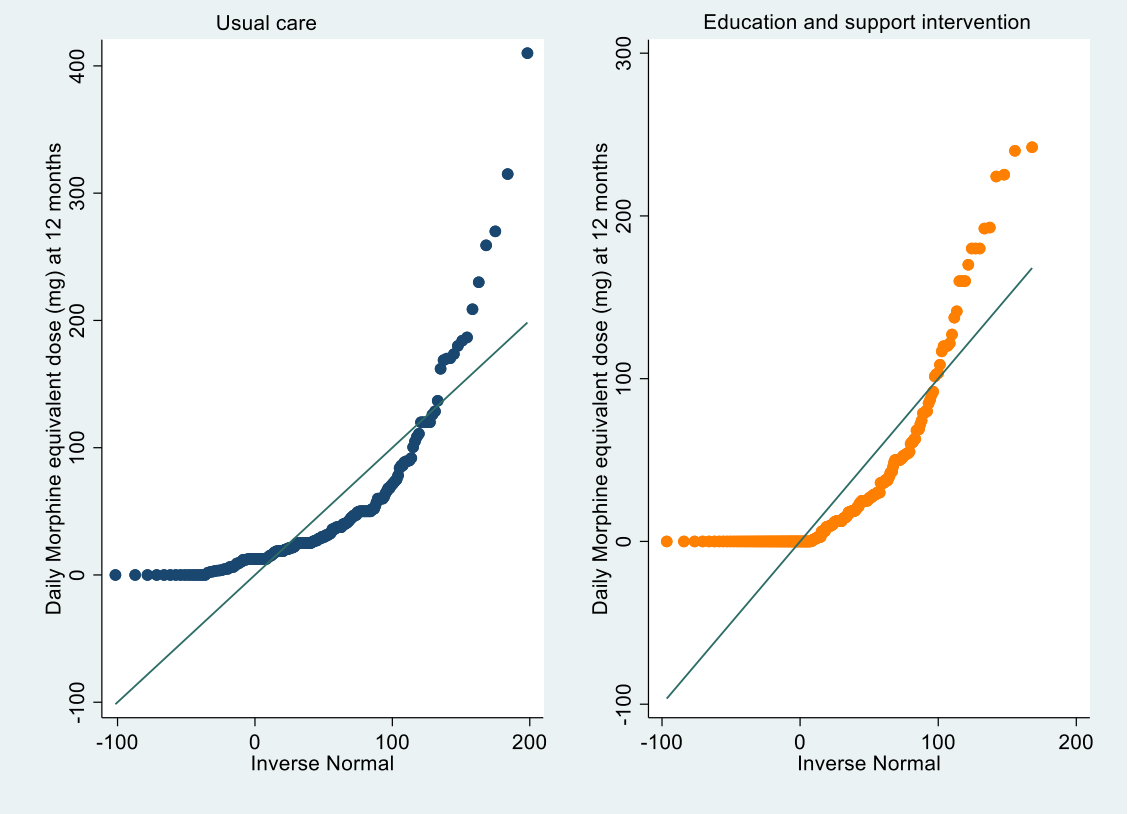
Section 6: Modelling Decision Change for Opioid use Primary Outcome

Exploratory analysis of the daily morphine equivalent dose (MED) at 12 months showed a zero-inflated and positively skewed distribution (eFigure 2 and eFigure3), with Shapiro-Wilk test statistic 0.75 ($p<0.001$). Removal of the zero values was not an option, as the aim of the trial was to taper participants off opioids, therefore the zero values were the aim of the trial and true values. There are numerous non-parametric ways to handle non-normal continuous data (log transform, poisson, negative binomial), however the literature of how to handle non-parametric zero-inflated data was scarce. The best option was to use a Gamma GLM model, however, this method fitted the positive skewness, but still did not account for the inflation of zero-values. After discussion with an independent statistician, Data monitoring committee, Trial steering committee and I-WOTCH Trial management group, it was decided to categorise the continuous opioid use measure into two groups: MED=0 and MED>0, i.e. participants who have fully tapered off opioids, and those still taking opioids. This was discussed to be the most clinically meaningful interpretation of the data, and also solved the zero-inflation modelling issues. It was therefore decided to model the opioid use outcome using a mixed effects partially nested logistic model, using individual clusters of size 1 for the usual care arm, with the same adjustments as previously defined in the SAP, in order to account for the clustering in the intervention arm.[4, 5] We also performed a sensitivity analysis on the opioid use outcome, modelling it as a continuous outcome using the same modelling as for the PROMIS-8A outcome (eTable 28). Included is also a post-hoc analysis of the participants still taking opioids at 12 months (MED>0) using a Gamma GLM model (eTable 29).

eFigure 1: Histogram of daily morphine equivalent dose at 12 months by treatment group. Green line shows normal distribution for reference



eFigure 2: QQ-plot of daily morphine equivalent dose at 12 months by treatment group. Blue line shows normal distribution for reference



eTable 30: Exploratory Opioid use at 12 months (continuous outcome)

	I-WOTCH Intervention	Control	TOTAL	Adjusted estimate (95% CI)*; p-value
Opioid use: Total MED of all opioid pain killers taken over the last 4 weeks (daily MED) at 12 months				-17.1 (-25.38 to -8.84); p<0.001
Median (IQR)	18 (0 to 50) [n=225]	29 (14 to 53) [n=208]	25 (6 to 50) [n=433]	

**Based on a heteroscedastic partially nested mixed-effect model with corrected degrees of freedom, adjusted for age, gender, baseline pain intensity, geographical location and baseline MED. The education support group was used as the cluster variable for the education and support intervention arm, with clusters of size 1 used for each participant in usual care.*

eTable 31: Exploratory Summary of Opioid use for participants who have not tapered at 12 months (MED>0), using Gamma GLM method

	I-WOTCH Intervention N=160	Control	TOTAL	Adjusted estimate (95% CI)*; p-value
Opioid use: Total MED of all opioid pain killers taken over the last 4 weeks (daily MED) IF taking opioids at 12 months				-0.21 (-0.35, -0.08); **p=0.002**
Median (IQR)	28.8 (14.2, 61.9) [n=160]	32.1 (18.8, 60) [n=193]	30 (17.9, 60.0) [n=353]	

***Gamma GLM model with log link, adjusted for age, gender, geographical location, baseline pain intensity and baseline opioid use (MED)*

eTable 32: Exploratory Sensitivity analysis using generalised estimating equation model on daily morphine equivalent dose at 12 months

	I-WOTCH Intervention N=160	Control	Adjusted estimate (95% CI)*; p-value
Opioid use: Total MED of all opioid pain killers taken over the last 4 weeks (daily MED) at 12 months			6.00 (3.06 to 11.74); p<0.001
Fully tapered off opioids (MED=0)[a]	65/225 (29%)	15/208 (7%)	

a Daily morphine equivalent dose (MED) over previous four weeks. Reported are those who fully tapered off opioids (MED=0mg). See eTable 2 for equivalences used

**Based on generalised estimating equation model adjusted for age, gender, geographical location, baseline pain intensity and baseline opioid use (MED), using logit link.*

Section 7: Indicative meaningful differences in I-WOTCH patient reported outcome measures

Background

The I-WOTCH trial has used a range of patient reported outcomes measures as secondary outcomes. To help set the findings on these measures in context post hoc exercise was performed, after findings were known, to decide on what might be a meaningful between group difference on each measure.

Complexities in generating and interpreting this information include; what might be deemed a worthwhile benefit is context specific driven by the nature of the disorder and by the intensity of the intervention. Such potential benefits and dis-benefits from interventions are commonly referred to as minimally clinically important differences (MCIDs). Published values for MCIDs include a mixture of estimates derived using anchor-based methods, typically what an individual participant might interpret as a minimal change from baseline, or distribution methods typically reflecting what between group difference might be worthwhile. However, there is commonly a lack of clarity as to whether what is referred to as an MCID is the worthwhile between group difference, the outcome reported in randomised controlled trials, or the minimal benefit or dis-benefit for an individual study participant. Nevertheless, despite these caveats, these data can serve as a starting point for interpreting patient reporting outcome measures.

Methods

For each of the secondary outcomes we searched PubMed in February 2023 using search terms of pain, MCID, and outcome measure of interest. For completeness we also included our joint primary outcome Patient-Reported Outcomes Measurement Information System (PROMIS) Pain Interference Short Form (8A) (PROMIS-PI-SF-8A). No published MCID data were available for pain populations when we designed the trial. Where no studies defining MCIDs were identified from studies of painful disorders for each outcome we expanded the search to cover all disorders. Once, starting from most recent studies, multiple studies of painful disorders reporting MCIDs for a particular patient reported outcomes had been identified, the search was discontinued. Where systematic review data were available these were used in preference to reporting individual studies. We then collated the headline figures from each study identified. Only if headline data were reported separately, did we identify MCID estimates derived using different approaches in our summary. We did not critically appraise any of the included studies. We also used two approaches grounded in our data. Since it is well established in the pain literature that a 10-point difference on a 0-100 scale is a useful benchmark for a worthwhile between group difference we presented 10% of the scale range. Finally, we estimated what effect sizes on each measure equated to a standardised mean difference of 0.3 which can be interpreted as a small to moderate difference.

We then presented these data to the clinical members of the study team to agree which value to use for each measure to indicate a worthwhile benefit from the I-WOTCH intervention (eTable 33).

Results

Pain interference (PROMIS 8A)

We identified nine studies, all published after the trial was started. We excluded one study from calculations because the upper end of reported values was extreme. After taking average of midpoints, we arrived at a value of 4.76. Omitting the surgical studies that tend to have higher reported MCID values the mean value was 4.1. A standardised mean difference of 0.3 came to a difference of 2.86, and ten percent of scale range to 3.6. These data were not strong enough to change the value of 3.5 we used for our original sample size calculation (eTable 33).

eTable 33: Pain interference

		Midpoint
Rheumatoid arthritis	2-3	2.5
6 item, group level meaningful change[6]		
Carpal tunnel release[7]	4.1 to 9.7	6.9
Spine clinic [8]*	3 to 24	-
Back pain physiotherapy clinic[9]	4.22	4.22
Rotator cuff repairs [10]	7.5	7.5
Adhesive capsulitis[11]	4.16	4.16
Carpal tunnel release[12]	7.8	7.8
Hand clinic[13]	4.3	4.3
Surgical cervical deformity[14]	5.5	5.5
Mean		5.36
Mean non-surgical		4.1
Mean difference if SMD=0.3		1.84
10% scale range		3.6
Target difference for trial		3.5

*outlier omitted

Pain intensity (PROMIS-3A)

We did not identify any studies specifically addressing the MCID of the Pain intensity (PROMIS-3A) measure. A standardised mean difference of 0.3 from within our data comes to 2.0. Applying the established benchmark of a 10% difference in scale range in pain measures to this measure with a scale range of 36.3 to 81.8 gives worthwhile difference of 3.53. After rounding this comes to 3.5, equivalent to a standardised mean difference of 0.57. This is the indicative value we used for a worthwhile benefit on pain intensity.

SF-12 physical and mental component scores

We identified six potentially relevant studies. Five of which were of surgical interventions. Taking the mean of the mid-points of the estimates from each of studies produced values of 8.61 and 7.11 respectively for the physical and mental component scores. These are implausibly large values for a worthwhile between group difference for a behavioural intervention, standardised mean differences were 1.06 and 0.64 respectively. We therefore based our worthwhile difference on Diaz-Arriba's study of back pain in the community.[15] After rounding 3.8 and 3.3 respectively. Equivalent to standardised mean differnces of 0.47 and 0.30 (eTable 34)

eTable 34: SF-12 physical and mental component scores

	PCS		*MCS*	
Population		'midpoint'		'midpoint'
Spinal surgery Distribution based[16]	1.9 to 12.7	14.6	-	-
Spinal surgery Anchor based[16]	6.4 to 16.5	11.45	-	-
Knee replacement[17]	4.5 (95% CI 3.9 to 5.2)	4.5	-	-
Lumbar fusion[18]	6.1 to 12.6	9.35	2.4 to 10.8	6.6
Revision lumbar fusion[19]	3.2 to 6.1	9.3	-	-
Surgery for lumbar stenosis[20]	2.5 to 12.1	7.3	7.0 to 15.9	11.45
Low back pain [15]	>3.77	3.77	>3.29	3.29
Mean		*8.61*		7.11
Mean surgical		*9.42*		9.0
10% scale range		10		10
Mean difference if SMD=0.3		2.43		3.33
Indicative meaningful difference		3.8		3.3

Pittsburgh sleep quality index

We found one study empirically estimating the MCID for Pittsburgh Sleep Quality Index, as 4.4, in a population undergoing rotator cuff repair.[21] One other study reported a personal communication from D Buysse the developer of the Pittsburgh Sleep Quality Index that the MCID was 3.0.[22] This later value has been used to define the MCID of the Pittsburgh Sleep Quality Index in Multiple subsequent studies. A standardised mean

difference of 0.3 equates to 1.26, and 10% of scale range is.2.1. Based on its widespread use in other studies we selected 3.0 as an indicative meaningful difference for this study.

Hospital Anxiety and Depression Scale (HADS)

We did not identify any relevant studies for Hospital Anxiety and Depression Scale from the pain literature. We identified five studies from the cardiorespiratory literature. Estimates for the MCID ranged from 0.81 to 5.2 for anxiety and 0.5 to 5.6 for depression. The mean of midpoint estimates were 1.8 and 1.81 respectively. Scores of 1.53 and 1.38 respectively equate to a standardised mean difference of 0.3 in our data. The study by Lemay stood out because as well as reporting empirical data they had conducted a Delphi exercise to establish final MCID figures of 1.7 for both anxiety and depression. [23] This equates to a standardised mean difference of 0.33 for anxiety and 0.37 for depression. These values have been used in sample estimations for several subsequent studies. We have selected these MCIDS as indicative meaningful differences for our study.

eTable-35: Hospital Anxiety and Depression Scale

	Anxiety		*Depression*	
Population		'midpoint'		'midpoint'
Bronchiectasis[24]	2 (range 1-4)	2	2 (Range 1-3)	2
Cardiovascular disease[23]	0.81 to 5.21	1.7	0.5 to 5.57	1.7
COPD[25]	1.3 to 1.8	1.55	1.5 to 1.7	1.6
COPD[26]	Around 1.5	1.5	Around 1.5	1.5
Survivors of acute respiratory failure[27]	2.0 to 2.5	2.25	2.0 to 2.5	2.25
Mean		*1.80*		1.81
10% scale range		2.1		2.1
Mean difference if SMD=0.3.3		1.53		1.38
Indicative meaningful difference		1.7		1.7

Pain Self efficacy

We found one systematic review that had identified two studies estimating the MCID for the PSEQ-10, used in the I-WOTCH study, in people with low back pain.[28] Estimates of the MCID ranged from 5.5 to 8.5 (mid-point 7.0). A standardised mean difference of 0.3 in our data is 3.95, 10% of scale range is 6.0. For pain self-efficacy we selected 7.0 as our indicative meaningful change, the midpoint of the empirical estimates of the MCID, a standardised mean difference of 0.53.

EQ-5D

We identified 19 studies of the MCID for the EQ-5D, including a systematic review of 17 studies in hip replacement.[29] As with other PROMS the sizes of the calculated MCIDS were substantially larger of the surgical studies than non-surgical studies (eTable 36). We therefore focussed on the 10 non-surgical studies, nine of which reported in EQ-5D utility score, one of which was itself a secondary analysis of eight previous studies.[30] One study reported an MCID for the EQ-5D visual analogue score. Some caution is needed when interpreting this result since the included studies have variously used the EQ-5D-3L and the EQ-5D-5L and have been carried out in different jurisdictions that use different value sets to score EQ-5D measures. As the authors of one included study observed discordance abounds when interpreting these data.[29] The mean MCID in non-surgical studies for the EQ-5D was 0.072, 10% of score range is 0.1, and a standardised mean difference of 0.3 was 0.08. Rounding the EQ-5D score gives an indicative meaningful difference of 0.07, equivalent to a standardised mean difference of 0.25.

Only one non-surgical study reported on the EQ-5D visual analogue score, a measure commonly not reported in studies using EQ-5D-3L or -5L.[31] They reported an MCID of 8.0. A standardised mean difference of 0.3 equates to a difference on visual analogue score of 6.4. For consistency with the indicative meaningful difference for the main EQ-5D score we have set the value for the visual analogue score as 7.0.

eTable-36: EQ-5D

	Utility		*VAS*	
Population		'midpoint'		'midpoint'
Hip replacements	IQR 0.18 to 0.36	0.27	IQR 12-23	20
Systematic review[29]	17 studies	(median)	16 studies	(median)
Spinal Surgery[18]	0.14 to 0.24	0.19	-	-
Spinal Surgery [32]	0.13 to 0.18 22%	0.16	-	-
Spinal surgery[33]	0.129	0.129	-	-
Spinal surgery[34]	0.19	0.19	-	-
Spinal Surgery[19]	0.35 (QALY)	0.35	-	-
Spinal Surgery[35]	0.15 to 0.54	0.35	-	-
Spinal Surgery[36]	0.24 (QALY)	0.24		
Hip/Knee osteoarthritis Surgical[37]	0.32	0.32		
Knee replacement[38]	0.085 (95% CI 0.042 to 0.127)	0.085	6.41 (95% CI 3.49 to 9.32)	6.41

Lumbar spondylolisthesis[39]	0.2	0.2	-	-
Older adults with falls. Intervention[40]	0.028 to 0.059	0.044	-	-
Older adults with falls. control[40]	0.007 to 0.051	0.029	-	-
Chronic obstructive pulmonary disease [41]	0.028 (range 0.017 to 0.033)	0.028	-	-
Hip/Knee osteoarthritis Non-surgical[37]	0.07	0.07		
Chronic obstructive pulmonary disease[31]	-	-	8	8
Multiple sclerosis[42]	0.050 to 0.084	0.067	-	-
Post traumatic Stress Disorder Anchor based[43]	0.05 to 0.08	0.07	-	-
Post traumatic Stress Disorder Distribution based[43]	0.04 to 0.10	0.07	-	-
Secondary analysis of eight previous studies[30]	Mean 0.074 (range 0.011 to 0.140)	0.074	-	-
Mean non-surgical studies		*0.072*		8
10% scale range		0.1		10
Mean difference if SMD=0.3		0.08		6.4
Indicative meaningful difference		0.07		7

Short Opioid Withdrawal Score

We did not find any studies reporting on MCIDs or meaningful differences in the Short Opioid Withdrawal Score. It is conceptually different from our other measures in that is designed to assess opioid withdrawal side effects meaning anchor based approaches to estimating the MCID based on the minimum within person improvement are not appropriate. Estimating an MCID based on a standardised mean difference is also problematic since the baseline value for the standard deviation is in a group not withdrawing from opioids. We have therefore set our indicative meaningful difference as 10% of the scale range, i.e., 3.0.

eTable 37: Summary of findings

Measure	Indicative meaningful difference
PROMIS 8A Pain interference	3.5
PROMIS 3A pain intensity	3.5
SF-36 mental component score	3.3
SF-36 physical component score	3.8
Pittsburgh Sleep Quality Index	3.0
HADS Anxiety	1.7
HADs Depression	1.7
Pain Self-efficacy	7.0
EQ-5D-5L index score	0.07
EQ-5D VAS	7.0
ShOWs	3

Discussion

There are many weaknesses in the approach used here. Few, if an of the MCIDs identified are directly relevant to our population of interest, and no quality appraisal has been made of the original papers. No effort has been made, except where the original authors have done this in their headline results, to separate out results derived from anchor based and distribution based methods. Crucially individual papers have not been examined in sufficient detail to identify those estimates that are relevant to a within person change, and those that are relevant to what would be a worthwhile between group difference for this trial. Our view is that what might be a worthwhile between group, on each of these measures is context specific, both for the population of interest and the interventions being compared. A more robust approach would have been to do this work a priori and involve various stakeholders, including patients in assessing what we might consider to be a meaningful between group difference. Just one included study, assessing MCIDs for HADS in people with cardiovascular disease used any external consensus to arrive at its conclusions.[23] Just one study considered what a minimally important deterioration might be, all other just considered minimally important improvements.[37] Nevertheless this work does give some benchmarks to allow us to present some indicative values for what might be a worthwhile benefit in each measure. For presentational purposes within the paper we refer to these indicative meaningful differences as indicative MCIDs.

References

1. Wolfe, F., et al., *The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity.* Arthritis Care Res (Hoboken), 2010. **62**(5): p. 600-10.
2. Carnes, D., et al., *Chronic musculoskeletal pain rarely presents in a single body site: results from a UK population study.* Rheumatology (Oxford), 2007. **46**(7): p. 1168-70.
3. Cella, D., Gershon, R, Bass, M, Rothrock, N. *Assessment Centre*. 2013. Accessed July 15, 2021 Available from: https://www.assessmentcenter.net.
4. Baldwin, S.A., et al., *Evaluating models for partially clustered designs.* Psychol Methods, 2011. **16**(2): p. 149-65.
5. Candlish, J., et al., *Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study.* BMC Med Res Methodol, 2018. **18**(1): p. 105.
6. Beaumont, J.L., et al., *Meaningful Change Thresholds for Patient-Reported Outcomes Measurement Information System (PROMIS) Fatigue and Pain Interference Scores in Patients With Rheumatoid Arthritis.* J Rheumatol, 2021. **48**(8): p. 1239-1242.
7. Bernstein, D.N., et al., *Minimal Clinically Important Differences for PROMIS Physical Function, Upper Extremity, and Pain Interference in Carpal Tunnel Release Using Region- and Condition-Specific PROM Tools.* J Hand Surg Am, 2019. **44**(8): p. 635-640.
8. Hung, M., et al., *What Are the MCIDs for PROMIS, NDI, and ODI Instruments Among Patients With Spinal Conditions?* Clin Orthop Relat Res, 2018. **476**(10): p. 2027-2036.
9. Khutok, K., et al., *Responsiveness of the PROMIS-29 Scales in Individuals With Chronic Low Back Pain.* Spine (Phila Pa 1976), 2021. **46**(2): p. 107-113.
10. Alben, M.G., et al., *Minimal clinically important difference (MCID) and substantial clinical benefit (SCB) of upper extremity PROMIS scores following arthroscopic rotator cuff repairs.* Knee Surg Sports Traumatol Arthrosc, 2023.
11. Alben, M.G., et al., *Minimal clinically important difference (MCID), substantial clinical benefit (SCB), and patient acceptable symptomatic state (PASS) of upper extremity PROMIS scores in idiopathic adhesive capsulitis.* Clin Rheumatol, 2023. **42**(2): p. 579-589.
12. Jiang, E.X., et al., *What are the Minimum Clinically Important Difference Values for the PROMIS and QuickDASH After Carpal Tunnel Release? A Prospective Cohort Study.* Clin Orthop Relat Res, 2022.
13. Lee, D., et al., *Responsiveness of the Brief Michigan Hand Outcomes Questionnaire and Patient-Reported Outcomes Measurement Information System Pain Interference in the Context of Patient Mental Health.* J Hand Surg Am, 2022.
14. Passias, P.G., et al., *Establishing the minimal clinically important difference for the PROMIS Physical domains in cervical deformity patients.* J Clin Neurosci, 2022. **96**: p. 19-24.
15. Diaz-Arribas, M.J., et al., *Minimal Clinically Important Difference in Quality of Life for Patients With Low Back Pain.* Spine (Phila Pa 1976), 2017. **42**(24): p. 1908-1916.
16. Nie, J.W., et al., *Establishing minimum clinically important difference for patient-reported outcome measures in patients undergoing lateral lumbar interbody fusion.* Acta Neurochir (Wien), 2023. **165**(2): p. 325-334.
17. Clement, N.D., D. MacDonald, and A.H. Simpson, *The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty.* Knee Surg Sports Traumatol Arthrosc, 2014. **22**(8): p. 1933-9.
18. Parker, S.L., et al., *Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis.* Spine J, 2012. **12**(12): p. 1122-8.
19. Parker, S.L., et al., *Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease.* J Neurosurg Spine, 2012. **16**(1): p. 61-7.
20. Parker, S.L. and M.J. McGirt, *Determination of the minimum improvement in pain, disability, and health state associated with cost-effectiveness: introduction of the concept of minimum cost-effective difference.* Neurosurgery, 2012. **71**(6): p. 1149-55.
21. Longo, U.G., et al., *Minimal Clinically Important Difference and Patient Acceptable Symptom State for the Pittsburgh Sleep Quality Index in Patients Who Underwent Rotator Cuff Tear Repair.* Int J Environ Res Public Health, 2021. **18**(16).
22. Hughes, C.M., et al., *Acupuncture and reflexology for insomnia: a feasibility study.* Acupunct Med, 2009. **27**(4): p. 163-8.
23. Lemay, K.R., et al., *Establishing the Minimal Clinically Important Difference for the Hospital Anxiety and Depression Scale in Patients With Cardiovascular Disease.* J Cardiopulm Rehabil Prev, 2019. **39**(6): p. E6-E11.

24. Wynne, S.C., et al., *Anxiety and depression in bronchiectasis: Response to pulmonary rehabilitation and minimal clinically important difference of the Hospital Anxiety and Depression Scale.* Chron Respir Dis, 2020. **17**: p. 1479973120933292.
25. Smid, D.E., et al., *Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary Rehabilitation: A Prospective Analysis.* J Am Med Dir Assoc, 2017. **18**(1): p. 53-58.
26. Puhan, M.A., et al., *The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease.* Health Qual Life Outcomes, 2008. **6**: p. 46.
27. Chan, K.S., et al., *Distribution-based estimates of minimal important difference for hospital anxiety and depression scale and impact of event scale-revised in survivors of acute respiratory failure.* Gen Hosp Psychiatry, 2016. **42**: p. 32-5.
28. Dube, M.O., P. Langevin, and J.S. Roy, *Measurement properties of the Pain Self-Efficacy Questionnaire in populations with musculoskeletal disorders: a systematic review.* Pain Rep, 2021. **6**(4): p. e972.
29. Deckey, D.G., et al., *Discordance Abounds in Minimum Clinically Important Differences in THA: A Systematic Review.* Clin Orthop Relat Res, 2022.
30. Walters, S.J. and J.E. Brazier, *Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D.* Qual Life Res, 2005. **14**(6): p. 1523-32.
31. Zanini, A., et al., *Estimation of minimal clinically important difference in EQ-5D visual analog scale score after pulmonary rehabilitation in subjects with COPD.* Respir Care, 2015. **60**(1): p. 88-95.
32. Nakarai, H., et al., *Minimal clinically important difference in patients who underwent decompression alone for lumbar degenerative disease.* Spine J, 2022. **22**(4): p. 549-560.
33. Vanhorn, T.A., Z.O. Knio, and T.J. O'Gara, *Defining a Minimum Clinically Important Difference in Patient-Reported Outcome Measures in Lumbar Tubular Microdecompression Patients.* Int J Spine Surg, 2020. **14**(4): p. 538-543.
34. Burgstaller, J.M., et al., *Evaluating the Minimal Clinically Important Difference of EQ-5D-3L in Patients With Degenerative Lumbar Spinal Stenosis: A Swiss Prospective Multicenter Cohort Study.* Spine (Phila Pa 1976), 2020. **45**(18): p. 1309-1316.
35. Parker, S.L., et al., *Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis.* J Neurosurg Spine, 2011. **14**(5): p. 598-604.
36. Parker, S.L., et al., *Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article.* J Neurosurg Spine, 2013. **18**(2): p. 154-60.
37. Bilbao, A., et al., *Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness.* Qual Life Res, 2018. **27**(11): p. 2897-2908.
38. Yapp, L.Z., et al., *Meaningful values of the EQ-5D-3L in patients undergoing primary knee arthroplasty.* Bone Joint Res, 2022. **11**(9): p. 619-628.
39. Asher, A.L., et al., *Defining the minimum clinically important difference for grade I degenerative lumbar spondylolisthesis: insights from the Quality Outcomes Database.* Neurosurg Focus, 2018. **44**(1): p. E2.
40. Jehu, D.A., et al., *Establishing the minimal clinically important difference of the EQ-5D-3L in older adults with a history of falls.* Qual Life Res, 2022. **31**(11): p. 3293-3303.
41. Bae, E., et al., *Validity of EQ-5D utility index and minimal clinically important difference estimation among patients with chronic obstructive pulmonary disease.* BMC Pulm Med, 2020. **20**(1): p. 73.
42. Kohn, C.G., et al., *Estimating a minimal clinically important difference for the EuroQol 5-Dimension health status index in persons with multiple sclerosis.* Health Qual Life Outcomes, 2014. **12**: p. 66.
43. Le, Q.A., et al., *Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT).* Health Qual Life Outcomes, 2013. **11**: p. 59.