BJA

British Journal of Anaesthesia, xxx (xxx): xxx (xxxx)

doi: 10.1016/j.bja.2023.10.020
Advance Access Publication Date: xxx
Clinical Investigation

CLINICAL INVESTIGATION

Responsiveness of multiple patient-reported outcome measures for acute postsurgical pain: primary results from the international multi-centre PROMPT NIT-1 study

Jan Vollert[1,2,†], Daniel Segelcke[1,†], Claudia Weinmann[3], Kathrin Schnabel[1], Fabian Fuchtmann[1], Daniela C. Rosenberger[1], Marcus Komann[3], Timo Maessen[1], Lena Sauer[1], Eija Kalso[4], Dominique Fletcher[5], Patricia Lavand'homme[6], Ulrike Kaiser[7], Hiltrud Liedgens[8], Winfried Meissner[3] and Esther M. Pogatzki-Zahn[1,*]

[1]Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Muenster, Muenster, Germany, [2]Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK, [3]Jena University Hospital, Department of Anesthesiology and Intensive Care Medicine, Friedrich Schiller University Jena, Jena, Germany, [4]Department of Anaesthesia, Intensive Care and Pain Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland, [5]Anaesthesia and Intensive Care Department, Ambroise Paré Hospital, APHP, Université Paris-Saclay, Boulogne, France, [6]Department of Anesthesiology and Acute Postoperative & Transitional Pain Service, Cliniques universitaires Saint Luc - University Catholic of Louvain, Brussels, Belgium, [7]Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Schleswig-Holstein, Campus Luebeck, Luebeck, Germany and [8]Grünenthal GmbH, Aachen, Germany

*Corresponding author. E-mail: pogatzki@anit.uni-muenster.de
[†]These authors contributed equally.

Abstract

Background: Postsurgical outcome measures are crucial to define the efficacy of perioperative pain management; however, it is unclear which are most appropriate. We conducted a prospective study aiming to assess sensitivity-to-change of patient-reported outcome measures assessing the core outcome set of domains *pain intensity* (at rest/during activity), *physical function*, *adverse events*, and *self-efficacy*.

Methods: Patient-reported outcome measures were assessed preoperatively, on day 1 (d1), d3, and d7 after four surgical procedures (total knee replacement, breast surgery, endometriosis-related surgery, and sternotomy). Primary outcomes were sensitivity-to-change of patient-reported outcome measures analysed by correlating their changes (d1–d3) with patients' global impression of change and patients' specific impression of change items as anchor criteria. Secondary outcomes included identification of baseline and patient characteristic variables explaining variance in change for each of the scales and descriptive analysis of various patient-reported outcome measures from different domains and after different surgeries.

Results: Of 3322 patients included (18 hospitals, 10 countries), data from 2661 patients were analysed. All patient-reported outcome measures improved on average over time; the median calculated sensitivity-to-change for all patient-reported outcome measures (overall surgeries) was 0.22 (range: 0.07–0.31, scale: 0–10); all changes were independent of baseline data or patient characteristics and similar between different procedures.

Conclusions: Pain-related patient-reported outcome measures have low to moderate sensitivity-to-change; those showing higher sensitivity-to-change from the same domain should be considered for inclusion in a core outcome set of patient-reported outcome measures to assess the effectiveness and efficacy of perioperative pain management.

Keywords: acute pain; core outcome set; pain assessment; psychometric properties; sensitivity-to-change; surgery

Received: 23 May 2023; **Accepted:** 26 October 2023

Editor's key points

- Patient-reported outcome measures refer to a patient's health condition, reported by them and without interpretation by a clinician or anyone else.
- This article explores how pain-related patient-reported outcome measures are modified according to changes in analgesic status after selected surgical procedures.
- For postoperative pain status, patient-reported outcome measures should reflect multiple, clinically and patient-relevant domains.

If acute postsurgical pain is not managed optimally, it causes suffering, impaired recovery after surgery, and long-term complications such as chronic postsurgical pain.[1–3] Patient-reported outcome measures (PROMs) are necessary to assess the highly individual experience of pain and pain-related symptoms. At its simplest, such a PROM captures *pain intensity* levels on rating scales such as the numeric rating (NRS) or visual analogue scale.[4] *Pain intensity* can differ at rest and during activity, constituting a distinct burden. The most appropriate PROMs to measure pain intensity are unknown. Although widely used, there is increasing evidence that rating scales are not sufficient to capture all relevant aspects of pain and successful pain management after surgery.[5–7] For pain-related outcome domains beyond intensity, there is a great inconsistency of domains and PROMs used across perioperative pain trials.[8,9] Harmonisation of outcome assessment is essential to increase the comparability of results, enhance transparency of reporting, and improve treatment decisions in clinical routine. To develop a harmonised core outcome set (COS), two steps are needed: agreement on the relevant domains, and the definition of ideal PROMs to assess each domain.[9–11]

Recently, the Innovative Medicines Initiative (IMI)-PainCare defined with a multidisciplinary panel including patients a COS of domains to assess postoperative pain (including *pain intensity at rest* and *pain during activity*, *physical function*, *adverse events*, and *self-efficacy*) for research and clinical practice related to postoperative pain in an international consensus process (Fig. 1a).[10] The next important step is to decide on suitable PROMs for each domain; here psychometric properties such as content validity and reliability and responsiveness (or sensitivity-to-change) of PROMs for each of the consented domains is part of this process.[9,10] Within IMI-PainCare, we searched for PROMs related to the domains and assessed their content validity and, if required, reliability (compare Fig. 1b). Responsiveness of a PROM needs to be determined as well; responsiveness (or sensitivity-to-change) is defined as its ability to detect a change in the patients' clinical condition over time on group level,[11,12] which no study has evaluated yet for pain after surgery. Sensitivity-to-change of a PROM is an important measure of longitudinal validity; the higher the responsiveness (e.g. measured as a correlation to a gold standard), the better the longitudinal validity of a PROM. In this line, responsiveness (in a study population similar to the study sample) of all PROMs used in a clinical trial was taken up recently by the CONSORT-Outcomes 2022 extension of the CONSORT 2010 statement as one outcome-specific item that should be reported in all published clinical trial reports.[13] Here, we performed a prospective multicentre international study related to PROMs for postoperative pain treatment non-interventional trial 1 (PROMPT NIT-1) evaluating the sensitivity-to-change of PROMs related to outcome domains agreed in the consensus process[10] after four surgical procedures (total knee arthroplasty, breast surgery, surgery related to endometriosis, and sternotomy). Those PROMs showing higher sensitivity-to-change from the same domain are better suited than those with lower sensitivity-to-change for being included in a COS of PROMs to assess the effectiveness and efficacy of perioperative pain management.

Methods

Below is a short overview of the methods. A detailed description can be found in the Supplementary material.

Ethics approval/Registration

The PROMPT NIT-1 study was designed as a prospective, multicentre, observational cohort study collecting PROMs and clinical data from patients before and after surgery (registered on February 8, 2019, on clinicaltrials.gov, ID: NCT03834922). It was conducted in accordance with the latest version of the Declaration of Helsinki and approved by the ethics committee of Jena University Hospital, Jena, Germany (Ref. 2019-1298-Bef, dated February 6, 2019) and the local ethics committees of all participating hospitals.

Overall study design

This study was part of the subproject PROMPT (WP2) of IMI-PainCare (https://www.imi-paincare.eu/PROJECT/PROMPT/). As in all IMI projects, research topics are defined as 'pre-competitive' and identified by the participating European Federation of Pharmaceutical Industries and Associations (EFPIA) partners. Research questions are then designed and carried out in co-leadership between academic and industry partners. All work is co-funded by the EU (under Horizon 2020) and the participating EFPIA partners (through in-kind contributions). The overall aim of PROMPT WP2 is to enable consensus on a COS of PROMs related to acute postoperative pain by using a stepwise approach following COMET (Core Outcome Measures in Effectiveness Trials) and COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) guidance;[11,12] the entire process is shown in Figure 1. This study aimed to investigate the sensitivity-to-change of specific PROMs in the assessment of acute postsurgical pain outcomes in the routine care of patients after four different surgical procedures (total knee arthroplasty, breast surgery, endometriosis surgery, and sternotomy).

In the present study (Fig. 1c), we investigated PROMs for four domains (with two subdomains) previously consented on as COS of domains: (1) pain intensity with the subdomains (a) *pain intensity at rest* and (b) *pain intensity during activity*; (2) *physical function*; (3) *adverse events*; and (4) *self-efficacy*.[10] The domains were identified in a multistep process beforehand (Fig. 1a).[10] PROMs for each domain were completed in PROMPT NIT-1 by patients three times, once on day 1 (d1), once on day 3 (d3), and once on day 7 (d7) after surgery, along with anchors (the patients' global impression of change [PGIC] item and patients' specific impression of change [PSIC] item) to assess the responsiveness of PROMs on d3 (day 3: primary outcome) and d7 (day 7: secondary outcome).

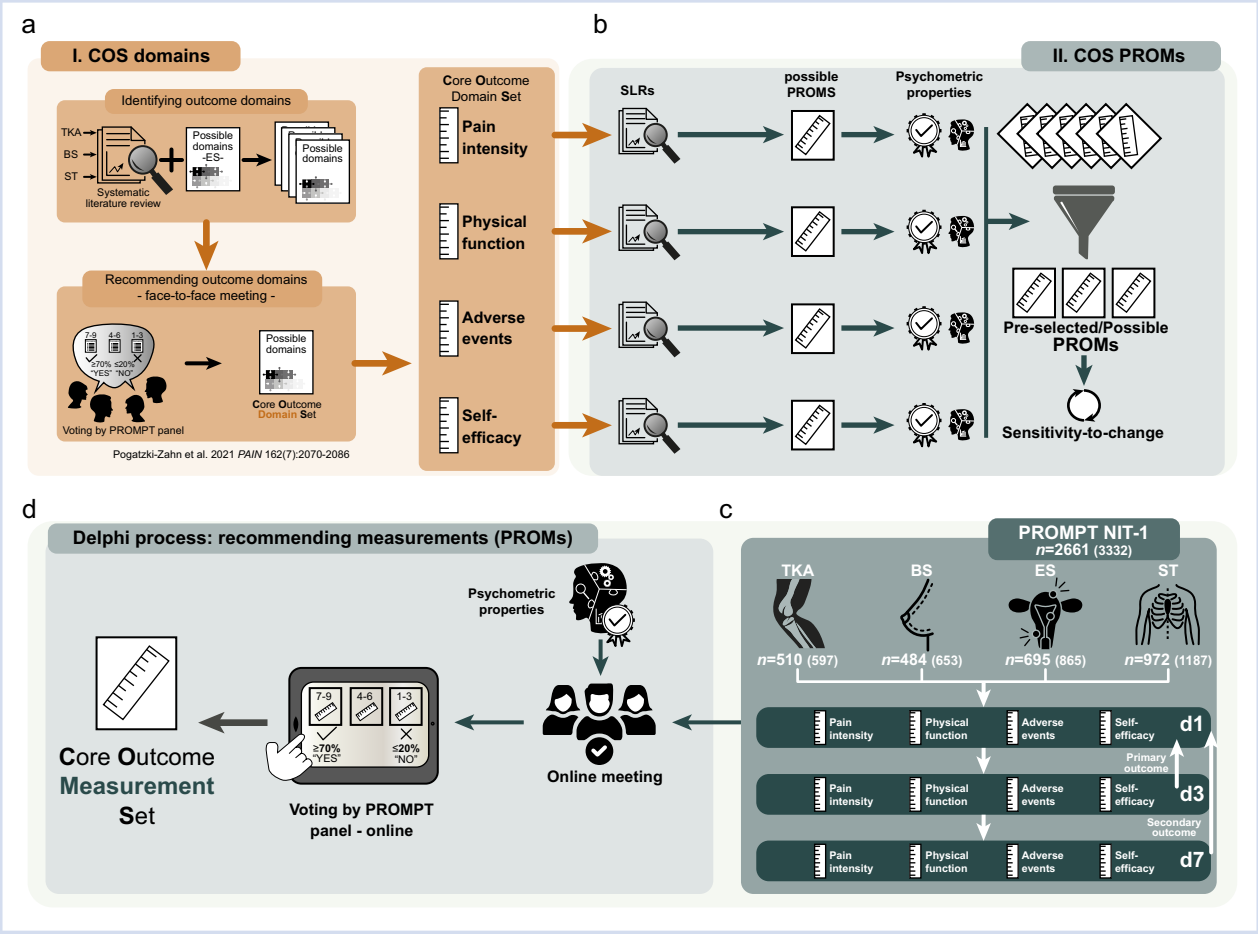

Fig 1. Study design. The study enables consensus on a core outcome set (COS) of patient-related outcome measures (PROMs) related to acute postoperative pain. (a) The IMI-Pain Care PROMPT consensus panel defined and recommended overarching outcome domains for all the included surgeries. The domains consist of pain intensity with the subdomains 'in general', 'at rest' and 'during activity', physical function, adverse events, and self-efficacy. (b) Systematic literature reviews (SLRs) were conducted to identify relevant, reliable, and valid PROMs for each of the specific outcome domains. Based on SLRs, instruments have been identified meeting the requirements (psychometric properties) for most of the domains. We adapted them where necessary for acute pain after surgery. (c) Identified PROMs included as patient questionnaires to be completed three times (day [d]1, d3, d7) after surgery within the large multicenter trial PROMPT NIT-1, together with anchors (the patients' global impression of change [PGIC] item and patients' specific impression of change [PSIC] items) to assess their responsiveness on d3 (primary outcome) and d7 (secondary outcome). (d) Results of these analyses will reach a consensus on a COS of PROMs that can be recommended for future standardisation of outcome assessment in clinical trials and for clinical practice to sensitively and validly assess individualised pain management after surgery. BS, breast surgery; ES, surgery related to endometriosis; NIT-1, non-interventional trial 1; PROMPT, PROMs for postoperative pain treatment; ST, sternotomy; TKA, total knee arthroplasty.

Eligibility criteria

Detailed inclusion and exclusion criteria are shown in Figure 2.

Patient-reported outcome measures and surveyed data

At baseline, after informed consent, data on patient characteristics, comorbidities, and treatment with opioids and other analgesics were collected. In addition, baseline questionnaires/PROMs were used to assess patient-relevant parameters such as quality of life (using the European Quality of Life Five Domains[14]), depression and anxiety (using the Hospital Anxiety and Depression Scale [HADS][15]), pain sensitivity (using the Pain Sensitivity Questionnaire[16]), pain expectancy, pain catastrophising (using the Pain Catastrophizing Scale [PCS][17]), preoperative pain (using the Brief Pain Inventory[18,19]), and neuropathic qualities of preoperative pain (using the Douleur Neuropathique en 4 Questions and the Neuropathic Pain Symptom Inventory[20,21]).

For d1, d3, and d7, PROMs addressing all outcome domains previously agreed in a consensus process,[10] were assessed. PROMs were pre-selected by systematic literature research for each domain. Those best fulfilling psychometric properties or at least face validity, feasibility assessments, or both were chosen by the steering committee of IMI-PainCare PROMPT to be included in PROMPT-NIT-1. PROMs related to *pain intensity* were evaluated via five questions by using an 11-point NRS based on previous recommendations[4,22,23] and the validated PAIN-OUT questionnaire[19] with the anchors 'no pain' and

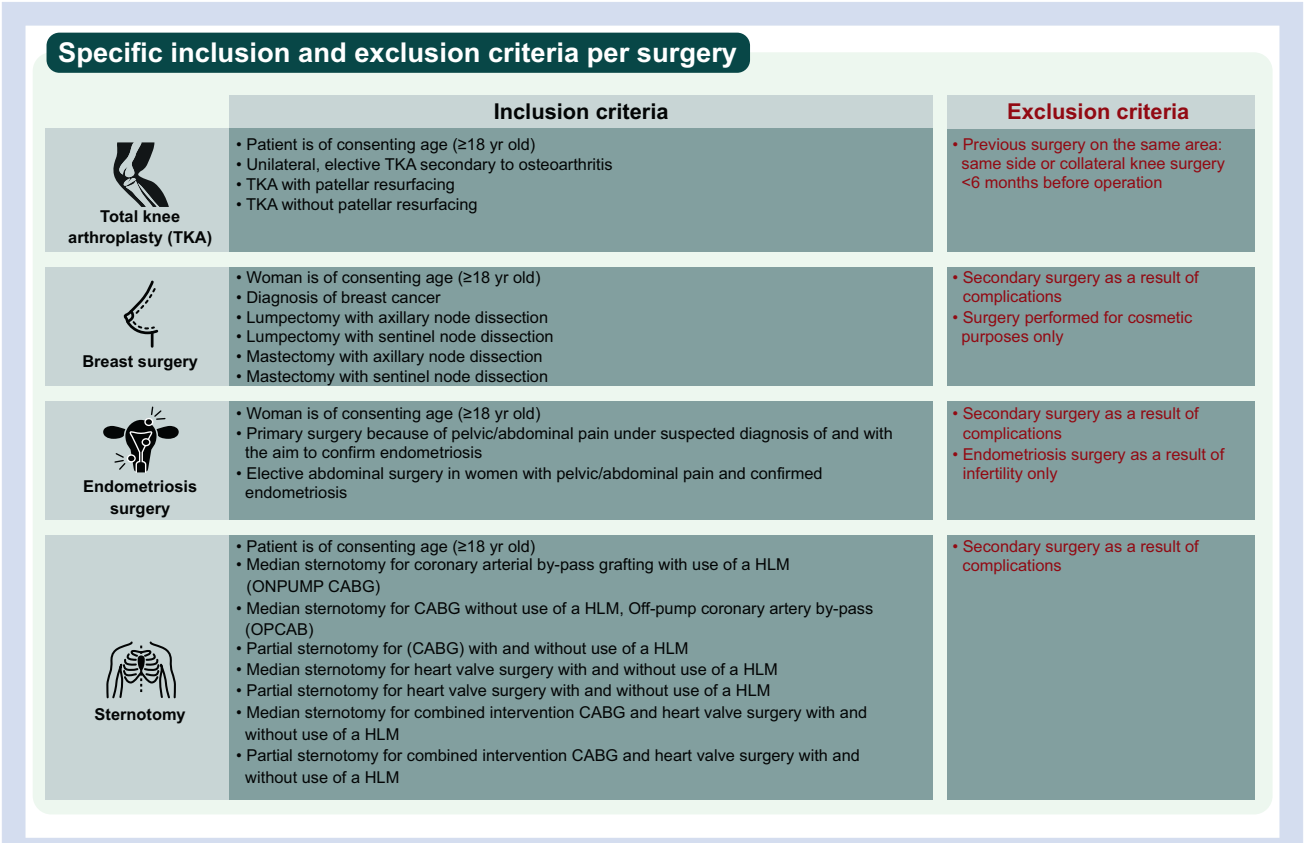

Fig 2. Specific inclusion (green) and exclusion (red) criteria per surgery for PROMPT NIT-1. For breast surgery and endometriosis-related surgery, only female participants were eligible. For the other two surgical groups, all sexes were included. HLM, heart-lung machine; NIT-1, non-interventional trial 1; PROMPT, PROMs for postoperative pain treatment; TKA, total knee arthroplasty.

'worst pain imaginable'. *Physical function* was assessed by rating *interference with physical activities because of pain* adapted from the validated PAIN-OUT questionnaire[19] on an 11-point Likert scale with the anchors 'did not interfere' and 'completely interfered'. No specific PROMs related to *self-efficacy* for the postoperative period were identified by using a systematic literature review. Thus, *self-efficacy* was assessed by 11-point Likert scale using a modified version of the Arthritis Self-Efficacy Scale.[24,30] *Adverse events* by counts were assessed using an adapted version of the validated Symptom Distress Scale,[25] adding an additional item (motor function) and skipping the categorisation after each item for feasibility reasons. For the sensitivity-to-change analysis, PROMs on perceived changes since d1 regarding *pain intensity*, *pain interference*, *self-efficacy*, *adverse events (PSIC items)*, and the PGIC since d1 were assessed on d3 and d7, using a 7-point Likert scale with the general question 'Since the first day after surgery until now, how would you describe the change in […]'. Anchors 'very much improved', 'much improved', 'minimally improved', 'no change', 'minimally worse', 'much worse', and 'very much worse'. All PROMs included in the questionnaires are shown in the Supplementary material.

Translations, where necessary, were carried out using a forward–backward procedure based on the WHO 'Process of translation and adaptation of instruments'. For more details related to the PROMs assessed, refer to the Supplementary material.

Data capturing and management

A total of 18 study sites in 10 countries participated. Cooperation agreements were concluded with each site, and local ethics approvals and GDPR approvals were obtained. Data were captured using OpenClinica (for baseline process and patient characteristic data, the patient-reported outcome surveys on d1 and d3) and LimeSurvey (for patient-reported outcome surveys from d7 on).

Primary and secondary outcome

Primary outcome was sensitivity-to-change of PROMs (assessed as change of each PROM between d1 and d3 correlated with PGIC and PSIC on d3). Secondary outcomes were sensitivity-to-change of PROMs (change of each PROM between d1 and d7 correlated with PGIC and PSIC on d7). In addition, we investigated if baseline characteristics (pain catastrophising, anxiety, depression, pre-existing pain, pre-study consumption of opioids) or patient characteristics (age, gender) affected the correlation of change in PROMs and PSIC/PGIC by comparing correlation coefficients across subgroups of the above mentioned variables. Finally, European-wide data on a wide

range of PROMs up to 7 days after surgery related to the domains *pain intensity*, *physical function, adverse events,* and *self-efficacy* are described on postoperative pain quality by assessing the most important domains as recommended recently.[6]

Statistical analyses

We present descriptive data of all PROMs as medians and interquartile range (IQR), graphically displayed in violin plots. PROMs related to *pain intensity* were further analysed using NRS ≥4 as a threshold based on Gerbershagen and colleagues.[26] To confirm the sensitivity-to-change/responsiveness of each PROM, we aimed to perform a criterion approach which is in accordance with the COSMIN recommendations.[11] PGIC and PSIC on d3 (primary outcome) and d7 (secondary outcome) assessing a patient's view on global and several specific changes related to the situation on d1 were used as anchors.[27]

As none of the analyses presented here are null hypothesis testing, we refrain from reporting *P*-values.

Results

Patient characteristics

A total of 3322 participants completed the study (Fig. 1). Patients who did not fill in questionnaires on d1, d3, or both, and any incomplete datasets, were excluded, leaving 2661 patients in this analysis. Of these, 972 (37%) underwent sternotomy, 695 (26%) surgery related to endometriosis (laparoscopy, complex surgery, hysterectomy), 510 (19%) total knee arthroplasty, and 484 (18%) breast surgery (conservation or mastectomy). The mean age of all patients was 54 yr (range: 18–91 yr). Overall, 65% of patients were females, which was biased by breast cancer and endometriosis-related surgery where only female patients were included.

Overall, 44% of patients reported persistent pain before surgery (knee arthroplasty 91%, breast surgery 15%, endometriosis surgery 68%, sternotomy 16%) and 6% opioid use (knee arthroplasty 15%, breast surgery 1%, endometriosis surgery 8%, sternotomy 2%). HADS scores for anxiety and depression were elevated for 14% (knee arthroplasty 12%, breast surgery 16%, endometriosis surgery 27%, sternotomy 5%) and 6% (knee arthroplasty 10%, breast surgery 5%, endometriosis surgery 9%, sternotomy 2%) of patients, respectively. PCS scores were elevated in 11% of all patients (knee arthroplasty 16%, breast surgery 5%, endometriosis surgery 22%, sternotomy 4%, more details in Supplementary material).

Descriptive analysis of patient-reported outcome measures on day 1 and day 3

All PROMs assessed on d1 and d3 related to four domains (*pain intensity*, *interference of physical function as a result of pain*, *adverse events*, and *self-efficacy*) after the four surgical procedures are shown in Figure 3 as violin plots with a highlighted median. Measures of pain intensity (*pain at rest, pain during procedure-specific activity, pain during physiotherapy, pain average,* and *pain worst*) were reduced on d3 compared with d1. Pain intensity was higher after knee arthroplasty and endometriosis surgery than after breast surgery on all pain intensity-related PROMs, with sternotomy in-between. Interference of physical function as a result of pain was highest on d1 and decreased 2 days later; this was similar for all four surgical procedures. Interestingly, interference as a result of pain in bed and during procedure-specific activity was similar for all surgical procedures (with medians between 2 and 7 [IQRs: 3–5]), whereas impairment of physical function during physiotherapy had a median of 0 (IQRs: 2–5) on d1 and d3 for breast surgery and endometriosis surgery while showing medians between 2 and 6 (IQRs: 2–4) for knee arthroplasty and sternotomy (Fig. 3). The number of adverse events was highest on d1 and decreased on d3 for all surgical procedures, but with sternotomy. Levels of self-efficacy increased or remained stable between days. PROMs for d7 after surgery continued to improve slightly for all surgical procedures (Supplementary material).

Global impression of change

On the 7-item PGIC scale, the overall median response on d3 was 'much improved' (IQR: one item wide). Per procedure, the median response was 'much improved' for knee arthroplasty and sternotomy, and 'minimally improved' for breast surgery and endometriosis surgery.

Relation of calculated changes of patient-reported outcomes measures and reported change with specific change items

Primary outcome was the assessment of sensitivity-to-change from d3 to d1 on 11-item scales for pain intensities, interference with physical function, self-efficacy, and the presence or absence of 11 adverse events. Median changes ranged from 0 (pain at rest for breast surgery, IQR: 1) to 4 (pain during physiotherapy for endometriosis surgery, IQR: 6) for *pain intensity*, from 0 (interference with physiotherapy for breast surgery and endometriosis surgery, IQRs: 2) to 2 (interference in bed and during activity, multiple surgeries, IQRs: 2–3) for *physical function*, from 0 to 1 (IQR: 2) for number of *adverse events*, and from 0 (multiple items, multiple surgeries, IQRs: 1–3) to 1 (multiple items, endometriosis surgery, IQRs: 1–3) for *self-efficacy* (Fig. 4). PSIC items for *pain intensity*, *physical function*, *adverse events*, *self-efficacy*, and PGIC were assessed on d3 (Fig. 5). The overall median response to *pain intensity* change items on d3 ranged from 'much improved' to 'very much improved' (IQR: two items wide) with a similar range for *physical function* PROMs, *adverse events*, and *self-efficacy* (Fig. 5).

Calculated changes were correlated with corresponding specific change items (seven-item scale from 'very much worse' to 'very much improved') to assess sensitivity-to-change for each PROM (Table 1). The average correlation across surgeries for specific change items was 0.22, with a minimum of 0.07 for change in *pain intensity during physiotherapy*, and a maximum of 0.31 for *pain intensity during activity*. For specific surgical procedures, the minimum was 0.04 for change in *pain intensity during physiotherapy* for knee arthroplasty and *self-efficacy* (daily activities) for breast surgery, and the maximum was 0.43 for change in interference with specific *physical function* for endometriosis surgery.

Correlation with PGIC averaged of 0.15, with a minimum of 0.04 for change in pain during physiotherapy and a maximum of 0.25 for worst pain. This relationship was not dependent on baseline data (e.g. preoperative pain catastrophising, anxiety, depression, pre-existing pain, pre-study consumption of opioids) or patient characteristics (e.g. age, gender) (Supplementary material).

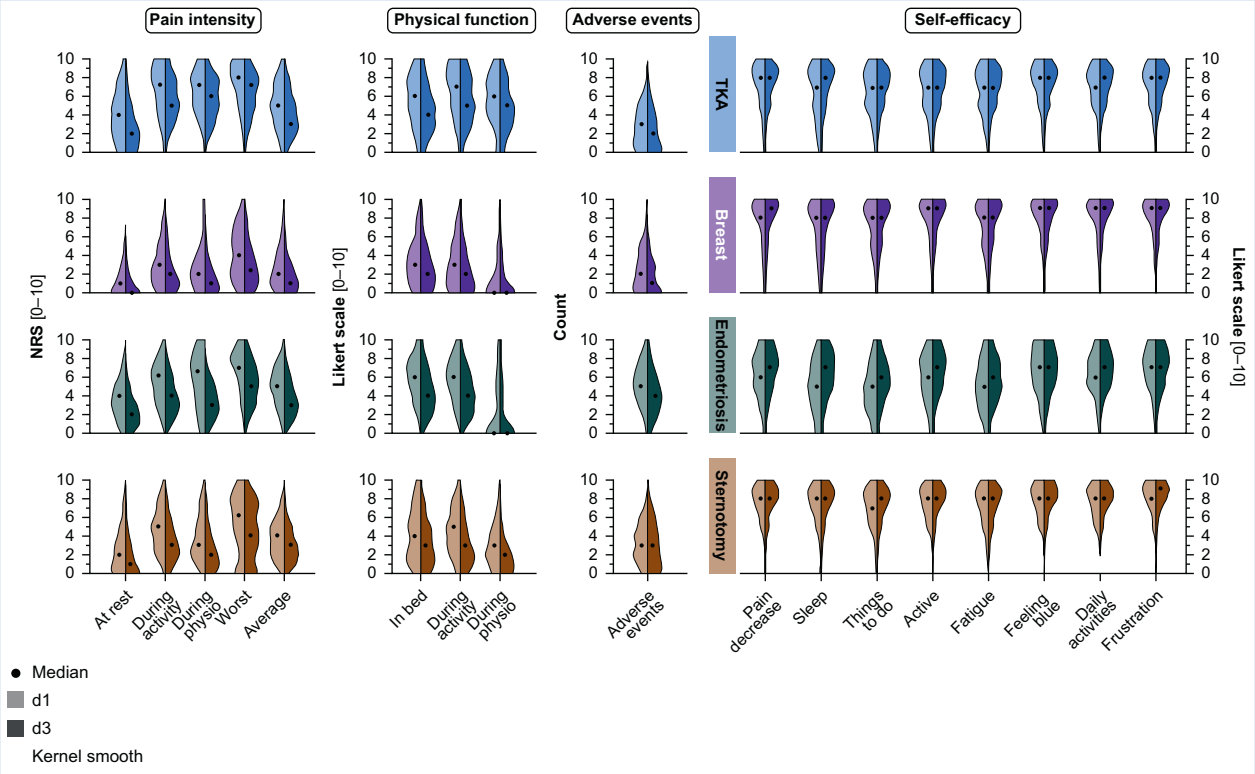

Fig 3. Descriptive analysis of patient-reported outcome measures (PROMs) on day 1 (d1) and day (d3). All PROMs assessed on d1 and d3 in relation to the five different domains (pain intensity at rest, pain intensity during activity, impairment of physical function as a result of pain, adverse events, and self-efficacy) with the different subdomains after the four surgical procedures are shown as split-violin plots (with Kernel smoothing; left side represents d1, right side represents d3) with the median highlighted (black dot) on different scales (numeric rating scale [NRS, 0–10] for pain intensity; Likert scale [0–10] for physical function and self-efficacy; numerical counting adverse events; n=510 total knee arthroplasty [TKA], n=484 breast surgery, n=695 surgery related to endometriosis, and n=972 sternotomy).

Discussion

Here, we present the primary results of the prospective international multicentre PROMPT NIT-1 study, including data from 18 sites across 10 European countries. Our overall response rate was good, with 80% of patients fully completing all three questionnaires. In 2661 patients who underwent one of four different surgical procedures, pain-related outcomes generally improved from d1 to d3 and further to d7 after surgery. This held true for measures of *pain intensity*, *physical function*, and *adverse events*, whereas levels of *self-efficacy* remain partly unchanged. The improvement is reflected in the PGIC at d3 and d7, and in the PSICs for *pain intensity*, *adverse events*, and *physical function*. Consequently, PGIC (and PSIC as secondary outcome) were shown to reflect calculated changes between days relevant for the PROMs, consistently between surgeries. Many (but not all) calculated changes of PROMs from d1 to d3 (and d1 to d7) correlated highly with PGIC and PSIC showing differences in sensitivity-to-change of PROMs assessed in our project. Individual patients' responses vary broadly within time points, often covering the full scale of instruments (11- or 7-point scales) for different scales. Thus, evaluating effects of postsurgical pain management should be based on PROMs corresponding to multiple, clinically and patient-relevant domains.

Sensitivity-to-change of investigated patient-reported outcome measures

The sensitivity-to-change of multiple PROMs had weak to moderate correlations. These correlations should not be underestimated, as they pertain to different scales (*7 vs 22* points) and overlapping, but not identical items. There were differences in the responsiveness of PROMs even within the same domain: *pain intensity during physiotherapy* showed, for example, virtually no correlation with the PSIC, whereas other PROMs related to this domain showed weak but present correlations. This is true for the analysis including all surgeries together and separate analysis related to each surgery. *Pain intensity during* (procedure-specific) *activities* (such as bending the knee for knee arthroplasty or lifting the arm proximal to breast surgery) was weak to moderately correlated with reported changes across all four surgical procedures. These PROMs may reflect important aspects relevant for the patients and are easy to interpret. Somewhat surprisingly, *pain intensity on average* and *pain intensity at rest* correlated as well weak to moderately with PGIC/PSICs. Ratings of *pain intensity at rest* are usually lower and exhibit less variance than *worst pain intensity* after surgery or *pain intensity during activity*. Recent analysis on outcome assessment after surgery have reported mainly on the highest *pain intensity* ratings, such as *worst pain intensity*.[28,29] Thus, our

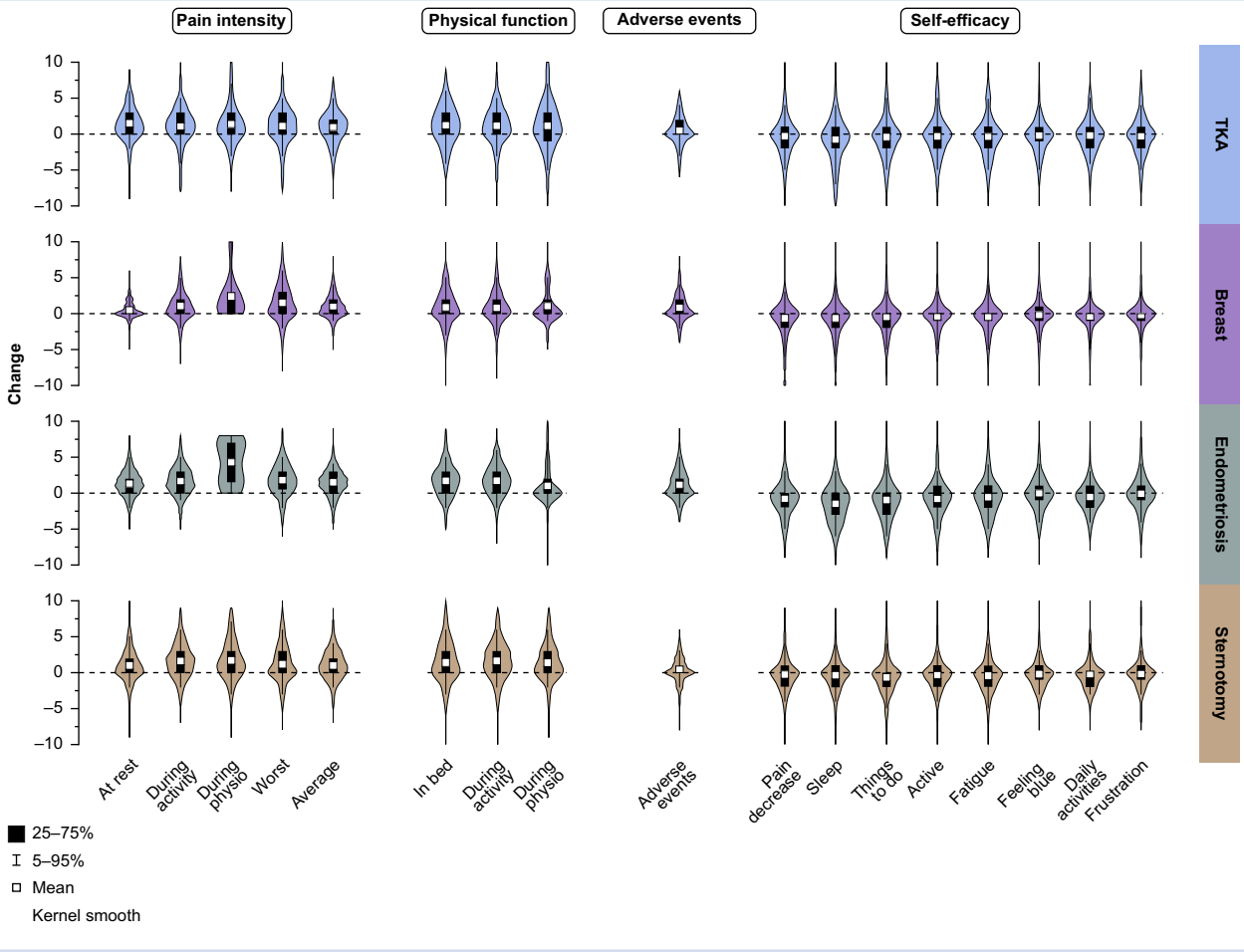

Fig 4. Calculated changes in patient-reported outcome measures (PROMs). Calculated changes from day 3 to day 1 for all PROMs are presented here as violin plots (Kernel smoothing, white dots represent mean, black boxes represent 25–75% quartile, and whiskers represent 5–95% percentile, n=510 total knee arthroplasty [TKA], n=484 breast surgery, n=695 surgery related to endometriosis, n=972 sternotomy).

data show that *pain intensity at rest*, or *pain intensity on average*, may be of higher importance to patients than thought and need to be considered in a COS to assess the patients' perspective related to a change in *pain intensity* after surgery. The reasons behind this need further investigations. Interestingly, correlations between calculated and reported changes for the domain *pain intensity* (and most other domains) were on average higher in patients undergoing endometriosis surgery—this may reflect a higher level of literacy in using questionnaires in these patients. Of course, we cannot rule out that these differences might be a methodological artifact.

PROMs related to *physical function* showed moderate responsiveness. Especially, procedure-specific *physical function* items correlated well in all surgical procedures and might serve as change-sensitive PROM in a COS assessing the efficacy and effectiveness of pain management after surgery. The challenge in estimating change in scales assessing *self-efficacy* is critically discussed and broadly refers to the item formulation (generic vs. specific).[30] Change in the number of adverse events between d1 and d3 correlated high with both PGIC and the PSIC for *adverse events*. Overall, correlations between calculated PROM changes and PSIC were higher than those between PROM changes and PGIC, as expected: the PGIC encompasses multiple domains, and our findings suggest that patients were able to separate specific from global changes.

Next steps in defining a core outcome set of patient-reported outcome measures for assessing effectiveness and efficacy of perioperative pain management

The present results provide important information about one property of a PROM required to be assessed for selecting the most appropriate PROMs for a domain in a COS for assessing effectiveness and efficacy of perioperative pain management.[11] These results will be presented to the next IMI-PainCare PROMPT consensus panel as indices for the quality of different PROMs and used (together with other measurement qualities assessed) to decide which PROM is best suited for a COS of PROMs related to perioperative pain.

Table 1 Correlations between reported change and calculated change including 95% confidence intervals. Calculations are between change items (either global impression of change or specific impression of change in the item noted, on a 7-item scale ranging from 'very much worse' to 'very much improved') and calculated changes between day 1 and day 3 in 11-item numeric rating scales (with anchors 'no pain' and 'worst pain imaginable' for pain intensities, 'does not interfere' and 'completely interferes' for interference, and 'very uncertain' and 'very certain' for self-efficacy items). Adverse events are changes in several reported out of 11 multiple-choice items. *Procedure-specific activity items were used.

Change item	Calculated change in	All	Breast surgery	Endometriosis surgery	Total knee arthroplasty	Sternotomy
Global change item (PGIC)	Pain at rest	0.19 (0.17–0.22)	0.23 (0.17–0.29)	0.27 (0.22–0.32)	0.22 (0.16–0.28)	0.13 (0.08–0.18)
	Pain average	0.23 (0.20–0.26)	0.25 (0.19–0.31)	0.30 (0.25–0.35)	0.27 (0.21–0.33)	0.16 (0.11–0.21)
	Pain worst	0.25 (0.22–0.28)	0.26 (0.20–0.32)	0.35 (0.30–0.40)	0.30 (0.24–0.36)	0.17 (0.12–0.22)
	Pain during activity*	0.24 (0.21–0.27)	0.22 (0.16–0.28)	0.30 (0.25–0.35)	0.27 (0.21–0.33)	0.12 (0.07–0.17)
	Pain during physiotherapy	0.04 (−0.01–0.08)	0.03 (−0.23–0.29)	0.02 (−0.47–0.49)	0.02 (−0.06–0.09)	0.07 (0.01–0.13)
	Interference in bed	0.20 (0.17–0.22)	0.24 (0.18–0.30)	0.33 (0.28–0.37)	0.21 (0.15–0.27)	0.07 (0.02–0.12)
	Interference during activity*	0.22 (0.19–0.24)	0.23 (0.17–0.29)	0.31 (0.26–0.36)	0.21 (0.15–0.28)	0.10 (0.05–0.15)
	Interference during physiotherapy	0.10 (0.06–0.13)	0.01 (−0.11–0.12)	0.06 (−0.02–0.14)	0.13 (0.06–0.20)	0.08 (0.02–0.13)
	Self-efficacy pain decrease	0.12 (0.09–0.15)	0.08 (0.01–0.15)	0.15 (0.10–0.21)	0.10 (0.04–0.17)	0.16 (0.10–0.21)
	Self-efficacy sleep	0.14 (0.11–0.16)	0.12 (0.05–0.19)	0.20 (0.14–0.25)	0.15 (0.09–0.21)	0.10 (0.04–0.16)
	Self-efficacy things to do	0.14 (0.11–0.17)	0.08 (0.02–0.15)	0.19 (0.13–0.24)	0.12 (0.05–0.18)	0.13 (0.07–0.19)
	Self-efficacy active	0.13 (0.10–0.16)	0.07 (0.00–0.13)	0.13 (0.07–0.18)	0.16 (0.10–0.22)	0.17 (0.11–0.23)
	Self-efficacy fatigue	0.15 (0.12–0.18)	0.06 (−0.01–0.13)	0.18 (0.12–0.23)	0.17 (0.10–0.23)	0.16 (0.10–0.22)
	Self-efficacy feeling blue	0.12 (0.09–0.15)	0.00 (−0.06–0.07)	0.09 (0.04–0.15)	0.15 (0.09–0.21)	0.21 (0.16–0.27)
	Self-efficacy daily activities	0.14 (0.11–0.17)	0.04 (−0.03–0.10)	0.17 (0.11–0.22)	0.18 (0.12–0.24)	0.21 (0.15–0.27)
	Self-efficacy frustration	0.14 (0.11–0.17)	0.01 (−0.06–0.07)	0.16 (0.11–0.22)	0.19 (0.13–0.25)	0.19 (0.13–0.25)
	Adverse events	0.17 (0.14–0.20)	0.12 (0.06–0.18)	0.17 (0.12–0.22)	0.24 (0.18–0.30)	0.19 (0.15–0.23)
Pain at rest (PSIC)	Pain at rest	0.27 (0.25–0.30)	0.30 (0.24–0.36)	0.41 (0.36–0.45)	0.29 (0.23–0.35)	0.15 (0.10–0.20)
Pain average PSIC	Pain average	0.29 (0.26–0.32)	0.29 (0.22–0.35)	0.41 (0.36–0.46)	0.34 (0.28–0.39)	0.16 (0.11–0.21)
Pain during activity* PSIC	Pain during activity*	0.31 (0.28–0.33)	0.27 (0.21–0.33)	0.42 (0.37–0.46)	0.28 (0.22–0.34)	0.17 (0.12–0.21)
Pain during physiotherapy PSIC	Pain during physiotherapy	0.07 (0.03–0.12)	0.05 (−0.27–0.36)	0.14 (−0.42–0.62)	0.04 (−0.03–0.12)	0.09 (0.04–0.15)
Pain worst PSIC	Pain worst	0.29 (0.26–0.32)	0.29 (0.23–0.35)	0.48 (0.43–0.52)	0.35 (0.29–0.40)	0.14 (0.09–0.19)
Interference in bed PSIC	Interference in bed	0.26 (0.23–0.28)	0.29 (0.23–0.35)	0.39 (0.34–0.44)	0.28 (0.22–0.34)	0.11 (0.06–0.16)
Interference during activity* PSIC	Interference during activity*	0.29 (0.27–0.32)	0.31 (0.24–0.37)	0.43 (0.38–0.47)	0.24 (0.18–0.30)	0.15 (0.10–0.20)
Interference during physiotherapy PSIC	Interference during physiotherapy	0.15 (0.11–0.18)	0.03 (−0.11–0.17)	0.07 (−0.02–0.16)	0.20 (0.13–0.27)	0.14 (0.08–0.19)
Self-efficacy PSIC	Self-efficacy pain decrease	0.20 (0.17–0.23)	0.13 (0.06–0.19)	0.26 (0.20–0.31)	0.26 (0.20–0.32)	0.19 (0.13–0.25)
	Self-efficacy sleep	0.17 (0.14–0.20)	0.16 (0.10–0.23)	0.25 (0.20–0.31)	0.20 (0.14–0.26)	0.14 (0.08–0.20)
	Self-efficacy things to do	0.20 (0.17–0.22)	0.12 (0.05–0.19)	0.27 (0.21–0.32)	0.22 (0.16–0.28)	0.16 (0.10–0.22)
	Self-efficacy active	0.19 (0.16–0.22)	0.11 (0.04–0.18)	0.18 (0.13–0.24)	0.29 (0.23–0.35)	0.19 (0.13–0.25)
	Self-efficacy fatigue	0.20 (0.17–0.23)	0.11 (0.04–0.18)	0.24 (0.19–0.29)	0.29 (0.23–0.35)	0.17 (0.11–0.23)
	Self-efficacy feeling blue	0.21 (0.18–0.24)	0.09 (0.03–0.16)	0.21 (0.15–0.26)	0.27 (0.21–0.33)	0.22 (0.16–0.28)
	Self-efficacy daily activities	0.21 (0.18–0.24)	0.04 (−0.03–0.10)	0.27 (0.22–0.32)	0.29 (0.23–0.35)	0.26 (0.20–0.31)
	Self-efficacy frustration	0.20 (0.17–0.23)	0.06 (−0.01–0.13)	0.24 (0.19–0.29)	0.27 (0.21–0.33)	0.21 (0.16–0.27)
Adverse events PSIC	Adverse events	0.28 (0.26–0.31)	0.25 (0.19–0.31)	0.30 (0.25–0.34)	0.33 (0.27–0.38)	0.30 (0.26–0.34)

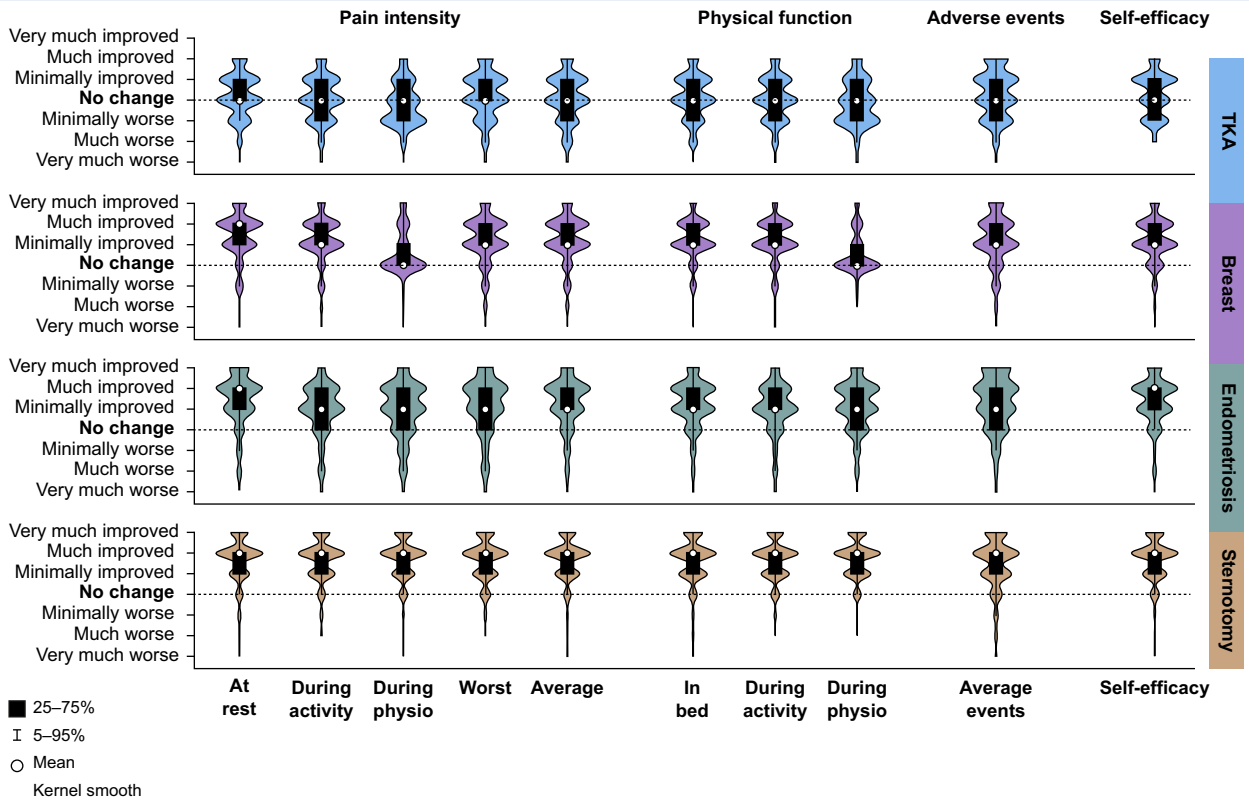

Fig 5. Specific impression of change (PSIC) items related to all domains assessed on day 3 (d3). PSIC items for pain intensity, impairment of physical function, adverse events, and self-efficacy were assessed on d3 with the five pain-related domains assessed on the 7-item scale ranging from 'very much improved' to 'very much worsened' and displayed as violin plots (Kernel smoothing, white dots represent mean, black boxes represent 25–75% quartile, and whiskers represent 5–95% percentile, n=510 total knee replacement [TKA], n=484 breast surgery, n=695 surgery related to endometriosis, n=972 sternotomy).

Similar to the Standardised Endpoints in Perioperative Medicine (StEP) initiative which developed the Core Outcome Measures for Perioperative and Anaesthetic Care (COMPAC) COS guiding stakeholders for selecting outcomes for future perioperative clinical trials,[31] the IMI-PainCare PROMPT initiative will provide additional and specific PROMs for being used in clinical trials and follow-up of treatment success in clinical practice related to perioperative pain management. Future analysis will follow to assess, for example, the minimal clinically important differences of these chosen PROMs, which is the number of changes required to inform healthcare professionals about the smallest change in score that still constitutes a meaningful change.[32]

Limitations of the study

The PROMs investigated here were identified by literature searches within IMI-PainCare and selected based on their available psychometric properties as recommended by COSMIN. However, choosing appropriate PROMs for perioperative pain management was considerably hampered because of different reasons (unavailability of PROMs for domain or target population, missing evidence for content validity as one prerequisite[11] of sound PROMs, lack of transculturally adapted PROMs for different languages). The use of PROMs reflects the general situation of PROM development and use. The currently limited assessment of psychometric properties of all available PROMs implies that for PROMs which shall be included in a COS, reassessment of psychometric properties, transculturally translation and adaptation, or even redevelopment of validated scales is urgently required. Data were not equally distributed between sites and countries, and we performed monitoring but could not visit or audit study sites. Another limitation refers to the sensitivity-to-change approach. As advised by COSMIN,[33] there are two different ways of detecting sensitivity-to-change: the anchor-based and the gold standard approach. For the latter, existing gold standard instruments would be applied; unfortunately, these were missing for our setting. Therefore, we chose the anchor approach, using the validated global impression of change as anchor. Finally, we only assessed four surgical procedures within IMI-PainCare PROMPT NIT-1; however, within the IMI-PainCare PROMPT Delphi panel consensus process on domains for perioperative pain management,[10] we accentuated the question of developing either separate COS for perioperative pain management after each surgical procedure or an overarching COS comprising all. By using a wide spectrum of surgeries with different requirements of perioperative pain management relevant for these procedures, the panel did see a

great overlap and considered the core outcome domains not as procedure-specific. In addition, they did not see the need for procedure-specific PROMs for the domains *pain intensity*, *adverse events*, and *self-efficacy*. However, they emphasised the need to assess *physical function* specific to the type of surgery (procedure-/condition-specific).[10]

Conclusions

Our results can inform a COS consensus panel about patient-reported outcome measures recommended for outcome assessment in clinical trials and for clinical practice related to perioperative pain management. A range (COS) of domains already consented on by a multidisciplinary panel covering different aspects of patients' individual experience with postoperative pain[10] seems to be necessary for reliably estimating efficacy or effectiveness of perioperative pain management.

Authors' contributions

Study design: EMPZ, WM, HL
Experiments: EMPZ, WM, CW, MK, DCR, FF, TM, KS, LS, DF, PLH, EK
Data analysis plan/data analysis: JV, EMPZ, UK, DS
Writing first manuscript draft: EMPZ, JV, DS
Figures: JV, DS, EMPZ
Manuscript revision: WM, CW, MK, DCR, FF, TM, KS, LS, DF, HL, PLH, EK, UK

Acknowledgements

We would like to thank all the clinical investigators in the additional centres who participated in PROMPT NIT-1: Benno Rehberg (Division of Anaesthesiology, Department of Anaesthesiology, Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland); Viktoria Frebault, Charles Fiquet and Philippe Neyret (Orthopedy Unit, Infirmerie Protestante de Lyon, France); Caterina Aurilio, Maria Caterina Pace, and Pasquale Sansone (Department of Women, Child and General and Specialized Surgery, University of Campania Luigi Vanvitelli, Naples, Italy); Dragana Unic-Stojanovic (Institute for Cardiovascular Diseases Dedinje, Medical faculty University of Belgrade, Belgrade, Serbia); Germano Cardoso, Patricia Redondo, Sofia Dias, Sofia Travisco, Maria Amelia Gomes, Rita Canotilho, Cátia Ribeiro, and Ana Margarida Correia (Surgery, Instituto Português de Oncologia Porto, Portugal); Julien Cabaton (Anesthésie – Réanimation en Orthopédie, Centre Paul Santy – Hôpital Privé Jean Mermoz, Lyon, France); Maria A. Perez-Herrero (FEA Anestesiologia y Reanimación, Hospital Clínico Universitario de Valladolid, Saint, Spain), Rainer Sabatowski (Comprehensive Pain Center, Medical Faculty and University Hospital Carl Gustav Carus, TU Dresden, Dresden, Germany); Svetlana Dinić, Mina Đaković, Milena Jović, Nemanja Dimić, and Lazar Bralušić (Department of Resuscitation and Intensive Care at the Institute for Orthopedic Diseases 'Banjica', Belgrade, Serbia); Theresa Wodehouse and Kristin Ullrich (Barts Health NHS Trust, St. Bartholomew's Hospital, London, UK); Thomas Volk (Department of Anaesthesiology, Intensive Care and Pain Therapy, Saarland University Medical Centre, Homburg, Germany); and Ulrike Stamer (Department of Anaesthesiology and Pain Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland). We also thank Jena University Hospital's Centre for Clinical Studies.

Declarations of interest

JV received payments for advisory boards and talks outside of the submitted work from Embody Orthopaedics, UK, and Casquar, Germany. EK has received fees for advisory board or lecture activities from Oriona Pharma, Finland, Pfizer, Sweden, and GSK, Finland. WM received payments for advisory boards and talks outside of the submitted work from Kyowa Japan, Mundipharma Singapore, Grünenthal, Germany Ethypharm, France, and Spectrum Therapeutics, Germany. His institution received research support from Pfizer, USA, Mundipharma, Germany, and Grünenthal, Germany. EMPZ received financial support from Grünenthal, Germany, for research activities and advisory and lecture fees from Grünenthal, Germany, Novartis, Germany, and Medtronic, Switzerland. In addition, she receives scientific support from the German Research Foundation (DFG), the Federal Ministry of Education and Research (BMBF), the Federal Joint Committee (G-BA), and the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777500. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. All money went to the institutions (WWU/UKM) EMPZ is working for. All the other authors have no conflicts of interest to declare.

Funding

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777500. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA Companies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bja.2023.10.020.

References

1. Fletcher D, Stamer UM, Pogatzki-Zahn E, et al. Chronic postsurgical pain in Europe: an observational study. *Eur J Anaesthesiol* 2015; **32**: 725–34
2. van Boekel RLM, Steegers MAH, Verbeek-van Noord I, van der Sande R, Vissers KC. Acute pain services and postsurgical pain management in The Netherlands: a survey. *Pain Pract* 2015; **15**: 447–54
3. Rosenberger DC, Pogatzki-Zahn EM. Chronic post-surgical pain – update on incidence, risk factors and preventive treatment options. *BJA Educ* 2022; **22**: 190–6
4. Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing numerical rating scales, Verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 2011; **41**: 1073–93
5. Komann M, Baumbach P, Stamer UM, et al. Desire to receive more pain treatment - a relevant patient-reported outcome measure to assess quality of post-operative pain

management? Results from 79,996 patients enrolled in the pain registry QUIPS from 2016 to 2019. *J Pain* 2021; **22**: 730–8
6. Schwenkglenks M, Gerbershagen HJ, Taylor RS, et al. Correlates of satisfaction with pain treatment in the acute postoperative period: results from the international PAIN OUT registry. *Pain* 2014; **155**: 1401–11
7. van Dijk JFM, Kappen TH, Schuurmans MJ, van Wijck AJ. The relation between patients' NRS pain scores and their desire for additional opioids after surgery. *Pain Pract* 2015; **15**: 604–9
8. Bigalke S, Maeßen TV, Schnabel K, et al. Assessing outcome in postoperative pain trials: are we missing the point? A systematic review of pain-related outcome domains reported in studies early after total knee arthroplasty. *Pain* 2021; **162**: 1914–34
9. Maeßen TV, Austenfeld E, Kaiser U, et al. Systematic review on pain-related outcome domains after sternotomy: a first step toward the development of a core outcome set. *Value Health* 2021; **24**: 1203–12
10. Pogatzki-Zahn EM, Liedgens H, Hummelshoj L, et al. Developing consensus on core outcome domains for assessing effectiveness in perioperative pain management: results of the PROMPT/IMI-PainCare Delphi Meeting. *Pain* 2021; **162**: 2717–36
11. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; **19**: 539–49
12. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; **27**: 1159–70
13. Butcher NJ, Monsour A, Mew EJ, et al. Guidelines for reporting outcomes in trial reports: the CONSORT-outcomes 2022 extension. *JAMA* 2022; **328**: 2252–64
14. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; **20**: 1727–36
15. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**: 361–70
16. Ruscheweyh R, Marziniak M, Stumpenhorst F, Reinholz J, Knecht S. Pain sensitivity can be assessed by self-rating: development and validation of the Pain Sensitivity Questionnaire. *Pain* 2009; **146**: 65–74
17. Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess* 1995; **7**: 524–32
18. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief pain inventory. *Ann Acad Med Singap* 1994; **23**: 129–38
19. Rothaug J, Zaslansky R, Schwenkglenks M, et al. Patients' perception of postoperative pain management: validation of the International Pain Outcomes (IPO) questionnaire. *J Pain* 2013; **14**: 1361–70
20. Bouhassira D, Attal N, Fermanian J, et al. Development and validation of the neuropathic pain symptom inventory. *Pain* 2004; **108**: 248–57
21. Bouhassira D, Attal N, Alchaar H, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005; **114**: 29–36
22. Chiarotto A, Deyo RA, Terwee CB, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J* 2015; **24**: 1127–42
23. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005; **113**: 9–19
24. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum* 1989; **32**: 37–44
25. Chiou CP. Development and psychometric assessment of the physical symptom distress scale. *J Pain Symptom Manage* 1998; **16**: 87–95
26. Gerbershagen HJ, Rothaug J, Kalkman CJ, Meissner W. Determination of moderate-to-severe postoperative pain on the numeric rating scale: a cut-off point analysis applying four different methods. *Br J Anaesth* 2011; **107**: 619–26
27. Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. *Trials* 2017; **18**: 280
28. Gerbershagen HJ, Aduckathil S, van Wijck, Peelen LM, Kalkman CJ, Meissner W. Pain intensity on the first day after surgery: a prospective cohort study comparing 179 surgical procedures. *Anesthesiology* 2013; **118**: 934–44
29. Gerbershagen HJ, Pogatzki-Zahn E, Aduckathil S, et al. Procedure-specific risk factor analysis for the development of severe postoperative pain. *Anesthesiology* 2014; **120**: 1237–45
30. Hewlett S, Cockshott Z, Kirwan J, Barrett J, Stamp J, Haslock I. Development and validation of a self-efficacy scale for use in British patients with rheumatoid arthritis (RASE). *Rheumatology (Oxford)* 2001; **40**: 1221–30
31. Myles PS, Boney O, Botti M, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. *Br J Anaesth* 2018; **120**: 705–11
32. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. *J Rheumatol* 2001; **28**: 406–12
33. de Vet HCW, Terwee CB, Mokkink L, Knol DL. *Measurement in medicine: a practical guide*. Cambridge, New York, Port Melbourne, New Delhi: Cambridge University Press; 2011

Handling Editor: Paul Myles