Received: 21 July 2022 | Revised: 23 August 2022 | Accepted: 4 September 2022

DOI: 10.1002/ejp.2033

ORIGINAL ARTICLE

A novel methodology to integrate outcomes regarding perioperative pain experience into a composite score: Prediction model development and validation

Bailin Jiang[1] | **Yanhong Liu**[2] | **Yaqing Wu**[1] | **Weidong Mi**[2] | **Yi Feng**[1]

[1]Department of Anesthesiology, Peking University People's Hospital, Beijing, China

[2]Anesthesia and Operation Center, the First Medical Center, Chinese PLA General Hospital, Beijing, China

Correspondence
Weidong Mi, Anesthesia and Operation Center, the First Medical Center, Chinese PLA General Hospital, No. 28 Fuxing Road, Haidian District, Beijing 100853, China.
Email: wwdd1962@163.com

Yi Feng, Department of Anesthesiology, Peking University People's Hospital, No.11 Xizhimen South Street, Xicheng District, Beijing 100044, China.
Email: doctor_yifeng@sina.com

Funding information
Seventh Framework Programme, Grant/Award Number: 223590; Research and Development, Grant/Award Number: 2018YFC2001905

Abstract

Background: An integrated score that globally assesses perioperative pain experience and rationally weights each component has not yet been developed.

Methods: A development dataset specific to adult Chinese patients undergoing orthopaedic surgery was obtained from PAIN OUT (1985 qualified patients of 2244). A more recent validation dataset obeying the same conditions was obtained from the Chinese Anaesthesia Shared-database Platform (1004 qualified patients of 1032). Outcomes were assessed using the International Pain Outcomes Questionnaire (IPO-Q), which comprises key patient-level outcomes of perioperative pain management, including pain experience and perceptions of care. Using principal component analysis and regression models, a composite score (CS) was inferred to integrate pain experience. The discrimination of the CS for dissatisfaction and desire for more pain treatment was compared with that of the worst pain score.

Results: A CS was developed from the 12 items of the IPO-Q regarding pain experience. The weight for calculating the CS was worst pain 11, least pain 17, time spent in severe pain 11, interference with activity in bed 9, interference with breathing deeply or coughing 10, interference with sleep 9, anxiety 12, helplessness 12, nausea 0, drowsiness 2, itch 5 and dizziness 2. In external validation, the CS indicated superior discrimination to the worst pain in predicting dissatisfaction ($p < 0.001$) and desire for more pain treatment ($p < 0.001$).

Conclusions: This study introduced a methodology to integrate outcomes regarding perioperative pain experience into a CS, which was based on the weight of each item.

Significance: This novel methodology sheds additional light on the riveting issue of carefully integrating several measures into a composite endpoint, which may be useful for quality improvement purposes when addressing the impact of a change in clinical practice.

1 | INTRODUCTION

A precise assessment of perioperative pain forms the basis for appropriate pain management (ASATFAPM., 2012). Although several outcome domains have been highlighted to be assessed practically, an 11-point numerical rating scale (NRS, 0 = no pain, 10 = worst pain possible) of average or worst pain was still the most common approach for assessing perioperative pain experience (Gerbershagen et al., 2011). However, pain is multidimensional (Zaslansky et al., 2018). Overtreatment or insufficient treatment ensues from an inappropriate assessment (Gerbershagen et al., 2011). Integrating several outcomes into a composite score (CS) is attractive and promising for both clinicians and investigators (Benzon et al., 2017). Compared with separate comparisons of each outcome, a CS generally has higher statistical power and correctly controls type I error rates (Dai et al., 2013; Silverman et al., 1993). Anesthesiologists are endeavouring to develop an integrated score for the assessment of pain experience (Benzon et al., 2017; Dai et al., 2013; Myles et al., 2016; Silverman et al., 1993). Unlike the assessment of an individual outcome that guides the specific treatment, an integrated score allows a clear conclusion regarding the total benefit provided by a regimen or policy. Nevertheless, to the best of our knowledge, no method that weighs each component deliberately has been developed. As a common methodology for assessing perioperative pain experience has never been reached (Benzon et al., 2017; Silverman et al., 1993), it is a challenge to find a standard to develop and validate this new score.

This study aimed to introduce a method for developing a weighted integrated score that can assess pain experience. We hypothesized that if pain experience could be globally tested using an integrated score, this latent score would be more associated with the patient perception of care compared with an individual outcome that embraces only partial information. Unlike chronic pain, acute pain is less affected by mental factors but by the pain experience itself (Gerbershagen et al., 2011). Patient perception might address subjective pain experience and was considered a global judgement of pain management in some studies (Komann et al., 2021; Stamer et al., 2021). Therefore, we divided patient-reported outcomes (PROs) into pain experience and patient perception. A CS was developed from the outcomes of the pain experience and tested based on the outcomes of the patient perception. Two crucial outcomes of patient perception, satisfaction and desire for more pain treatment were taken together as the makeshift pain experience to test this new score. Worst pain, an extensively used assessment (Greimel et al., 2018; Schnabel et al., 2020), was used as the individual outcome to verify our hypothesis. The discrimination of the CS for predicting these two outcomes of patient perception was compared with that of the worst pain.

2 | METHODS

2.1 | Study design and data source

This is a prognostic study for the development and validation of prediction models. Data were collected before the development of the CSs. The methodology used for data collection was provided by PAIN OUT (www. Pain-out. EU), an international quality improvement and perioperative pain registry project (Zaslansky et al., 2015). The methodology was described before (Zaslansky et al., 2015) and was registered at ClinicalTrials. gov (NCT02083835). The PAIN OUT provided a standardized methodology to assess multidimensional pain-related PROs on postoperative day 1 (POD1) (Baca et al., 2021; Zaslansky et al., 2018) and trained surveyors from each center. All surveyors had to pass quizzes to achieve the standards of approaching patients, collecting data and entering them into a web-based password-secure portal. The International Pain Outcomes Questionnaire (IPOQ), a validated questionnaire, was used to assess the PROs (Rothaug et al., 2013).

Anonymized data for inferring the CS were obtained from the PAIN OUT as the development set. A more recent dataset was provided by the Chinese Anaesthesia Shared-database Platform (CASP) as the validation set. The same methodology and surveyors were employed in this project to ensure that the data were consistent with PAIN OUT data. All participating centers in this study obtained ethical approval from local institutional review boards (IRBs). Ethical approval for this study (approval number: 2018PHB050-01) was provided by the IRB of Peking University People's Hospital, Beijing, China (the principal investigator organization, Chairperson Prof Kaiyan Liu) on 11 May 2018. As per the requirements of the principal IRB, written informed consent was waived, but oral consent was obtained from all patients. This manuscript adheres to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statements.

2.2 | Patients

A dataset specific to adult Chinese patients undergoing orthopaedic surgery was selected from the PAIN OUT as the development set. The same conditions were used to obtain a population comparable with that of the PAIN OUT from the CASP as the validation set. Patients were eligible if they fulfilled the following inclusion

criteria: (a) underwent any kind of inpatient orthopaedic surgery; (b) were 18 years or older; (c) were on POD 1 and returned to the ward from the postanesthesia care unit for at least 6 h and (d) consented to the survey. Patients whose data about 'worst pain', 'satisfaction' or 'desire for more treatment' were missing were excluded from this study.

2.3 | Measured parameters

PROs were assessed using the IPO-Q (Rothaug et al., 2013), which evaluates the five outcome domains. Four of these were pain experiences, and the fifth was the patient perception of perioperative pain management (Zaslansky et al., 2018).

Pain experiences included (a) intensity of pain since surgery (worst, least pain and time spent in severe pain), (b) interference of pain with function (activities in and out of bed, sleeping and breathing deeply, or coughing), (c) emotional impairment due to pain (anxiety and helplessness) and (d) adverse effects (nausea, drowsiness, itching and dizziness).

Patient perception comprises (a) pain relief from treatment, (b) satisfaction with pain treatment, (c) desire for more pain treatment, (d) receipt of information about pain treatment options and (e) participation in decisions about pain treatment.

Most items used an 11-point numerical rating scale (NRS). Two consisting of 'desire for more treatment' and 'receipt of information' were assessed using a dichotomous yes/no scale. Two items addressing 'time spent in severe pain' and 'pain relief from treatment' were recorded using a percentage scale (0 to 100%), which was standardized using an 11-point NRS to facilitate analysis in this study. A dichotomous approach was used to divide 'satisfaction' ratings into two subsets to construct the regression model. According to an ABC analysis (Ultsch & Lötsch, 2015) of the baseline data in 10 countries from the PAIN OUT registry (the report has not yet been published) and a previous study (Vasta et al., 2022), the definition of dissatisfaction was determined before analysis. Patients with an NRS of satisfaction of 7 or less were defined as dissatisfied, and those with an NRS of more than 7 were defined as satisfied. Considering that some patients may not have been out of bed until the survey, pain interference with activities outside the bed was not analysed in this study.

2.4 | Data processing

2.4.1 | Missing data

The multiple imputation (MI) technique (Austin et al., 2021) was used to handle random missing data in the development set, and six MI datasets were established for the following analysis. For the validation set, all patients with missing values were excluded from the analysis to ensure concrete validation.

2.5 | Sample size calculation

The sample size was estimated according to satisfaction with the perioperative pain treatment. According to the analysis of the development set, moderate or severe pain (NRS > 3) in worst pain exhibited a sensitivity of 83.1% and specificity of 52.9% in discriminating whether patients reported dissatisfaction (NRS ≤ 7). We hypothesized that the discrimination of the CS would be superior to that of the worst pain NRS score. With a two-sided α level of 5%, a permissible error of 0.08, and an attrition rate of 10%, 94 patients who reported dissatisfaction and 142 patients who reported satisfaction were required to verify the discrimination of the CS.

2.6 | Composite score derivation

Principal components with a cumulative contribution rate > 70% were extracted from all the pain experience items as candidate predictors, and multiple linear regression was used to fit a 'satisfaction' model (Table S1). The partial regression coefficients of the principal components entering the model and the score coefficients of each perioperative pain item in the corresponding principal components were used to calculate the weight of each item, and a comprehensive linear score was obtained. In brief, we multiplied the partial regression coefficient of the significant principal component (*RC*) by the score coefficient (*SC*) of a certain item in the same principal component and then divided the product by the standard deviation (*SD*) of this item to standardize it. Next, the results of this item from each significant principal component were summarized to obtain the weight (W) for this item.

$$W = \sum \frac{RC_i \times SC_i}{SD_i}$$

Finally, we standardized the weight of each item to contribute a total of 100 points.

$$SW = \frac{W_i}{\sum W_i} \times 100$$

where *SW* is the standardized weight of each item.

The *CS* was the sum of each item's score, which was calculated by multiplying the item's standardized value (*SV*) by its weight divided by 100. The SV was calculated

by transforming the actual value into a 0–10 scale (NRS, 0 = null, 10 = most).

$$CS = \frac{\sum SV_i \times SW_i}{100}$$

The CS was a 0 to 10 scale (0 = least pain experience, 10 = worst) with two decimal places.

2.7 | Validation

The area under the receiver operating characteristic curve (AUC) for predicting both dissatisfaction and desire for more pain treatment using the CS was calculated in the development set for internal validation. Furthermore, the Delong test was used to compare the discrimination between the CS and worst pain.

A temporal validation strategy was used for the external validation. In the validation set, a more recent dataset, the CS was validated using the AUC to predict dissatisfaction and desire for more pain treatment. After adjusting for the influence of other variables of patient perception, including participation in treatment decisions, pain relief and received treatment information, the modified AUC was calculated in the validation set. The Delong test was conducted to compare the CS with the worst pain in the validation set. The worst pain NRSs and CSs were submitted as two anonymized scores to another independent statistician for comparison.

2.8 | Statistical analysis

Continuous variables are expressed as the mean ± SD, and categorical data are presented as numbers (proportions). Given the relatively large amount of data, we used effect size to describe the difference between the two sets. Cohen's *d* was used for continuous variables and is presented as a *d* value with 95% confidence intervals (95% CI). The *d* value equaling or exceeding ±0.2, 0.5 or 0.8 suggests a small, medium or large meaningful difference, respectively (Zaslansky et al., 2018). For categorical data, the odds ratios were converted to the same metric, Cohen's *d*, to facilitate the statement of clinical significance (Zaslansky et al., 2018). The correlation matrix method was used to analyse the principal components. A stepwise method was used to fit the multiple-linear model for predicting satisfaction with the aforementioned principal components. Logistic regression for predicting dissatisfaction or desire for more pain treatment was used to adjust the AUC of the CSs for the influence of other variables of patient perception. The Delong test was used to compare the AUC.

Two-sided *p* values <0.05 were considered statistically significant. Statistical analyses were performed using MedCalc (version 20.0; MedCalc Software Ltd.) and IBM SPSS Statistics (version 25.0; IBM Corp.).

3 | RESULTS

3.1 | Patient characteristics

From February 2014 to November 2020, a total of 2244 patients from 20 centers were approached, of whom 1985 patients qualified for the analysis as the development set; from May 2020 to September 2021, 1032 patients from 19 centers were approached, of whom 1004 patients qualified for the analysis as the validation set (Figure S1). In the development set, there were 5.6% missing data for participation in decisions, 3.0% for pain relief from treatment, 2.1% for receipt of information, 2.1% for helplessness, 1.6% for anxiety, and less than 1% for other variables of PROs. There were no missing data in the validation set, which was refined for precise validation.

The patients' characteristics are presented in Table 1.

3.2 | Patient-reported outcomes

Comparable pain intensities were reported in the development and validation sets. There were small-to-medium differences between the two sets in the interference of pain with function, emotional impairment and adverse effects. Large differences existed regarding parts of the patient perception of perioperative pain management, including pain relief from treatment, receipt of information and participation in decisions. Satisfaction was lower in the development dataset. The desire for more pain treatment in both datasets was similar. More patients in the development set were considered to be dissatisfied. The details are presented in Table 2.

3.3 | Composite score

Five principal components with a cumulative explained variance of 71.5% were extracted, of which the first two were included in the regression model (Table S1). The weight for the pain experience was worst pain 11, least pain 17, time spent in severe pain 11, interference with activity in bed 9, interference with breathing deeply or coughing 10, interference with sleep 9, anxiety 12, helplessness 12, nausea 0, drowsiness 2, itching 5 and dizziness 2 (Table 3). Table S2 provides an example regarding how to calculate the individual CS using these weights.

TABLE 1 Baseline characteristics of patients

Characteristic	Development set (n = 1985)	Validation set (n = 1004)	d value (95% CI)	Effect size
Age (year)	51.73 (16.39)	56.67 (16.37)	−0.30 (−0.38, −0.22)	Small
BMI (kg·m^{-2})	24.43 (3.76)	25.39 (3.63)	−0.26 (−0.33, −0.18)	Small
Male [n (%)]	1000 (50.4)	510 (50.8)	−0.01 (−0.09, 0.07)	Negligible
Type of surgery [n (%)]				
Joint replacement	594 (29.9)	335 (33.4)	−0.09 (−0.18, 0.00)	Negligible
Fracture fixation	548 (27.6)	141 (14.0)	0.47 (0.36, 0.58)	Small
Spine surgery	324 (16.3)	368 (36.7)	−0.60 (−0.70, −0.50)	Medium
Reconstruction	315 (15.9)	81 (8.1)	0.42 (0.28, 0.56)	Small
Others	204 (10.3)	79 (7.9)	0.16 (0.01, 0.31)	Negligible

Note: Continuous variables are expressed as mean ± SD; categorical data are presented as number (proportion).

TABLE 2 Patient-reported outcomes regarding pain experience and patient's perception in this study

Item	Development set (n = 1985)	Validation set (n = 1004)	d value (95% CI)	Effect size
Pain experience				
Composite score (NRS)	1.90 (1.44)	1.47 (1.07)	0.32 (0.25, 0.40)	Small
Pain intensity (NRS)				
Worst pain	4.32 (2.45)	4.14 (2.24)	0.08 (0.00, 0.15)	Negligible
Least pain	1.26 (1.47)	1.04 (1.33)	0.16 (0.08, 0.23)	Negligible
Time spent in severe pain[a]	2.04 (2.34)	2.16 (2.17)	−0.05 (−0.13, 0.03)	Negligible
Interference with function (NRS)				
Activities in bed	3.31 (2.75)	2.62 (2.35)	0.26 (0.19, 0.34)	Small
Breathing deeply or coughing	0.88 (1.68)	0.77 (1.58)	0.07 (−0.01, 0.14)	Negligible
Sleep	2.15 (2.85)	1.33 (2.34)	0.30 (0.23, 0.38)	Small
Emotional impairment (NRS)				
Anxiety	1.66 (2.37)	0.68 (1.44)	0.47 (0.39, 0.55)	Small
Helplessness	1.16 (2.12)	0.49 (1.29)	0.35 (0.28, 0.43)	Small
Adverse effects (NRS)				
Nausea	1.12 (2.04)	0.68 (1.56)	0.23 (0.16, 0.31)	Small
Drowsiness	1.59 (2.22)	0.45 (1.21)	0.59 (0.51, 0.66)	Medium
Itch	0.34 (1.07)	0.14 (0.47)	0.22 (0.15, 0.30)	Small
Dizziness	0.97 (1.82)	0.48 (1.11)	0.30 (0.23, 0.38)	Small
Patient's perception				
Dissatisfied with treatment [n (%)][b]	557 (28.1)	126 (12.5)	0.55 (0.44, 0.67)	Medium
Desiring more treatment [n (%)]	564 (28.4)	258 (25.7)	0.08 (−0.02, 0.17)	Negligible
Participation in treatment decisions (NRS)	5.26 (3.94)	8.46 (2.05)	−0.94 (−1.02, −0.86)	Large
Pain relief (NRS)[a]	3.21 (2.59)	7.88 (2.05)	−1.93 (−2.02, −1.84)	Large
Satisfaction (NRS)	8.22 (2.02)	8.82 (1.47)	−0.33 (−0.40, −0.25)	Small
Received treatment information [n (%)]	1064 (53.6)	906 (90.2)	−1.15 (−1.27, −1.02)	Large

Note: Continuous variables are expressed as mean ± SD; categorical data are presented as number (proportion).

Abbreviation: NRS, numeric rating scale.

[a]Standardized as a 0–10 scale.

[b]Defined as an NRS of satisfaction ≤7.

The CS in the development set was higher than that in the validation set (1.90 (1.44) vs. 1.47 (1.07) NRS, small effect size) (Table 2).

TABLE 3 The weight for the composite score

Item (NRS)	Weight
Pain intensity	
Worst pain	11
Least pain	17
Time spent in severe pain[a]	11
Interference with function	
Activities in bed	9
Breathing deeply or coughing	10
Sleep	9
Emotional impairment	
Anxiety	12
Helplessness	12
Adverse effects	
Nausea	0
Drowsiness	2
Itch	5
Dizziness	2
Total	100

Abbreviation: NRS, numeric rating scale.

[a]Standardized as a 0–10 scale.

3.4 | Validation

In the internal validation, the AUC for predicting dissatisfaction and desire for more pain treatment by the CS was 0.79 (95% CI: 0.78, 0.81) and 0.75 (95% CI: 0.73, 0.77), respectively. After adjusting for other items of patient perception, the adjusted AUC for predicting dissatisfaction and desire for more treatment by the CS was 0.84 (95% CI: 0.82, 0.85) and 0.79 (95% CI: 0.77, 0.81), respectively. The CS indicated superior discrimination to worst pain in predicting dissatisfaction (AUC of worst pain = 0.75 [0.73,0.77], $p < 0.001$) and desire for more pain treatment (AUC of worst pain = 0.74 [0.72,0.76], $p = 0.001$). See Figure S2.

In external validation, the AUC of the CS for predicting dissatisfaction and desire for more pain treatment was 0.78 (95% CI: 0.74, 0.82) and 0.76 (95% CI: 0.72, 0.79), respectively. After adjusting for other items of patient perception of perioperative pain management, the adjusted AUC for predicting dissatisfaction and desire for more treatment by the CS was 0.90 (95% CI: 0.87, 0.92) and 0.80 (95% CI: 0.77, 0.83), respectively. The CS indicated superior discrimination to the worst pain in predicting dissatisfaction (AUC of worst pain = 0.65 [0.60,0.70], $p < 0.001$) and desire for more pain treatment (AUC of worst pain = 0.67 [0.63,0.71], $p < 0.001$). See Figure 1.

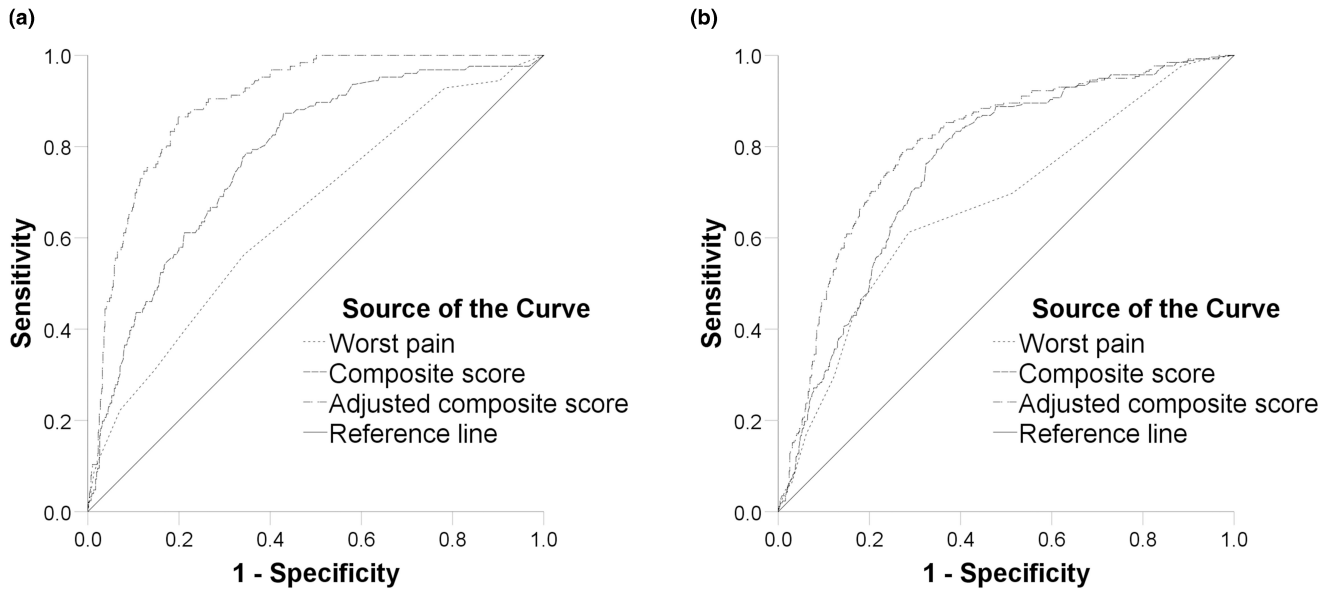

FIGURE 1 Comparison of the receiver operating characteristic curve between the composite score and the worst pain numeric rating scale in the validation set. (a) Predicting dissatisfaction with pain treatment; (b) predicting desire for more treatment. A composite score was calculated based on the weight (Table 3), and the value of each item was derived from the pain-related patient-reported outcomes. Adjusted composite score, after adjusting for other patient perceptions of perioperative pain management, including participation in treatment decisions, pain relief and receiving treatment information.

4 | DISCUSSION

A CS was developed from the 12 items of the IPO-Q, which comprise key patient-reported outcomes of perioperative pain experience, including pain intensity, physical and emotional functional interference and side effects. The CS indicated superior discrimination to the worst pain in predicting dissatisfaction and desire for more pain treatment. This novel CS may be useful for health centers to assess new strategies or techniques for perioperative pain treatment. Despite the lack of a gold standard for the assessment of pain experience, a tactical methodology was introduced to develop and validate this CS, which rationally weighs each component.

Although the items of PROs should be assessed separately to guide individual treatment, the total benefit provided by a new regimen may not be identified. Conversely, multiple outcome measures increase type I errors (Dai et al., 2013; Silverman et al., 1993). Establishing an algorithm to integrate various types of outcome measures is warranted (Dai et al., 2013). These scores will be useful in assessing a new strategy or technique engaged in pain treatment. Several integrated scores, such as 'quality of recovery scales' (Myles et al., 2016) and 'integrated pain/opioid consumption score' (Dai et al., 2013; Silverman et al., 1993), have been developed to address perioperative pain globally. However, these scores equally weighted the contributions of each item. A simple summary score might compromise the significance of some vital outcomes and may be less convincing. Moreover, some integrated scores (Dai et al., 2013; Silverman et al., 1993) did not take other important clinical information into account (Benzon et al., 2017). Pain intensity and side effects are appreciated, but functional and emotional status is rarely noted (Benzon et al., 2017). Because pain is multidimensional, several outcome domains should be assessed. The integration of several outcomes into a CS should be performed carefully. A rational score should comprise all crucial components and weigh them appropriately. The sum of the components does not make sense.

Because a gold standard in the assessment of acute pain experience has never been reached thus far (Silverman et al., 1993), it is a challenge to determine a methodology for developing and accessing this new score. We introduced principal component analysis to reduce dimensionality and a regression model to infer this score. PROs are divided into pain experience and patient perspective. We hypothesized that a rational score would be more associated with the patient's perspective, such as satisfaction and desire for more treatment, compared with an individual component, such as worst pain. The perspective is actually a description of the pain experience but is not equal to the experience. In PROs, the patient's perspective embraces five items, two of which are considered more crucial (Benzon et al., 2019; Greimel et al., 2018; Junewicz & Youngner, 2015; Komann et al., 2021; Rothaug et al., 2013; Stamer et al., 2021) and were used in this study.

We tactically took dissatisfaction and desire for more pain treatment together as a substitute for pain experience to test this score. Good discrimination for both dissatisfaction and desire for more treatment endorsed our hypothesis and methodology. This CS was used to assess pain experience.

The controversy surrounding the significance of satisfaction is evolving (Meißner et al., 2017). Although low satisfaction might be associated with high pain intensity (Benzon et al., 2019; Gerbershagen et al., 2011), inconsistency occurs at times (Greimel et al., 2017; Greimel et al., 2018; Roeb et al., 2017). Satisfaction is thought to be less associated with pain experience (Meißner et al., 2017). Actually, satisfaction encompasses two separate components: outcome satisfaction and process satisfaction (Benzon et al., 2019), although sometimes these two entities might not be distinguished explicitly. Process satisfaction depends on process-based parameters and could be more easily influenced by patient-specific factors. However, unlike chronic pain, acute pain is less affected by mental factors but by the pain experience itself (Benzon et al., 2019; Gerbershagen et al., 2011). Successful pain management is the major perioperative determinant of satisfaction (Benzon et al., 2019). Both pain and satisfaction are multidimensional and complex. The variables considered in these studies were possibly not enough to reveal the association between pain experience and satisfaction (Meißner et al., 2017). This suggests that it is not appropriate to assess the perioperative pain experience based on the worst pain or patient satisfaction only.

In this study, the good discrimination of CSs for dissatisfaction demonstrated that satisfaction was relevant to the pain experience (Meißner et al., 2017). Meanwhile, the CS was also strongly associated with the desire for more treatment, which was considered a global judgement of pain management in some studies (Komann et al., 2021; Stamer et al., 2021). The good and robust discrimination for both dissatisfaction and desire for more treatment across datasets might demonstrate that this score addressed the multidimensional nature of perioperative pain experience, and the hypothesis using dissatisfaction and desire for more treatment together as a substitute for pain experience was plausible.

Therefore, neither dissatisfaction nor desire should be considered as the outcome that this study aimed to predict. We aimed to predict pain experience. In contrast, we prefer the concept that dissatisfaction and desire are used to

predict pain experience. These two items both overlap the pain experience. We combined dissatisfaction and desire for more treatment to improve the pain experience, but neither was equal to pain experience and could not replace the CS.

This score could determine whether some changes in clinical practice benefit patients and whether some policies improve the capacity of healthcare centers for pain management. Accordingly, no cutoff for the CS was assigned because this score should be used to judge the effect of changes in medical practices. Because quality improvement is assessing changes in treatment patterns by comparing a baseline population of patients treated in a conventional manner to a group treated differently. The change in the score, but not the exact value, is important. Therefore, this score is not a convenient tool for individual treatment or satisfaction prediction but a relatively complex score embracing copious items derived from PROs.

It should be noted that this study did not aim to predict individual perioperative pain or guide relevant treatment. Consequently, the patient characteristics that would affect perioperative pain were not considered in this study. The aim of this study was not to describe the influence of patient characteristics on pain experience. Therefore, the item 'persistent pain before surgery' in the PROs was not included in the analysis.

Differences in the characteristics and PROs between the two sets were present to exhibit balance. This difference was never the aim. We expected a robust score that could work among a variety of populations and thus overcome an imbalance between the two datasets. This might indicate generalization of this score to other populations with different surgery types or experiencing severe or mild pain. Therefore, the variety of characteristics across or within the datasets was not important in this study and thus was not evaluated.

To further validate this new CS, controlled trials or cohort studies should be conducted to confirm causality, since the data collected in this study were cross-sectional. Furthermore, the minimally clinically important difference based on this CS should be ascertained (Khan & Butler, 2022). The clinically important threshold of change is meaningful to better interpret this score, which is not intuitive. The change could also be categorized into small, moderate or large deteriorations or improvements.

Health centers and policy makers would understand how different a new regimen or policy engaged in postoperative pain treatment has altered the patients' pain experience by comparing the CS reported by the trial and control population or after and before the new approach was used. A weighted CS might avoid problems caused by having multiple outcomes, especially when the effects of a change in clinical practice are inconsistent across these outcomes (Weinfurt & Reeve, 2022). However, conspicuous deteriorations in relevant items should also be noted and considered before a prudent decision is made (Benzon et al., 2017).

Additionally, data collection in this study was limited to the IPO-Q on POD1, which precluded addressing the impact on long-term outcomes. Concerns regarding the stability of observed changes in PROs have been raised (Khan & Butler, 2022). Analysing the results over time could help confirm that the outcomes observed are robust and sustained. Further studies comparing responses generated at 2 or more times are warranted. Moreover, if new questionnaires were engaged, another new score could be composited using the methodology introduced in this study. This might allow embracing more information and adapting the score for special purposes and situations.

4.1 | Limitations

It should be emphasized that the score is based on the IPO-Q and thus cannot be extrapolated to other situations. The intrinsic limitation of the IPO-Q, such as the lack of recording average pain and movement-evoked pain (Camiré et al., 2020), might limit the power of this score for the assessment. However, this study draws attention to a methodology that can develop an integrated score from several measures and has been validated across datasets. This study provides a ladder rather than an answer. This CS still requires further confirmation and improvement before its practical use.

As an intrinsic limitation, a CS removes the ability to distinguish the direction of the effect on multiple outcomes. A joint hypothesis-testing framework might be considered to ensure noninferiority in each outcome (Benzon et al., 2017; Dai et al., 2013).

Moreover, the findings of this study are relevant for Chinese patients undergoing orthopaedic surgery. Whether these results can be generalized to other populations remains unclear. However, the stability of discrimination between different sets encourages confidence. Furthermore, this study focused on satisfaction. However, psychological data were not collected. This may result in bias.

5 | CONCLUSIONS

This study introduced a methodology to integrate outcomes regarding pain experience into a CS, which was based on the weight of each item.

AUTHOR CONTRIBUTIONS

B Jiang: This author helped with data generation, collection, analysis, interpretation and the conception and

writing of the manuscript. Y Liu: This author helped with data generation, collection and writing of the manuscript. Y Wu: This author helped with data generation, collection and conception and writing of the manuscript. W Mi: This author helped with the conception and writing of the manuscript. Y Feng: This author helped with data analysis and interpretation and the conception and writing of the manuscript. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGEMENTS

The statistical review provided by Shenda Hong, PhD at the National Institute of Health Data Science, Peking University, Beijing, China, was greatly appreciated. The authors thank all the collaborators at PAIN OUT and CASP for the data analysed in this study. Furthermore, we appreciate the continuous support provided by PAIN OUT.

FUNDING INFORMATION

This work was supported by the National Key Research and Development Programme of China (grant number 2018YFC2001905). The PAIN OUT project was supported by the European Community's Seventh Framework Programme FP7/2007–2013 (grant number 223590).

CONFLICT OF INTEREST

None.

CLINICAL TRIAL NUMBER AND REGISTRY URL

The data collection methodology was registered before patient enrollment at www.clinicaltrials.gov (https://clinicaltrials.gov/) under the identifier NCT02083835 on March 11, 2014 (principal investigator: Winfried Meissner, MD). The statistical analysis plan of this study was admitted to PAIN OUT and approved by the Publication Board of PAIN OUT on 05 November 2019.

ORCID

Bailin Jiang https://orcid.org/0000-0002-9012-7472
Yi Feng https://orcid.org/0000-0002-3952-1351

REFERENCES

ASATFAPM. (2012). Practice guidelines for acute pain management in the perioperative setting: An updated report by the American Society of Anesthesiologists Task Force on acute pain management. *Anesthesiology*, *116*, 248–273.

Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing data in clinical research: A tutorial on multiple imputation. *The Canadian Journal of Cardiology*, *37*(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Baca, Q., Marti, F., Poblete, B., Gaudilliere, B., Aghaeepour, N., & Angst, M. S. (2021). Predicting acute pain after surgery: A multivariate analysis. *Annals of Surgery*, *273*, 289–298.

Benzon, H. T., Dunn, L. K., & Tran, D. Q. (2019). I Can't get no (patient) satisfaction. *Anesthesia and Analgesia*, *128*, 393–395.

Benzon, H. T., Mascha, E. J., & Wu, C. L. (2017). Studies on postoperative analgesic efficacy: Focusing the statistical methods and broadening outcome measures and measurement tools. *Anesthesia and Analgesia*, *125*, 726–728.

Camiré, D., Erb, J., Kehlet, H., Brennan, T., & Gilron, I. (2020). Movement-evoked pain versus pain at rest in postsurgical clinical trials and meta-analyses: Protocol for a follow-up systematic review. *JMIR Research Protocols*, *9*, e15309.

Dai, F., Silverman, D. G., Chelly, J. E., Li, J., Belfer, I., & Qin, L. (2013). Integration of pain score and morphine consumption in analgesic clinical studies. *The Journal of Pain*, *14*, 767–777. e768.

Gerbershagen, H. J., Rothaug, J., Kalkman, C. J., & Meissner, W. (2011). Determination of moderate-to-severe postoperative pain on the numeric rating scale: A cut-off point analysis applying four different methods. *British Journal of Anaesthesia*, *107*, 619–626.

Greimel, F., Maderbacher, G., Baier, C., Keshmiri, A., Schwarz, T., Zeman, F., Meissner, W., Grifka, J., & Benditz, A. (2018). Multicenter cohort-study of 15326 cases analyzing patient satisfaction and perioperative pain management: General, regional and combination anesthesia in knee arthroplasty. *Scientific Reports*, *8*, 3723.

Greimel, F., Maderbacher, G., Zeman, F., Grifka, J., Meissner, W., & Benditz, A. (2017). No clinical difference comparing general, regional, and combination anesthesia in hip arthroplasty: A multicenter cohort-study regarding perioperative pain management and patient satisfaction. *The Journal of Arthroplasty*, *32*, 3429–3433.

Junewicz, A., & Youngner, S. J. (2015). Patient-satisfaction surveys on a scale of 0 to 10: Improving health care, or leading it astray? *The Hastings Center Report*, *45*, 43–51.

Khan, M. S., & Butler, J. (2022). Stability of changes in health status: Next step in comprehensively assessing patient-reported outcomes. *JAMA*, *328*(10), 923–924. https://doi.org/10.1001/jama.2022.13551

Komann, M., Baumbach, P., Stamer, U. M., Weinmann, C., Arnold, C., Pogatzki-Zahn, E., & Meißner, W. (2021). 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. *The Journal of Pain*, *22*, 730–738.

Meißner, W., Komann, M., Erlenwein, J., Stamer, U., & Scherag, A. (2017). The quality of postoperative pain therapy in German hospitals. *Deutsches Arzteblatt International*, *114*, 161–167.

Myles, P. S., Myles, D. B., Galagher, W., Chew, C., MacDonald, N., & Dennis, A. (2016). Minimal clinically important difference for three quality of recovery scales. *Anesthesiology*, *125*, 39–45.

Roeb, M. M., Wolf, A., Gräber, S. S., Meißner, W., & Volk, T. (2017). Epidural against systemic analgesia: An international registry analysis on postoperative pain and related perceptions after abdominal surgery. *The Clinical Journal of Pain*, *33*, 189–197.

Rothaug, J., Zaslansky, R., Schwenkglenks, M., Komann, M., Allvin, R., Backström, R., Brill, S., Buchholz, I., Engel, C., Fletcher, D., Fodor, L., Funk, P., Gerbershagen, H. J., Gordon, D. B., Konrad, C., Kopf, A., Leykin, Y., Pogatzki-Zahn, E., Puig, M., … Meissner, W. (2013). Patients' perception of postoperative pain

management: Validation of the international pain outcomes (IPO) questionnaire. *The Journal of Pain*, *14*, 1361–1370.

Schnabel, A., Yahiaoui-Doktor, M., Meissner, W., Zahn, P. K., & Pogatzki-Zahn, E. M. (2020). Predicting poor postoperative acute pain outcome in adults: An international, multicentre database analysis of risk factors in 50,005 patients. *Pain Reports*, *5*, e831.

Silverman, D. G., O'Connor, T. Z., & Brull, S. J. (1993). Integrated assessment of pain scores and rescue morphine use during studies of analgesic efficacy. *Anesthesia and Analgesia*, *77*, 168–170.

Stamer, U. M., Bernhart, K., Lehmann, T., Setzer, M., Stüber, F., Komann, M., & Meissner, W. (2021). 'Desire for more analgesic treatment': Pain and patient-reported outcome after paediatric tonsillectomy and appendectomy. *British Journal of Anaesthesia*, *126*, 1182–1191.

Ultsch, A., & Lötsch, J. (2015). Computed ABC analysis for rational selection of Most informative variables in multivariate data. *PLoS One*, *10*, e0129767.

Vasta, S., Zampogna, B., Hartog, T. D., El Bitar, Y., Uribe-Echevarria, B., & Amendola, A. (2022). Outcomes, complications, and reoperations after meniscal allograft transplantation. *Orthopaedic Journal of Sports Medicine*, *10*, 23259671221075310.

Weinfurt, K. P., & Reeve, B. B. (2022). Patient-reported outcome measures in clinical research. *JAMA*, *328*, 472–473.

Zaslansky, R., Meissner, W., & Chapman, C. R. (2018). Pain after orthopaedic surgery: Differences in patient reported outcomes in the United States vs internationally. An observational study from the PAIN OUT dataset. *British Journal of Anaesthesia*, *120*, 790–797.

Zaslansky, R., Rothaug, J., Chapman, C. R., Bäckström, R., Brill, S., Fletcher, D., Fodor, L., Gordon, D. B., Komann, M., Konrad, C., Leykin, Y., Pogatski-Zahn, E., Puig, M. M., Rawal, N., Ullrich, K., Volk, T., & Meissner, W. (2015). PAIN OUT: The making of an international acute pain registry. *European Journal of Pain (London, England)*, *19*, 490–502.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jiang, B., Liu, Y., Wu, Y., Mi, W., & Feng, Y. (2022). A novel methodology to integrate outcomes regarding perioperative pain experience into a composite score: Prediction model development and validation. *European Journal of Pain*, *00*, 1–10. https://doi.org/10.1002/ejp.2033