Seminars in Arthritis and Rheumatism 75 (2025) 152842

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Measurement properties of the decisional conflict scale in people living with chronic non-cancer pain: a secondary analysis of the DECIDE-PAIN pan-Canadian survey

Florian Naye [a], Maxime Sasseville [b], Karine Toupin-April [c], France Légaré [d], Marie-Pierre Gagnon [b], Chloé Cachinho [a], Thomas Gérard [a], Valentin Vaillant [e], Olivia Dubois [a], Alison M. Hoens [f], Yannick Tousignant-Laflamme [a], Simon Décary [a,*]

[a] Université de Sherbrooke, Faculté de médecine et des sciences de la santé, École de réadaptation, 300112e Avenue Nord, Sherbrooke, Québec J1H 5N4, Canada
[b] Université Laval, Faculté des sciences infirmières, 1050, avenue de la Médecine, local 3645, Québec, Québec G1V 0A6, Canada
[c] University of Ottawa, Faculty of Health Sciences, School of Rehabilitation Sciences, 200 Lees Avenue, room 261, Ottawa, Ontario K1N 6N5, Canada
[d] Université Laval, Faculté de Médecine, Département de médecine de famille et de médecine d'urgence, 1050, avenue de la Médecine, local 4873, Québec, Québec G1V 0A6, Canada
[e] Institut de Kinésithérapie Podologie Orthopédie (IKPO) de la région sanitaire de Lille, 10 rue Saint Jean-Baptiste de la Salle, 59000, Lille, France
[f] University of British Columbia, Faculty of Medicine, Department of Physical Therapy, 2177 Wesbrook Mall, Vancouver, British Columbia V6T 1Z3, Canada

ARTICLE INFO

Keywords:
Decisional conflict scale
Chronic non-cancer pain
Shared decision-making
Measurement properties, OMERACT

ABSTRACT

Objectives: The Decisional Conflict Scale is a candidate instrument for the Core Outcome Measurement Set of the OMERACT Shared Decision-Making working group. However, its validity among chronic non-cancer pain populations is lacking. We explored measurement properties of the English version of the 16-item Decisional Conflict Scale in this population.
Methods: Informed by the COnsensus-based Standards for the selection of health status Measurement INstruments guidelines, we conducted a secondary analysis of a panCanadian cross-sectional survey. We assessed the distribution of score including ceiling and floor effects, readability, internal consistency, structural validity of an hypothesized five-factor model, and convergent and discriminant validities. We tested measurement invariances regarding age, sex, health literacy, and education.
Results: We collected data from 1103 respondents. We found ceiling effect at the item-level, while no ceiling and floor effects at the subscale and total score levels. Readability scores were below the sixth-grade reading level. All results of internal consistency were above the a priori threshold. The confirmatory factorial analysis resulted in a good model fit. We found good convergent, but poor discriminant validities. We confirmed measurement invariances for age, sex, health literacy, and education.
Conclusion: This secondary analysis of a panCanadian survey provides evidence of the measurement properties of the Decisional Conflict Scale in people living with chronic non-cancer pain. We propose that this instrument should be analyzed using its total score rather than its subscales due to a lack of discriminant validity. Further research is necessary to enhance its validity, possibly by re-examining its factor structure.

Introduction

Decisional conflict is a central construct in shared decision-making (SDM) studies as it is the most often reported outcomes of decision support tools trials (also known as SDM tools) [1,2]. Decisional conflict is defined as "a state of uncertainty about the course of action to take"

* Corresponding author at: Université de Sherbrooke, Faculté de médecine et des sciences de la santé, École de réadaptation, 3001, 12e Avenue Nord, Sherbrooke, Québec J1H 5N4, Canada.
E-mail addresses: florian.naye2@usherbrooke.ca (F. Naye), maxime.sasseville@fsi.ulaval.ca (M. Sasseville), ktoupina@uottawa.ca (K. Toupin-April), france.legare@fmed.ulaval.ca (F. Légaré), marie-pierre.gagnon@fsi.ulaval.ca (M.-P. Gagnon), chloe.cachinho2@usherbrooke.ca (C. Cachinho), thomas.gerard@usherbrooke.ca (T. Gérard), v.vaillant@ikpo.fr (V. Vaillant), olivia.dubois@usherbrooke.ca (O. Dubois), alison.hoens@ubc.ca (A.M. Hoens), yannick.tousignant-laflamme@usherbrooke.ca (Y. Tousignant-Laflamme), simon.decary@usherbrooke.ca (S. Décary).

https://doi.org/10.1016/j.semarthrit.2025.152842

Available online 3 October 2025
0049-0172/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

[3]. Globally, decisional conflict is understood within the Ottawa Decision Support Framework (ODSF) as a discrepancy between what people need to make a decision –termed decisional needs –and what is actually provided by healthcare providers [4]. Specifically, the ODSF purports that for an informed value congruent decision to be made, an individual needs to have access to the best evidence about the possible options (information), to be able to clarify what matters most to them (values and preferences), adequate support (including removing undue pressure) and have a sense of the favorable impact of the chosen option (impact) [5]. Any deficit in one of these factors will result in uncertainty about the choice to be made and results in future negative outcomes [6]. SDM aims to meet the people's decisional needs, thereby reducing decisional conflict and in the long term, decision regret [6]. This outlines its importance as a measurable construct in this field.

Decisional conflict is commonly measured with the Decisional Conflict Scale (DCS). Various versions have been developed for research and clinical practice, including the 16-item DCS, a 10-item low-literacy version, and the four-item SURE test [4,7]. The original English version was developed and validated in the context of prevention decisions (e.g., influenza immunization, breast cancer screening) and demonstrated good internal consistency, test-retest reliability, and known-group validity [3]. Subsequent studies extended the use of the DCS to other contexts, such as end-of-life care [8] and genetic testing [9], supporting its structural, convergent, divergent, known-group, and predictive validity as well as internal consistency. Most subsequent uses of the DCS across clinical areas have not included psychometric re-evaluations; researchers often rely on the original validation data, even when applying the instrument to new conditions and decision types [10]. To our knowledge, no previous study has specifically assessed the measurement properties of the DCS in chronic pain populations, where rheumatic and musculoskeletal disorders are prevalent, despite the complex, preference-sensitive nature of decisions in these contexts [11, 12].

OMERACT (Outcome Measures in Rheumatology) is an independent international initiative comprising researchers, clinicians and patients leading efforts in developing Core Outcome Sets [13]. A Core Outcome Set refers to a minimal standardized set of outcomes measures, which should be used and reported as a minimum in all clinical trials on a specific field [14,15]. OMERACT employs a rigorous step-by-step approach to developing Core Outcome Sets including: 1) defining a Core Domains Set that should be measured in all randomized controlled trials and longitudinal observational studies (i.e., *what* to measure in terms of outcomes, also called domains) [16] and 2) determining the Core Outcome Measurement Set (i.e., *how* to measure the domains) [17, 18].

The OMERACT SDM working group aims to identify a Core Outcome Set that should be measured all randomized controlled trials and longitudinal observational studies in SDM field. We previously reached consensus on the Core Domains Set [19–21]. We then conducted a scoping review to identify candidate instruments for assessing the core domains [17]. This scoping review focused on measurement instruments used in SDM trials involving people living with rheumatic and musculoskeletal diseases (RMDs). Among 220 measurement instruments identified across 99 studies, the DCS was the most frequently used (38 % of the studies), suggesting its potential as a candidate instrument [10]. However, this scoping review raised concerns as it revealed limited evidence of validity of the DCS in RMD populations [10]. It found that only internal consistency had been evaluated using data from people living with RMDs [22–24]. This concern could hinder the next step of the OMERACT process (using the OMERACT Filter 2.2), as robust measurement properties are critical for selecting of the Core Outcome Measurement Set [17]. This emphasizes the need to expand the knowledge base regarding the measurement properties of the DCS.

The aim of this study was to explore measurement properties of the English version of the 16-item DCS using a panCanadian dataset of people living with chronic noncancer pain, including RMDs.

Methods

Study design and settings

This study is a secondary analysis of a cross-sectional online survey of people from the 10 Canadian provinces (published protocol available [25]). We used the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) Risk of Bias checklist for Patient-Reported Outcome Measures [26] to inform the reporting of results.

Eligibility criteria and recruitment

In the primary study, we recruited community-dwelling people living with chronic non-cancer pain from the largest Canadian-owned market research and analytics company, Leger Marketing (https://leger360.com/). This company maintains a panel of 500,000 active members of Canadian society with Internet access from the 10 Canadian provinces. We used a stratified proportional random sampling based on the population and chronic pain prevalence of each province [27,28]. We recruited a sample of citizens, permanent residents or refugees living in Canada, aged $\geq$ 18 years old, with primary chronic pain (e.g., low back pain, fibromyalgia, migraine) or secondary pain (e.g., postsurgical pain, pain due to structural changes, pain associated with a disease of the nervous system). The International Association for the Study of Pain defined chronic primary pain as pain in one or more anatomical regions that persists or recurs for longer than three months and that cannot be better accounted for by another chronic pain condition [29], and chronic secondary pain as chronic pain that is linked to other diseases as the underlying cause, for which pain may initially be regarded as a symptom [30]. Respondents had to read, write, and understand one of the two official Canadian languages (i.e., French or English). We excluded individuals with cancer-related chronic pain because their decision-making context differs fundamentally from that of noncancer chronic pain. Cancer care often involves high-stakes decisions tied to life expectancy, curative intent, and palliative care, whereas noncancer pain management typically focuses on long-term symptom relief and quality of life. Cancer treatments such as chemotherapy or radiation introduce distinct risks and benefits, and the psychological impact of a life-threatening diagnosis may influence how conflict and regret are experienced [31]. In this context, choices may feel "obvious" or survival-driven—for example, accepting aggressive treatment despite side effects—which can reduce decisional conflict and regret [32]. In contrast, individuals with noncancer chronic pain face more preference-sensitive decisions involving uncertain outcomes and lifestyle changes. Including cancer-related pain could therefore introduce confounding factors and mask patterns specific to noncancer pain populations. Randomization method and strategy to achieve a representative sample are detailed in the published protocol [25]. For this study, we only selected English respondents from the dataset.

Ethical considerations

We obtained the ethics approval from the Research Ethics Board of the Research Centre at the Centre Hospitalier Universitaire de Sherbrooke (project #2022–4645). We conducted the survey following the Canadian Personal Information Protection and Electronic Documents Act [33] and the Marketing Research and Intelligence Association's Charter of Respondent Rights [34]. Respondents consented to participate in the study. All respondents had to give their consent before their participation and could request to be removed from the study at any time. Respondents also had to consent to Leger Marketing's terms of use and privacy policy [35].

Data collection

Instrument

The DCS is informed by the ODSF [3] and tell us about modifiable factors that make an inherently uncertain decision even more difficult [5]. The 16 items are grouped into five subscales, each designed to capture information on these modifiable factors, namely: uncertainty, lack of information, unclear values, feeling unsupported, and ineffective decision-making [4]. The items are rated on a 5-point Likert scale ranging from 0 = "strongly agree" to 4 = "strongly disagree" [3]. A continuous total score is obtained by the sum of the items' scores divided by 16 and multiplied by 25, giving a measure ranging from 0 = no decisional conflict to 100 = extremely high decisional conflict [36]. As well, it is possible to compute the scores of the subscales [36]. We gathered the DCS score related to the most difficult decision faced by the respondents across their care pathway. These decisions are mainly related to treatment options or diagnostic tests. The DCS is freely available on https://decisionaid.ohri.ca/eval_dcs.html .

Other variables assessed

We collected data on age, sex at birth, pain duration, education level, and if the respondents face difficulties understanding what healthcare providers said about their condition (i.e., functional health literacy).

Data analysis

We provide a summary of the assessed measurement properties, their definition, and the threshold used to interpret our results in Table 1. We performed the statistical analyses in R (version 4.4.1) and was supported by biostatisticians. We performed descriptive analysis (mean and standard deviation (SD), frequency and percentage) of the sample to summarize respondents' characteristics.

Distribution of scores

We used descriptive analyses of the DCS items. For each categorical item, we reported mean, SD, median, first quartile, third quartile, minimum, maximum, skewness, and kurtosis. We also calculated the proportion of respondents for each option response. For each item, we reported the proportion of respondents who selected 0 ('totally agree') and 4 ('totally disagree'). For the subscale and total scores, we reported the proportion of respondents who achieved a score of 0 or 100. A ceiling or floor effect was defined as at least 15 % of respondents scoring at the highest or lowest threshold [47].

Readability

To assess the readability of the scale, we used the Microsoft Word Readability test and reported Flesch-Kincaid Grade Level and the Flesch Reading Ease [48]. The Flesch-Kincaid Grade Level provides a score that corresponds to the US grade level, we expected a score $\langle$ 6 [49–51]. The Flesch Reading Ease scores ranged between 0 and 100, we expected a score $\rangle$ 80 indicating at least an easy style corresponding to a sixth-grade reading level [52,53].

Internal consistency

We calculated Cronbach's alpha and McDonald's omega for each factor, and for the overall scale [38]. We considered values above 0.70 as acceptable [39].

Structural validity

We assessed all relevant assumptions prior to conducting confirmatory factor analyses, including factorability, local independence, and multicollinearity. Details of these evaluations are provided in Supplementary Material 1. Previous studies reported the dimensionality of the DCS [3,4,54,55]. We performed a five-factor confirmatory factorial analysis (CFA), containing three items for factors 1 to 4 and four items for factor 5. Due to the ordinal nature of the items, we used the Weighted Least Squares Mean and Variance-adjusted estimation method [56]. We reported the following fit statistics and recommended criteria to assess the overall adequacy of the model: Comparative Fit Index (CFI) $>$ 0.95 [39], Tucker-Lewis Index (TLI) $>$ 0.95 [39], Standardized Root Mean Square Residual (SRMR) $<$ 0.08 [39], and Root

Table 1
Summary of the measurement properties, definition, and threshold used.

Measurement properties	Definition	Threshold
Reliability		
Internal consistency	The degree of the interrelatedness among the items [37]	Cronbach's alpha/Mc Donald's omega of the total score $\geq$ 0.70 [38,39]
Construct validity		
Structural validity	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured [37]	CFI $>$ 0.95 or LTI $>$ 0.95 or SRMR $<$ 0.08 or RMSEA $<$ 0.06 [39]
Measurement invariance	Measurement invariance assesses the (psychometric) equivalence of a construct across groups or measurement occasions and demonstrates that a construct has the same meaning to those groups or across repeated measurements [40]	ΔCFI $\leq$ 0.01 combined with ΔRMSEA $\leq$ 0.015 or ΔSRMR $\leq$0.01 was indicative of invariance [41]
Convergent validity	Convergent validity is related to whether a latent variable is well estimated with the selected indicators (i.e., whether a latent variable represents a construct which researchers intend to estimate) [42]	Average Variance Extracted $\geq$ 0.5 and Standard Regression Weight $\geq$ 0.5 and Composite Reliability $\geq$ 0.7 [43]
Discriminant validity	Discriminant validity is referring to the extent in which the construct is differing from one another empirically. It also measures the degree of differences between the overlapping construct [44]	Evidence of convergent validity is established, and no item cross-loads on other factors (factor loading $>$ 0.4 on several factors [45]) and the Average Variance Extracted or each construct greater than the squared inter-factor correlations and the correlation between two factors is assessed at three levels (0.85, 0.80, and 0.70) [38]
Interpretability		
Ceiling effect	Ceiling effect occurs when a considerable proportion of respondents endorse the worst score [46]	15 % of the respondents scoring at the highest threshold (i.e., 5 on the 5-point Likert scale for the item level, 100/100 for the subscale and total scores [47]
Floor effect	Floor effect occurs when a considerable proportion of respondents endorse the best score [46]	15 % of the respondents scoring at the lowest threshold (i.e., 0 on the 5-point Likert scale for the item level, 0/100 for the subscale and total scores [47]

CFI: Comparative Fit Index, DCS: Decisional Conflict Scale, LTI: Lewis-Tucker Index, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual.

Mean Square Error of Approximation (RMSEA) < 0.06 [39].

Convergent validity

From the five-factor CFA, we tested convergent validity. Based on the criteria proposed by Hair et al [38,43], we defined good convergent validity if the Standard Regression Weight ≥ 0.5, the Average Variance Extracted ≥ 0.5, and the Composite Reliability ≥ 0.7.

Discriminant validity

From the five-factor CFA, we tested discriminant validity using the recommendations by Cheung et al [38]. We defined good discriminant validity [38] if 1) evidence of convergent validity is established, 2) no item cross-loads on other factors (factor loading > 0.4 on several factors [45], 3) the Average Variance Extracted or each construct greater than the squared inter-factor correlations, and 4) the correlation between two factors is assessed at three levels (0.85, 0.80, and 0.70).

Measurement invariance

We used Multigroup Confirmatory Factorial Analysis (MCFA) to test for configural, metric, scalar, and strict invariances. We conducted this analysis with the Weighted Least Squares Mean and Variance (WLSMV)-adjusted estimation method [56]. However, during our data screening, we identified a response option of an item with a frequency of zero, which could lead to estimation issues with WLSMV. To address this issue and ensure the robustness of our results, we shifted to using the Maximum Likelihood (ML) estimation method for this analysis [57]. To ensure transparency and reproducibility, we clearly disclosed the estimation method alongside our results.

We examined measurement invariance across age, sex, health literacy, and education level that are commonly used to compare groups in SDM studies. Establishing measurement invariance for these variables can thus validate that differences observed in the DCS are due to actual differences in decision-making processes rather than artifacts of the methods. If the samples were different for relevant characteristics, we adjusted the MCFA accordingly. We concluded to configural invariance if the model had good model fit [58] (i.e., CFI > 0.95 [39], TLI > 0.95 [39], SRMR < 0.08 [39], and RMSEA < 0.06 [39]. Comparisons of successive constrained models and subsequent conclusions on invariance were made based on the guidelines provided by Chen [41]: ΔCFI ≤ 0.01 combined with ΔRMSEA ≤ 0.015 or ΔSRMR ≤ 0.01 was indicative of invariance. For this comparison step, metric invariance model was compared to configural invariance model, scalar invariance was compared to metric invariance model, and strict invariance model was compared to scalar invariance model [40].

Post-hoc power calculation

We used the *semPower.postHoc* function from the *semPower* package [59] to calculate the statistical power of the confirmatory factorial analysis. With a sample size of 1103 respondents, fixed alpha risk at 0.05 and expected root mean square error of approximation at 0.06 [39], we obtained a statistical power of 100 % for the confirmatory factorial analysis.

Results

Characteristics of the respondents

We describe the respondents' characteristics in Table 2. Of the 1103 respondents, mean age was 51 years old (SD =16.4) and pain duration ranged from 3 months to 59 years. Half of respondents were females (52.1 %), most have difficulties to understand what healthcare providers say about their condition (71.3 %), and 50 % have a diploma lesser than university degree. Participants reported a range of ethnocultural backgrounds, with the possibility of selecting multiple identities. The most frequently reported backgrounds were European (n = 423, 38.3 %), North American (n = 325, 29.5 %), and Asian (n = 132, 12.0 %). Approximately 3.8 % (n = 42) of participants identified as Indigenous. Among those reporting multiple backgrounds, the most common combinations were European and North American (n = 66, 6.0 %), Indigenous and North American (n = 8, 0.7 %), and Indigenous and European (n = 7, 0.6 %). Overall, 29 distinct identity profiles were observed, reflecting a diverse and multidimensional ethnocultural composition within the sample.

Distribution of scores

DCS items mean ranged from 1.08 to 1.66 and all the response options were selected at least once. Response options 4 and 5 were less selected as indicated with the skewness ranged between 0.27 and 0.80. We provided histograms of the distribution for each item in Supplementary Material 2. At the item level, all items except item #12 (i.e., *This decision is easy for me to take*) exhibited a floor effect, with 16.1 % to 30.2 % of respondents selecting 'strongly agree', but no ceiling effect were observed. At the subscale level, neither ceiling nor floor effects were present. At the total score level, 5.0 % of respondents achieved a DCS total score of 0, while none scored 100, indicating the absence of floor and ceiling effects. We provided more information on the distribution of scores in Table 3.

Readability

We obtained a Flesch-Kincaid Grade Level of 3.6 out of 12 and a Flesch Reading Ease score of 84.9 out of 100.

Internal. consistency

All results were above the threshold of 0.70. We provide the subscale and overall findings in Table 4.

Structural validity

All assumptions were met for confirmatory factor analysis, as detailed in Supplementary Material 1. With the WLSMV-adjusted estimation method, the five-factor confirmatory factorial analysis resulted in a good model fit. We obtained robust CFI = 0.958, robust TLI = 0.946, robust RMSEA = 0.081 (90 % CI [0.074; 0.084]), and SRMR = 0.026.

Convergent validity

We found Standard Weight Regression ranged from 0.71 to 0.90, Average Variance Extracted from 0.60 to 0.72, and Composite Reliability from 0.82 to 0.91, resulting in good convergent validity. We provide all the results in Table 5.

Discriminant validity

We found evidence of convergent validity, but cross-loadings between items and factors. Each factor's Average Variance Extracted estimate was lower than the squared inter-factor correlations, indicating that the factors shared more variance with each other than with their own items. All inter-factor correlations were greater than 0.85, suggesting a high degree of overlap between the factors, indicating poor discriminant validity. We provide details in Table 5 and Supplementary Material 3.

Measurement invariance

Due to the low proportion of intersex respondents (n = 4, 0.4 %), we

removed these respondents from the analysis. We also removed 111 respondents with missing data on the pain duration variable and 3 respondents with missing data in the education variable. We performed the measurement invariance analyses on 985 respondents. We provide all the measurement invariance results in Table 6. We provide details on subgrouping, adjustment, and reason on the choice of the estimation method for each measurement invariance in Supplementary Material 4.

Discussion

We assessed the measurement properties of the English version of the 16-item Decisional Conflict Scale (DCS) among individuals living with chronic non-cancer pain, using data from a pan-Canadian cross-sectional survey. Our analysis showed that the total DCS score did not exhibit floor or ceiling effects, and the scale was readable and internally consistent. We validated the hypothesized five-factor structure and found good convergent validity, although discriminant validity remained limited. We confirmed measurement invariance across age, sex, health literacy, and education level—key variables in shared decision-making research. These findings support the use of the English DCS in this population and context. Our findings led us to five observations.

First, our study found evidence supporting the structural and convergent validities of the DCS, as well as its internal consistency, indicating that it is a robust tool for research and clinical practice. Its readability below the sixth grade reading level ensures that respondents can provide accurate responses by easily understanding the questions [51,60]. However, the lack of discriminant validity suggests that the DCS is best interpreted using the total score rather than its subscales [61]. The DCS demonstrates measurement invariance for several variables, indicating that different groups within each variable interpret the decisional conflict construct in a consistent manner. Our measurement invariance results allow for valid comparisons across subgroups [40]. For instance, a recent survey identified age as a risk factor for decisional conflict in individuals living with chronic pain [62], highlighting the need for comparative interventions effectiveness analysis across different age groups. Our measurement confirms invariance of the DCS regarding age, researchers can then confidently interpret any observed differences—or the absence of differences—between these groups [40]. Our invariance measurement also enhances the generalizability of findings, because subgroups interpret the decisional conflict construct consistently [63]. However, we encountered issues with the covariance matrix in several models. After confirming that the matrix was positive definite (i.e., all eigenvalues were positive), we attributed these problems to high inter-factor. Readers should interpret the results with caution because these models might be sensitive to small changes in the data. The DCS appears to be an effective measurement instrument for assessing decisional conflict in people living with non-cancer chronic pain, enabling valid subgroup comparisons and supporting the generalizability of research findings.

Second, the DCS demonstrates an adequate level of readability. We found that the DCS has a readability level below the sixth grade, a benchmark commonly referenced in information science [49,50]. This finding outlines the DCS's suitability for broad implementation in SDM, where clear and accessible communication is essential for decision-making. SDM actively involves individuals in their health-related decisions, making effective information exchange and comprehension central to the process [64]. An implicit clinical bias may exist whereby individuals with limited formal education are perceived as less in need of SDM [65,66]. This perception can result in diminished engagement in decision-making processes and exacerbate health inequities, despite substantial evidence indicating that the most vulnerable populations derive the greatest benefit from SDM [67]. It is necessary to use easy-to-read measure to ensure that all individuals, regardless of their educational background, can participate fully in SDM. The DCS is an adapted measure to promote inclusivity in SDM across clinical and research settings.

Third, with a low item score at baseline, the DCS may struggle to detect small but meaningful changes, potentially leading to an underestimation of an intervention's effectiveness. Although no ceiling effect was observed in the total score, our item-level analysis revealed that over 60 % of respondents selected the agreement options (i.e., strongly agree or agree), while fewer than 5 % chose the disagreement options (i.e., disagree and strongly disagree). In this study, the DCS was applied to assess the most difficult decision respondents faced throughout their care pathway. Despite the challenging nature of these decisions, most items were rated at the lower end of the scale, suggesting that participants perceived relatively low decisional conflict overall. This positive skew may reflect adaptation in long-term care, but could also result from social desirability bias, retrospective rationalization, or limitations in how the DCS conceptualizes and captures conflict. Notably, the DCS relies on strong, absolute statements (e.g., feeling sure, knowing risks and benefits), which may discourage moderate responses and fail to capture nuanced or unresolved decisional tensions. This distribution poses a challenge for clinical trials of SDM interventions, where baseline scores close to the floor may restrict the ability to detect improvement.

Fourth, the DCS may require modifications to enhance its validity. Our findings do not support the discriminant validity of subscores, as the five factors of the scale do not sufficiently distinguish between constructs. High inter-factor correlations and cross-loadings, which result in low eigenvalues for four out of the five factors, confirm a lack of discriminant validity and raise concerns about the scale's ability to accurately assess distinct constructs of decisional conflict. These issues suggest that the subscale structure is suboptimal, with overlapping constructs likely resulting from inadequate concept elicitation or poorly defined domains. Originally, the DCS was developed based on three clinical considerations: decision uncertainty, factors contributing to uncertainty, and perception of effective decision-making [3]. Over time, the "factors contributing to uncertainty" domain was subdivided into three distinct domains: informed, values clarity, and support. This refinement may have introduced some degree of conceptual overlap, which has been noted in previous studies reporting concerns with the structure of the DCS [4,9,54,68,69]. Given these findings, researchers should interpret DCS results cautiously, as overlapping factors may conceal meaningful differences in decisional conflict. Further research is necessary to refine the scale by re-examining its factor structure and addressing the conceptual overlap among constructs.

This study contributes to the OMERACT initiative on instrument selection by addressing key elements of the Filter 2.2 process. Following the identification of the Decisional Conflict Scale (DCS) as a candidate instrument for the Core Outcome Measurement Set in shared decision-making (SDM) research, our study provides empirical data relevant to both Step 1 and Step 2 of the OMERACT Filter 2.2 [17]. In Step 1, we provide feasibility information, including a sixth-grade readability level that supports accessibility and no floor and ceiling effects at the total score level, suggesting adequate interpretability. In Step 2, we contribute evidence on several measurement properties of the DCS, including internal consistency, structural validity, convergent validity, and measurement invariance across key variables such as age, sex, education, and health literacy. Our findings provide a strong foundation to support the DCS's ongoing consideration within the OMERACT Core Outcome Measurement Set selection process and highlight the need for longitudinal assessment (e.g., test-retest reliability, measurement error) to complete its appraisal.

Strength and limitations

A significant strength of our study lies in the robustness of the sample used. This secondary analysis drew from a representative panCanadian survey, ensuring the inclusion of diverse respondents with varied demographic and clinical profiles. The substantial sample size facilitated rigorous testing of measurement invariance across several key variables,

with each subgroup containing a sufficient number of respondents to yield reliable results. However, we did not assess the model across other subgroups (e.g., type of decision, native language) due to concern for subgroup sample size. This study has limitations. One limitation is that we could not distinguish between specific chronic pain conditions, such as RMDs, neuropathic pain, or other etiologies. We collected participants' diagnoses through a self-reported survey question. However, during initial data cleaning and quality control procedures, we identified concerns regarding the accuracy of some responses. As we were unable to verify these diagnoses through medical records, we made the decision to exclude this variable from our analyses. As a result, we did not assess measurement invariance between RMDs and non-RMDs populations. However, during initial data cleaning and quality control procedures, we identified concerns regarding the accuracy of some responses. As we were unable to verify these diagnoses through medical records, we made the decision to exclude this variable from our analyses. As a result, we did not assess measurement invariance between RMDs and non-RMDs populations. However, individuals living with chronic pain—including those with RMDs—often face similar types of decisions, such as navigating diagnostic uncertainty, weighing the modest benefits and risks of long-term treatments, and adapting to recommended lifestyle changes. The DCS captures constructs that are relevant across these contexts. Given the high prevalence of RMDs in chronic pain populations, our sample likely includes a substantial proportion of individuals with RMDs [27,70]. Our findings therefore provide preliminary evidence for the applicability of the DCS in populations aligned with OMERACT's focus. The original design of this study did not focus on evaluating the measurement properties of the DCS. As responses to all DCS items were mandatory, we were unable to assess the applicability criterion related to missing data. Moreover, the cross-sectional design was not offering us to study the test-retest reliability, measurement error, and responsiveness of the DCS. Since our data collection relied on participants' recall of a past decision, the reported levels of decisional conflict may be subject to recall bias, which could compromise the accuracy of responses and influence the interpretation of our findings.

Conclusion

This secondary analysis of a panCanadian cross-sectional study in 1103 people living with chronic noncancer pain provides evidence of validity and applicability of the 16-item DCS. The DCS has a sixth-grade reading level measure that presents good pragmatic criteria, making it a valuable instrument to implement in SDM research. We found measurement invariances for several key SDM variables, establishing the DCS as an efficient measurement instrument for enabling valid comparisons across groups and enhancing generalizability of the findings. Further research is necessary to enhance its validity, possibly by re-examining its factor structure.

Ethical considerations

The Research Ethics Board of the Research Centre at the Centre Hospitalier Universitaire de Sherbrooke approved our study (approval #2022–4645) on May 05, 2022.

Consent to participate

Respondents gave written consent before starting the survey. Respondents provided informed consent to conduct the study and use any personal information. Data has been anonymized.

Consent for publication

Not applicable

Funding statement

Financial support for this study was provided entirely by a grant from Canadian Institutes of Health Research. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

CRediT authorship contribution statement

Florian Naye: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maxime Sasseville:** Writing – review & editing, Methodology, Conceptualization. **Karine Toupin-April:** Writing – review & editing, Methodology, Conceptualization. **France Légaré:** Writing – review & editing, Methodology, Conceptualization. **Marie-Pierre Gagnon:** Writing – review & editing, Methodology, Conceptualization. **Chloé Cachinho:** Writing – review & editing, Methodology, Conceptualization. **Thomas Gérard:** Writing – review & editing, Methodology, Conceptualization. **Valentin Vaillant:** Writing – review & editing, Methodology, Conceptualization. **Olivia Dubois:** Writing – review & editing, Methodology, Conceptualization. **Alison M. Hoens:** Writing – review & editing, Methodology, Conceptualization. **Yannick Tousignant-Laflamme:** Writing – review & editing, Methodology, Conceptualization. **Simon Décary:** Writing – original draft, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Samuel Lemaire-Paquette, biostatistician, for his support in this article.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2025.152842.

References

[1] Stacey D, Lewis KB, Smith M, Carley M, Volk R, Douglas EE, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2024;1(1):Cd001431.
[2] Garvelink MM, Boland L, Klein K, Nguyen DV, Menear M, Bekker HL, et al. Decisional conflict scale findings among patients and surrogates making health decisions: part ii of an anniversary review. Med Decis Mak 2019;39(4):315–26.
[3] O'Connor AM. Validation of a decisional conflict scale. Med Decis Mak 1995;15(1):25–30.
[4] Garvelink MM, Boland L, Klein K, Nguyen DV, Menear M, Bekker HL, et al. Decisional conflict scale use over 20 years: the anniversary review. Med Decis Mak 2019;39(4):301–14.
[5] O'Connor AM, Jacobsen MJ, Stacey D. An evidence-based approach to managing women's decisional conflict. J Obstet Gynecol Neonatal Nurs 2002;31(5):570–81.
[6] Hoefel L, O'Connor AM, Lewis KB, Boland L, Sikora L, Hu J, et al. 20th anniversary update of the ottawa decision support framework part 1: a systematic review of the decisional needs of people making health or social decisions. Med Decis Mak 2020;40(5):555–81.
[7] Légaré F, Kearing S, Clay K, Gagnon S, D'Amours D, Rousseau M, et al. Are you SURE?: assessing patient decisional conflict with a 4-item screening test. Can Fam Physician 2010;56(8):e308–14.
[8] Song MK, Sereika SM. An evaluation of the Decisional Conflict Scale for measuring the quality of end-of-life decision making. Patient Educ Couns 2006;61(3):397–404.
[9] Katapodi MC, Munro ML, Pierce PF, Williams RA. Psychometric testing of the decisional conflict scale: genetic testing hereditary breast and ovarian cancer. Nurs Res 2011;60(6):368–77.
[10] Naye F, Toupin-April K, de Wit M, LeBlanc A, Dubois O, Boonen A, et al. OMERACT Core outcome measurement set for shared decision making in rheumatic and

musculoskeletal conditions: a scoping review to identify candidate instruments. Semin Arthritis Rheum 2023;65:152344.
[11] Naye F, Légaré F, Cachinho C, Gérard T, Toupin-April K, Sasseville M, et al. People living with chronic pain in Canada face difficult decisions and decisional conflict concerning their care: data from the national DECIDE-PAIN survey. BMC Prim Care 2024;25(1):424.
[12] Matthias M.S., Henry S.G. Reducing frustration and improving management of chronic pain in primary care: is shared decision-making sufficient? Journal of general internal medicine. 2022;37 (1):227–8.
[13] Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials 2007;8:38.
[14] Clarke M, Williamson PR. Core outcome sets and systematic reviews. Syst Rev 2016;5:11.
[15] Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials 2007;8:39.
[16] Maxwell LJ, Beaton DE, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Core Domain Set Selection According to OMERACT Filter 2.1: the OMERACT Methodology. J Rheumatol 2019;46(8):1014–20.
[17] Maxwell LJ, Beaton DE, Boers M, D'Agostino MA, Conaghan PG, Grosskleg S, et al. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: filter 2.2. Semin Arthritis Rheum 2021;51(6):1320–30.
[18] Beaton D., Grosskleg S., Shea B.J., Tugwell P., Bingham III C.O., Conaghan PG, et al. The omeract handbook version 2.1.2021.
[19] Toupin-April K, Barton JL, Fraenkel L, Meara A, Li LC, Brooks P, et al. OMERACT development of a core domain set of outcomes for shared decision-making interventions. J Rheumatol 2019;46(10):1409–14.
[20] Toupin-April K, Décary S, de Wit M, Meara A, Barton JL, Fraenkel L, et al. Endorsement of the OMERACT core domain set for shared decision making interventions in rheumatology trials: results from a multi-stepped consensus-building approach. Semin Arthritis Rheum 2021;51(3):593–600.
[21] Décary S, de Wit M, Naye F, Barton JL, Fraenkel L, Li LC, et al. Consensus on the definitions and descriptions of the domains of the OMERACT Core Outcome Set for shared decision making interventions in rheumatology trials. Semin Arthritis Rheum 2024;65:152381.
[22] Gasteiger C, Groom KM, Lobo M, Scholz U, Dalbeth N, Petrie KJ. Is Three a Crowd? The influence of companions on a patient's decision to transition to a biosimilar. Ann Behav Med 2022;56(5):512–22.
[23] Meade T, Dowswell E, Manolios N, Sharpe L. The motherhood choices decision aid for women with rheumatoid arthritis increases knowledge and reduces decisional conflict: a randomized controlled trial. BMC Musculoskelet Disord 2015;16:260.
[24] Simon D, Kriston L, von Wolff A, Buchholz A, Vietor C, Hecke T, et al. Effectiveness of a web-based, individually tailored decision aid for depression or acute low back pain: a randomized controlled trial. Patient Educ Couns 2012;87(3):360–8.
[25] Naye F, Légaré F, Paquette JS, Tousignant-Laflamme Y, LeBlanc A, Gaboury I, et al. Decisional needs assessment for patient-centred pain care in Canada: the DECIDE-PAIN study protocol. BMJ Open 2023;13(5):e066189.
[26] Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res 2024;33(11):2929–39.
[27] Schopflocher D, Taenzer P, Jovey R. The prevalence of chronic pain in Canada. Pain Res Manag 2011;16(6):445–50.
[28] Statistics Canada. Population And Demography Statistics 2022 [Available from: https://www.statcan.gc.ca/en/subjects-start/population_and_demography.
[29] Nicholas M, Vlaeyen JWS, Rief W, Barke A, Aziz Q, Benoliel R, et al. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain 2019;160(1):28–37.
[30] Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP classification of chronic pain for the international classification of diseases (ICD-11). Pain 2019;160(1):19–27.
[31] Zaza C, Baine N. Cancer pain and psychosocial factors: a critical review of the literature. J Pain Symptom Manage 2002;24(5):526–42.
[32] Shrestha A, Martin C, Burton M, Walters S, Collins K, Wyld L. Quality of life versus length of life considerations in cancer patients: a systematic literature review. Psychooncology 2019;28(7):1367–80.
[33] Personal information protection and electronic documents act, (2000).
[34] Charter of respondents rights, (2006).
[35] Marketing L. Priv Not 2023 [Available from: https://leo.tech/privacy-statement/.
[36] O'Connor AM. User manual - Decisional Conflict scale ottawa. Hospital Research Institute; 2010 [Available from: http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf.
[37] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63(7):737–45.
[38] Cheung GW, Cooper-Thomas HD, Lau RS, Wang LC. Reporting reliability, convergent and discriminant validity with structural equation modeling: a review and best-practice recommendations. Asia Pac J Manag 2024;41(2):745–83.
[39] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011.
[40] Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev 2016;41:71–90.
[41] Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Multidiscip J 2007;14(3):464–504.
[42] Lee D. The convergent, discriminant, and nomological validity of the depression anxiety stress scales-21 (DASS-21). J Affect Disord 2019;259:136–42.
[43] Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis, 7th edition. Pearson; 2009.
[44] Hair JF, Hult GTM, Ringle CM, Sarstedt MA. Primer On Partial Least Squares Structural Equation Modeling (PLS-SEM), 2nd edition. SAGE Publications, Inc; 2016.
[45] Howard MC. A systematic literature review of exploratory factor analyses in management. J Bus Res 2023;164:113969.
[46] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4(4):293–307.
[47] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34–42.
[48] Lewis CC, Mettert KD, Stanick CF, Halko HM, Nolen EA, Powell BJ, et al. The psychometric and pragmatic evidence rating scale (PAPERS) for measure development and evaluation. Implement Res Pr 2021;2:26334895211037391.
[49] Weiss BD. Health literacy and patient safety: help patient understand. Manual for clinicians. 2nd edition. AMA Foundation; 2007.
[50] Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, et al. Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019;34(9):1884–91.
[51] Institute of Medicine Committee on Health L. Health literacy: a prescription to end confusion. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Washington (DC): National Academies Press (US) Copyright 2004 by the National Academy of Sciences. All rights reserved.; 2004.
[52] Flesch R. A new readability yardstick. J Appl Psychol 1948;32(3):221–33.
[53] Flesch R. How Write Plain Engl 1981.
[54] Lu C, Mu W, Jin YH, Shi YX, Li G, Li Y, et al. Cross-cultural adaptation and psychometric assessment of the statement format Decisional Conflict Scale for Mandarin version. BMC Health v Res 2019;19(1):873.
[55] Pecanac KE, Brown RL, Steingrub J, Anderson W, Matthay MA, White DB. A psychometric study of the decisional conflict scale in surrogate decision makers. Patient Educ Couns 2018;101(11):1957–65.
[56] Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods 2016;48(3):936–49.
[57] Koziol NA, Bovaird JA. The impact of model parameterization and estimation methods on tests of measurement invariance with ordered polytomous data. Educ Psychol Meas 2018;78(2):272–96.
[58] van de Schoot R, Lugtig P, Hox J. A checklist for testing measurement invariance. Eur J Dev Psychol 2012;9(4):486–92.
[59] Jobst LJ, Bader M, Moshagen M. A tutorial on assessing statistical power and determining sample size for structural equation models. Psychol Methods 2023;28(1):207–21.
[60] Paz SH, Liu H, Fongwa MN, Morales LS, Hays RD. Readability estimates for commonly used health-related quality of life surveys. Qual Life Res 2009;18(7):889–900.
[61] Haberman S, Sinharay S, Feinberg RA, Subscores Wainer H. A Practical Guide To Their Production And Consumption. Cambrifge University Press; 2024.
[62] Naye F, Sasseville M, Cachinho C, Tousignant-Laflamme Y, Gerard T, Décary S. Risk factors of clinically significant decisional conflict in people living with chronic pain. Ann Fam Med 2024;22(Supplement 1):6080.
[63] Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods 2000;3(1):4–70.
[64] Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999;49(5):651–61.
[65] Légaré F, Thompson-Leduc P. Twelve myths about shared decision making. Patient Educ Couns 2014;96(3):281–6.
[66] Muscat DM, Cvejic E, Smith J, Thompson R, Chang E, Tracy M, et al. Equity in Choosing Wisely and beyond: the effect of health literacy on healthcare decision-making and methods to support conversations about overuse. BMJ Qual Saf 2024.
[67] Durand MA, Carpenter L, Dolan H, Bravo P, Mann M, Bunn F, et al. Do interventions designed to support shared decision-making reduce health inequalities? A systematic review and meta-analysis. PLoS One 2014;9(4):e94670.
[68] Kim J, Kim S, Hong SW, Kang SW, An M. Validation of the decisional conflict scale for evaluating advance care decision conflict in community-dwelling older adults. Asian Nurs Res 2017;11(4):297–303.
[69] Lam WW, Kwok M, Liao Q, Chan M, Or A, Kwong A, et al. Psychometric assessment of the Chinese version of the decisional conflict scale in Chinese women making decision for breast cancer surgery. Health Expect 2015;18(2):210–20.
[70] Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006;10(4):287–333.