# RESEARCH

Psychometric properties of the Danish SDM-Q-9 questionnaire for shared decisionmaking in patients with pelvic floor disorders and low back pain: item response theory modelling

Mette Hulbaek<sup>1,4\*</sup>, Sofie Ronja Petersen<sup>2</sup> and Charlotte Ibsen<sup>3</sup>

# Abstract

**Background** Worldwide, involving patients in healthcare has become a focus point. Shared decision-making (SDM) is one element of patient involvement and, in many countries, including Denmark, requires culturally adapted and validated questionnaires to measure diverse patient populations' perceptions of this concept. SDM-Q-9, a widely used nine-item generic questionnaire, assesses patients' perception of nine elements during decision-making in consultations. The primary aim of this study is to assess the psychometric performance of the Danish version of the SDM-Q-9 through item response theory (IRT). Additionally, to assess the questionnaire's generic applicability among patients with pelvic floor disorders or low back pain.

**Methods** After treatment decisions, Danish patients with pelvic floor disorders or low back pain rated the level of SDM by completing the SDM-Q-9 questionnaire. litem response theory (the Graded Response Model by Samejima) was applied to assess each item's psychometric performance and the questionnaire's generic applicability (among others discriminative ability, precision and item differential functioning).

**Results** The study invited 825 patients for participation and comprised 758 patients for analysis;73% were women, with a mean age of 52 years and a mean SDM score of 3.87. Discrimination parameters (a-scores) for the model ranged from 2.39 (item 1) to 4.48 (item 8). Analysis of the item-information function curves reflected that item 8 demonstrated the highest maximum, indicating higher precision, while items 1, 2 and 9 showed the lowest maxima. Chi<sup>2</sup>-test statistics showed no significant differential item functioning at the 0.01-significance level for any item between the two patient groups. A ceiling effect was observed as most patients selected the highest score, while a low information load was identified in the SDM's upper load for each item and the overall instrument.

**Conclusions** The Danish SDM-Q-9 demonstrates strong overall performance, with the ability to differentiate between the distinct levels of the underlying construct of SDM. However, the high ceiling effect is a critical limitation.

\*Correspondence: Mette Hulbaek mette.hulbaek@rsyd.dk

Full list of author information is available at the end of the article







While the SDM-Q-9 could serve as a generic questionnaire across samples with varying demographic composition, further exploration of these findings is warranted, particularly across patient samples encompassing more diverse decisions, e.g. patients with life-threatening diseases.

**Keywords** SDM-Q-9, Shared decision making, Psychometrics, Item response theory, Patient involvement, Pelvic floor disorder, Low back pain

# Introduction/background

Shared Decision Making (SDM) has been increasingly advocated as an ideal model for making decisions during a medical encounter, offering different treatment options [1]. SDM is a joint process of sharing information between patients and clinicians to make evidencebased healthcare decisions together [2, 3]. Furthermore, SDM is a conceptual construction based on partnership and incorporating essential elements of person-centered healthcare such as patient values and preferences, options, patient participation, patient education, benefits/risks (pros/cons), and deliberation and negotiation [4]. Despite slow implementation worldwide, SDM has been shown to benefit patients, clinicians and healthcare systems [5]. SDM is indicated if there is more than one reasonable and evidence-based option [2, 3]. The implementation of SDM into healthcare necessitates the development of reliable and practical assessment methods.

Various measurement instruments, including the Shared Decision Making Questionnaire (SDM-Q-9) (Supplementary Material 2), are available to assess SDM in clinical practice [6]. SDM-Q-9, a widely used generic questionnaire with nine items, explores patients' perspectives on essential elements of the decision making process and is designed for broad application across diverse populations [7, 8]. SDM-Q-9 has been translated into numerous languages and validated in various settings and patient populations [7, 9-20]. Numerous of these validation studies using classical factor analysis support the SDM-Q-9 as a questionnaire with a one-factorial construct. Several studies suggest that the best indices of fit would be achieved by excluding item 1 ('My doctor made clear that a decision needs to be made') and, in some settings, item 9 ('My doctor and I reached an agreement on how to proceed') [7-12, 20]. Additional insight into the performance of each of the nine items was gained through some studies that applied item response theory [13–15]. Overall, the studies reported that the SDM-Q-9 demonstrated good fit for a unidimensional latent structure, with items 6 ('My doctor asked me which treatment option I prefer'), 7 ('My doctor and I thoroughly weighed the different treatment options') and 8 ('My doctor and I selected a treatment option together') identified as the most relevant items for SDM. In contrast, item 1 exhibited the lowest loading and scalability indices.

SDM has been the subject of increasing attention in the Danish healthcare system [21, 22]. Several SDM initiatives have been launched in Denmark, where SDM implementation among diverse patient populations has been the focal point. These initiatives have increased the demand in Denmark for culturally adapted and validated instruments to measure SDM [22]. Thus, in 2018, a Danish translation and validation of the SDM-Q-9 (Supplementary Material 1) was published and tested on a group of patients with pelvic floor disorders (PFDs) [23]. This study indicated a one-factorial construct for the Danish version, differentiating with item 1 ('My doctor made clear that a decision needs to be made') compared to the underlying construct [23]. Additional validation with item response theory was suggested to understand these nine items further and to conduct additional validation studies with larger and more diverse samples of the Danish population [23]. Subsequently, the Danish version of the SDM-Q-9 has been used in several studies to assess the level of perceived SDM e.g. in patients with Low Back pain (LBP) [24].

Thus, the primary aim of this study was to assess the psychometric performance of the individual items of the Danish version of the SDM-Q-9 using item response theory (IRT) with patients diagnosed with either a pelvic floor disorder or low back pain. The secondary aim was to assess the Danish questionnaire's generic applicability by investigating the differential item functioning between the two patient groups.

# Methods

### The study sample

The patient sample for this methodological study included data from two groups - PFD and LBP. The PFD group was data from the validation study from 2017 (PFD sample 2017) [23] supplemented with additional, new data from 2022 (PFD sample 2022). The data from the LPB group was from a study in 2018 to promote patient-centred care in patients with LBP (the LBP sample) [24].

# Participants

Patients were recruited from outpatient clinics at five Danish hospitals. The PFD-group was recruited from November 2016 to March 2017 and December 2021 to July 2022. The LBP-group was recruited from November 2017 to August 2018. The PFD-group was recruited from gynecological departments and a pelvic floor unit, whilst the LBP-group was recruited from a Spine Center. All patients had been referred to specialized multidisciplinary consultations (one, two or three consultants with one nurse specialist and often a physiotherapist) by general practitioners.

Inclusion criteria for the PFD-group were (i) referral to a gynecological department or pelvic floor unit with a diagnosis of PFD, e.g. urogenital prolapse, urinary incontinence, fecal incontinence or pelvic floor pain, (ii) age  $\geq$  18, (iii) sufficient knowledge of the Danish language.

Inclusion criteria for the LBP-group were: (i) referral to the Spine Center with a primary diagnosis of LBP with or without leg pain symptoms (sciatica), (ii) aged 18–60 years and (iii) capable of reading and speaking Danish. Exclusion criteria for the LBP-group were patients with neck pain and pain in the upper back.

The PFD-group were presented with the following treatment options: pelvic floor muscle training, bladder training, lifestyle modifications, treatment with electrical tibial nerve stimulation, pharmacological treatment or surgery. The treatment options for the LBP-group were rehabilitation, surgery or a training program.

# The SDM-Q-9

The SDM-Q-9 assesses patients' perception of nine elements of SDM during consultations (Supplemental material 1). Patients rate the nine items on a 6-point Likert scale.

The response categories represent the patients' level of agreement with the statement of the item from 0 (completely disagree) to 5 (completely agree) with a total sum score from 0 to 45 [8]. A higher score indicates that patients have a higher perception of involvement and SDM [25]. The validated Danish SDM-Q-9 was culturally adapted with cognitive interviews (n = 11) and an expert panel consensus before use in the multidisciplinary setting of PFD [26]. Here, the phrase 'the doctor' was replaced with 'one of the team members'. The culturally adapted SDM-Q-9 team version showed high acceptance in a pretest (n = 50) among the patient group (ibid.).

# Procedures

Immediately after consultation, patients from the PFDgroup and the LBP-group were asked to participate.

After consent, the PFD-group were handed a paper version of the SDM-Q-9 questionnaire to answer anonymously and return to the department's mailbox. Patients who did not return the questionnaire were considered non-participants. The reporting of non-participants in the PFD-sample from 2017 was not possible. The LBPgroup received a link to an online version of SDM-Q-9 immediately after their consultation. Data was obtained through SurveyXact<sup>®</sup>. Patients who did not respond received up to three written reminders and one phone call. SDM was not implemented at the time of data collection for the PFD sample 2017 but all clinicians at the pelvic floor unit had received a blended learning course of SDM for the PFD sample 2022. The LBP sample was collected during the ongoing implementation of SDM at the Spine Center.

## Statistical analysis

Descriptive statistics were used to characterize the study sample. For categorical data, descriptive statistics were expressed as absolute and relative frequencies (%) and numerical data were either reported as means with standard deviation or medians with interquartile range, depending on the distribution of the variable.

### **Psychometric properties**

As construct validity had already been examined and unidimensionality of the questionnaire was found investigation of construct validity was not part of this analysis [23].

Samejima's Graded Response Model (GRM) [27] was applied; a parametric model that applies item response theory (IRT). The model selection was based on the Bayesian information criteria, Akaike's information criteria and the sample-adjusted Bayesian information criteria.

The theoretical basis of the IRT-model is that a person's latent trait (in this case SDM), quantified by  $\theta$ , will increase or decrease the probability of giving a particular response to a given item. Conditionally on the latent trait, response probabilities to items are independent of each other, unlike in classical test theory (CTT). There are a number of advantages of IRT compared to CTT, including independency between item scores and scale norms, acknowledgement of differential scaling of items and heteroscedasticity [28]. The GRM that we applied is a polytomous model, since there are 6 response options per item. The GRM is a generalized form of the partial credit model, in that the GRM allows unequal spacing between response options. Its equation is denoted

$$m{P}_{m{i}}\left( heta
ight.)=rac{1}{1+e^{-a}{}_{-}^{m{i}\left( heta
ight.-b_{m{i}m{k}}
ight)}}$$

and can be interpreted as the probability of a person with latent trait  $\theta$  to respond with option k or higher. In the SDM-Q-9, k can be any of the integers {0,1,2,3,4,5}.  $\theta$  is traditionally fixed on a scale between  $\pm 3$  with 0 reflecting the mean trait level. The parameter  $a_i$  denotes the discriminating ability (slope) of an item i, while  $b_{ik}$  describes its requirement parameter (the required amount of  $\theta$  at which the probability to respond  $\geq k$ equals 50%). Furthermore, the GRM calculates the amount of information contained in an item as a function of  $\theta$  ( equation not shown). That information is described by the item information function.

To investigate the psychometric properties within the complete study sample, we extracted the following aspects from the Graded Response Model [29, 30]:

1) discriminative abilities (Are the individual items able to differentiate the distinct degree of the underlying construct of SDM?) using discrimination parameters a. The interpretation of the parameters is based on the recommendation by Baker and Kim [31].

2) categorical response behaviors (Are the responses related to the degree of the underlying construct of SDM?) using item trace plots and requirement parameters b.

3) precision at item and scale levels (Are the individual items consistently informative for the underlying construct of SDM across items?) using item information functions.

4) local independence (Do the individual items measure the underlying construct of SDM in a unidimensional space?) using item residual correlation. No cut-off points were used because (1) it is an arbitrary approach [32] and (2) unidimensionality for the Danish version was confirmed in a previous validation study [23].

Infit and outfit statistics were used to assess model fit to avoid Type-I errors, as the sample size was over 200 [33].

To investigate the generic applicability of SDM-Q-9, potential differential item functioning (DIF) was assessed to determine if the items measured equally between genders and the two sub-samples (the patient groups with LBP and PFDs), controlling for the SDM level using ordinal logistic regression. DIF was considered present if the same items were consistently flagged as statistically significant at the  $\alpha = 0.01$  level by the likelihood-ratio Chi<sup>2</sup>-test [34].

Sample size was estimated according to simulated scenarios showing that a sample size of 500–1000 is sufficient for obtaining accurate parameter estimates for the Graded Response Model [35].

All the statistical analyses were performed in R v4.2.1 using the libraries *mirt* and *lordif* [36].

# **Missing data**

Non-participants and participants with a blank SDM-Q-9 questionnaire (all SDM scores missing) were excluded from the final analysis. For participants with one up to eight missing SDM scores, item mean imputation was carried out as recommended by Dai [37]. Furthermore, we conducted a sensitivity analysis of our model on the complete case population.

# Results

A total of 825 patients were invited and 758 patients were analysed (Fig. 1).

In the PFD-group, 376 filled in the questionnaire with scores for all or some items; and 382 from the LBP-group.



# Table 1 Patient characteristics

	PFD group	LBP group	Total
Patient characteristics			
Invited; n (%)	384 (46.5)	441 (53.5)	825
Non-participants; (% total)	8 (2.0)	57 (12.9)	65 (7.9)
For analysis; n (%)	376 (49.6)	382 (50.4)	758
Gender			
Women; n (% women)	371 (96.6)	235 (53.3)	605 (73.3)
Age			
Age, women; mean, years (SD)	60.0 (17.8)	45.3 (10.4)	51.5 (15.7)
Age, men; mean, years	61.25 (17.7)	46.3 (10.8)	47.2 (11.8)
SDM			
SDM, women; mean score (SD)	4.12 (1.07)	3.43 (1.10)	3.87 (1.13)
SDM, men; mean score, (SD)	3.67 (1.56)	3.46 (1.19)	3.47 (1.21)
Missings, SDM total score; n	19	59	78

The PFD-group consisted of 97% females, ranging in age from 17 to 93 years, with a mean age of 60.0 years. The LBP-group consisted of 53% females, ranging in age from 18 to 60 years, with a mean age of 45.3 years (Table 1). The mean scores for SDM (range 0–5) were relatively high; 3.47 (men) and 3.87 (women) (Table 1).

# Discriminative ability and categorical response behavior

Results from the Graded Response Model for the Danish SDM-Q-9 show the probability for the response categories of the single items in relation to the latent trait, SDM level ( $\theta$ ) on the continuum [-3; 3]. The model reached generally high discrimination parameters (a-scores) ranging from 2.39 (item 1) to 4.48 (item 8) (Table 2). Items 1, 2, 5 and 9 have the lowest a-scores (with a = 2.39; a = 3.33; a = 3.92 and a = 3.30 respectively). The requirement parameters b1-b5 range from -2.14 to 0.46 for  $-3 < \theta < 3$  (Table 2). Requirement parameters (b-scores) have no disordering present since all are in increasing order, reflecting a successive increase in higher category response probability with increasing  $\theta$ .

# Table 2 Item parameters for the SDM-Q9

The item trace lines show that items 1 and 9 have the most overlap in response probability (Fig. 2).

# Precision

The item-information function curves reflect that item 8 had the highest maximum, indicating higher precision, but within a relatively narrow range (Fig. 3). Items 1, 2 and 9 show the lowest maxima for their information functions. Many items show a sudden loss of information at  $\theta \approx 1$ , which indicates a ceiling effect of the responses, as most patients have chosen the highest answer category (5=completely agree). The decreasing information amount is visible throughout the test, with increasing  $\theta$  the test loses information. The lower information load in the SDM range's upper end could be detected for each item and the overall instrument. Therefore, the SDM-Q-9 is more informative in patients/situations with a lower SDM level.

# Local independence

With the exception of a single correlation coefficient, all others are weaker than  $\pm 0.5$  (Table 3). Specifically, items 8 ("My doctor and I selected a treatment option together") and 4 ("My doctor precisely explained the advantages and disadvantages of the treatment options") show a strong negative correlation of -0.52.

# Item fit

Each item had a mean square statistic (i.e. outfit and infit statistic) between 0.67 and 1.30 (Table 4). Items 1 ("My doctor made clear that a decision needs to be made", outfit = 1.30) and 5 ("My doctor helped me understand all the information", outfit = 0.67) exhibit relatively high and low outfit statistics, respectively. However, this number is stabilized in the infit-statistic, which is less sensitive to outliers than the outfit statistic.

Items	Discrimination parameter	Requirement parameters				
	a	b1	b2	b3	b4	b5
1. My doctor made clear that a decision needs to be made	2.39	-2.14	-1.79	-1.35	-0.68	0.27
2. My doctor wanted to know exactly how I want to be involved	3.33	-2.03	-1.72	-1.29	-0.56	0.30
3. My doctor told me that there are different options for treating my medical conditions	4.05	-1.81	-1.42	-1.01	-0.51	0.20
4. My doctor precisely explained the advantages and disadvantages of the treatment options	4.13	-1.69	-1.43	-0.85	-0.31	0.38
5. My doctor helped me understand all the information	3.92	-2.02	-1.69	-1.25	-0.74	0.05
6. My doctor asked me what treatment option I prefer	4.18	-1.72	-1.45	-0.93	-0.39	0.29
7. My doctor and I thoroughly weighed the different treatment options	4.22	-1.63	-1.39	-0.73	-0.22	0.46
8. My doctor and I selected a treatment option together	4.48	-1.63	-1.44	-0.91	-0.47	0.25
9. My doctor and I reached an agreement on how to proceed	3.30	-1.94	-1.65	-1.31	-0.82	-0.11



Fig. 2 Item trace lines for SDM-Q-9. The graphs of P1 to P6 correspond to the six Likert scale options (fully disagree to fully agree)

# Differential item functioning (DIF)

The Chi<sup>2</sup>-test statistics from the ordinal logistic regression indicated DIF at the 0.01-significance level for item 5 between the PFD-group and LBP-group (Fig. 4). The Chi<sup>2</sup>-test statistics showed no DIF at the 0.01-significance level for any other items between the PFD-group and the LBP-group. For gender, no DIF were found, indicating equal item functioning between men and women.

# Sensitivity analysis

The results from the complete case analysis (n = 709) showed no major deviations from the main results,

except for local dependency, which was slightly higher for most items and substantially higher for items 4 and 8 (Supplementary Material 3).

# Discussion

This was the first study to investigate the psychometric properties of the Danish SDM-Q-9 questionnaire using item response theory. Generally, we found that the questionnaire adequately measured the underlying construct of SDM in a representative sample of patients with PFD and LBP. Nevertheless, individual items, particularly items 1, 2, and 9, exhibit poor performance regarding the



Fig. 3 Item information functions for SDM-Q-9

Table 3 Residuals and correlation coefficients for the S	DM-(	Q-9
--	------	-----

	item1	item2	item3	item4	item5	item6	item7	item8	item9
item1	NA	0.26	-0.24	-0.23	-0.21	-0.24	-0.32	-0.28	0.18
item2		NA	-0.27	-0.30	-0.23	-0.25	-0.27	-0.27	-0.20
item3			NA	-0.27	-0.27	-0.19	-0.22	-0.34	-0.28
item4				NA	-0.26	-0.33	0.25	-0.52	-0.28
item5					NA	-0.33	-0.30	-0.32	0.22
item6						NA	0.23	0.24	-0.25
item7							NA	-0.31	-0.30
item8								NA	-0.31
item9									NA

Items	Outfit	z.outfit	Infit	z.infit
item1	1.30	2.14	1.09	1.15
item2	0.97	-0.17	1.06	0.75
item3	0.91	-0.25	0.96	-0.49
item4	0.84	-0.82	0.99	-0.09
item5	0.67	-1.07	0.99	-0.05
item6	0.79	-0.91	1.00	-0.04
item7	0.81	-1.11	0.99	-0.08
item8	0.96	-0.03	1.09	1.15
item9	0.73	-0.74	0.99	-0.04

Table 4 Item fit statistics from the GRM

underlying construct of SDM. Specifically, item 1 performs inadequately, and consideration should be given to removing this item from the instrument. Furthermore, the Danish version of the SDM-Q-9 can be used as a generic tool for measuring SDM in different populations. Our study shows that the validation of the instrument is largely unaffected by the characteristics of the two patient groups and their responses. Thus, it seems that patients' different diagnoses do not significantly influence the functionality of the questionnaire and its items.

The investigation of the individual items showed overall good discrimination abilities indicated by the items' good performance at differentiating the distinct degree of the underlying construct of SDM. In particular, item 8 had a high discrimination ability (4.50), but items 4, 6 and 7 also discriminated well with parameters higher than 4.0. Ballesteros et al. found similar good discrimination abilities for items 6, 7, and 8 by looking at the performance of specific items using Rasch analysis [13]. However, items 4

and 8 showed a strong negative correlation of -0.52, possibly indicating a minor construct in addition to the SDM construct. The item correlation structure was within an acceptable range, although items 4 and 8 were correlated by a coefficient of 0.6, suggesting a departure from unidimensionality and in contrast to findings from Hulbæk et al. [23], De las Cuevas et al. [9] and Kriston et al. [8]. However, we kept the items for further testing, especially given their relatively high information load. Fit statistics were in accordance with the model parameters.

Unidimensionality was found in the previous validation of the Danish version when classical test theory was applied. However, we found a negative correlation between items 4 and 8, which would contradict the assumption of unidimensionalty. This was even more pronounced in the complete case analysis. We reflect that this could be explained by the differing patient roles represented by the two items. In item 4, the patient occupies a more passive position, where the clinician plays the active role (My doctor (The team) precisely explained...). In contrast, item 8 reflects a more active role for the patient, positioning them as a partner in decision-making (My doctor (The team) and I selected ... together). As item 8 had a higher discrimination and high precision, it was assessed to be the most informative of the underlying SDM construct compared to all other items. We like to interpret this as the one item representing the core of patients' perception of SDM. Item 8 expresses the active partnership ("My doctor and I selected a treatment option together"), which influences patients' scores fundamentally, reflecting item performances. Other



Fig. 4 Test Characteristic Curves (TCC) for patient group DIF in the Danish version of the SDM-Q9. The TCCs show the relation between the true score (sum of all items, left and DIF item 5, right) and the trait level

elements in the concept of SDM are outweighed by active partnership since the active partnership is seen to empower patients more than information about options [38] as represented by e.g. item 3. Patients often see themselves as vulnerable in the decision making process, and we believe item 8 encompasses 'team-talking', which can restore patients' autonomy [39]. The contrasting roles between items 4 and 8 may help explain the negative correlation, as they are linked to the same context (testlet dependency). This could also suggest a hidden lack of unidimensionality which was not found earlier. However, there is no existing research to directly support this assumption. Qualitative research might further explore the construct and individual items of the SDM-Q-9 to better understand relationships between items. Finally, since our assumption of unidimensionality was based on a previous Danish study [23], a new validation study using a bi-factor analysis may be necessary to assess the underlying structure of the questionnaire specifically within the population of this study.

Item 1 had the weakest performance with the significantly lowest a-parameter of all items (2.4), with low precision and the lowest ability to differentiate the distinct degree of the underlying construct of SDM. Ballesteros et al. also reported in a sample of patients with multiple sclerosis that this item did not perform well [13]. The Rasch analysis performed by Wu et al. among breast cancer patients showed that item 1 was misfitting in general, leading them to exclude this item from further testing in their analysis [14]. Similarly to Wu et al., we conducted an additional analysis using a more parsimonious version of the questionnaire (28 p. 82), excluding the poorly fitting items 1 and 9. The additional analysis resulted in a more precise description of the underlying trait and improved item response modelling parameters.

In addition, our findings, particularly concerning item 1, underpin previous psychometric investigations of the Danish version of SDM-Q-9 [23]. Factor analysis showed that item 1 had the lowest correlation to the total scale, the lowest correlation coefficient between any two scale items, the highest uniqueness and the lowest factor loading, explaining the lowest percentage of the total variance. Repeated findings of item 1's poor performance despite using different methodological testing [9, 13, 14, 20, 23, 40] lead to speculations upon a more pragmatic use of SDM-Q-9 excluding item 1 in clinical practice. However, for research purposes item 1 should be included to improve comparisons with other international and national research studies. We therefore suggest a validation of the instrument using data from a sample that exclusively responds to items 2-9 as part of a development of the questionnaire to assess whether this approach would improve its efficiency.

The discrimination ability of items is apparent in the b-parameters at the lower end of the theta-space from - 3 to 3. Participants tend to respond with relatively high scores on the Danish version of the questionnaire, resulting in a less nuanced and informative interpretation at the upper end, presenting a ceiling effect. The predominantly negative requirement parameters indicate that a high-perceived level of SDM is relatively easy to achieve. Thus, we found that the SDM-Q-9 has critical limitations in capturing variations in the higher range of SDM, and its optimal utility and discriminative capacities are best observed when assessing lower levels of SDM. As reported in a review by Doherr et al. on the performance of the SDM-Q-9 [41], coupled with the insight from the previous validation study of the Danish SDM-Q-9 [23], a recurrent ceiling effect poses a critical challenge in differentiating and measuring SDM in settings characterized by a pre-existing high level of SDM. This issue is evident in our study, as most patients opted for the highest response category (completely agree), which leads to a relatively high mean SDM score in both subsamples. Furthermore numerous items exhibit a sudden loss of information around  $\theta \approx 1$ .

Kriston et al.'s initial examination with DIF of the SDM-Q-9 as a generic questionnaire revealed inherent differences in item function across different subgroups with different diseases [8]. However, we detected no substantial differences in the analysis of psychometric assessments across subsamples, and we found that the SDM-Q-9 has a similar item functioning in the two patient groups despite demographic differences between the samples, e.g. age and gender. Nejati et al. in their research tested the DIF in subsamples of males and females and they found no differences, too [15]. Thus, we assume that the Danish version can be used as a generic questionnaire across different patient groups as well as across genders.

A key limitation to this study is that we did not collect information on patients' educational levels. Generally, highly educated patients opt for greater involvement in decision making than less educated patients [42-44]. Moreover, highly educated patients tend to have a greater capacity to obtain, read, and understand basic health information and services necessary for making appropriate health decisions [45, 46]. Thus, to properly understand if educational level could have biased our results, these data need to be collected and analyzed in future research, as it is well-known that education correlates with SDM engagement. The overall percentage of women in our sample was notably high (73.3%). The high proportion of women may have influenced the SDM level in our sample, as a previous study reported that women were more likely to engage actively in SDM than men [42]. On the other hand, a study from the Arabian Gulf Region

found no differences in the perception of SDM between male and female patients [47]. However, an Arab context is not directly comparable with a Danish context. The authors suggest that Arab female patients may experience discomfort expressing their preferences and expectations towards SDM due to cultural, social, and religious sensitivities within the physician-patient relationship [47]. A more representative patient sample for the Danish population could be drawn from a Spanish study by Jimenez-Fonseca et al.in which the SDM-Q-9 was applied to male and female oncology patients [48]. In the study, female patients exhibited a significantly different prevalence of dissatisfaction with SDM in the decision-making process compared to their male counterparts [48].

Besides gender, our subsamples differ in age due to data being collected in different study protocols [23, 24]. The age difference poses a challenge for transferability of the mean SDM in general. However, we argue that subsamples, which differ across various demographic variables, could improve assessment of the questionnaire's functional generalizability. This could be achieved by analysing subgroup results and identifying potential differences in how items function across subgroups. We believe that we enhanced the robustness of the study through the diverse subgroups composed of varying gender ratios and ages and facing different decisions and options. Further, we collected data over an extended period and in settings with different phases of SDM implementation.

To strengthen this study, all answers from the two different study protocols (subsample LBP-group and subsample PFD-group) were collected within sufficient timespan to avoid recall bias.

Data from the LBP-group was collected electronically and data from the PFD-group was collected in paper version. This difference in administration modes of the questionnaires have to be considered regarding possibility of introducing response bias to results.

However, we found high response rates in both groups and very few missing values. Further, our DIF analysis showed no significant DIF between the two groups. To support the assumption of no response bias, the two meta analytic studies by Gwaltney et al. and Muehlhausen et al. conclude that subjectively reported outcome measures collected on paper are quantitatively comparable with measures collected electronically [49, 50].

This study presents some important strengths. One siginificant strength is the large sample size (n=758), which is essential for accurately estimating item parameters. However, the most significant strength lies in our methodological approach since it has been argued that IRT offers several advantages over classical test theory [29, 51]. IRT enables the examination of individual items of the instrument, a precise estimation of the relationship between the latent trait and item responses, and a

targeted evaluation of potential problems across various domains. These attributes make it particularly well-suited for assessing the functionality of a questionnaire. Due to IRT's independence from the underlying trait, sample heterogeneity (e.g. different scoring due to disease groups, recruitment periods and clinics), will not affect the reliability and generalizability of results.

# Conclusion

The Danish SDM-Q-9 demonstrates overall strong performance, with the ability to differentiate between the distinct levels of the underlying construct of SDM. It demonstrates utility and discriminative capability in the presence of lower SDM levels but exhibits critical limitations at higher levels of SDM.

Item 1 exhibits weaknesses and the lowest ability for discrimination and precision. Consequently, item 1 might be excluded in the clinical use of SDM-Q-9. However, for research purposes item 1 should be included to allow for comparison with other research studies.

While the SDM-Q-9 could serve as a generic questionnaire across samples with varying demographics, further exploration of these findings is warranted, particularly across diverse patient samples encompassing decisions beyond those among PFD- and LBP patients, e.g., decisions among patients with life-threatening diseases.

#### Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12911-025-03023-6.

Supplementary Material 1 Supplementary Material 2 Supplementary Material 3

#### Acknowledgements

The authors would like to thank all participating patients and healthcare professionals. We would also like to thank the statisticians Therese Koops Grønborg and Andreas Kristian Pedersen for consulting advice and research assistent Caroline M. Moos for her thorough proof reading of the final manuscript.

#### Author contributions

MH and CI made substantial contributions to the conception and design of the work; SRP did substantial statistical analysis, interpretation and drafting of the results section, MH and CI made substantial interpretation of data in the clinical context; MH and CI drafted the work and substantively revised it. All authors (MH, SRP and CI) approved the submitted version and agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

#### Funding

Open access funding provided by University of Southern Denmark MH received funding from the Region of Southern Denmark (J.nr.: 21/17487 Efond: 1003). SRP received no funding. CI received funding by the Health Research Fund of Central Denmark Region; DEFACTUM, Central Denmark Region; Spine Centre of Southern Denmark, Hospital Lillebaelt; Aarhus University; The Danish Health Authority and The Danish Rheumatism Association.

### Data availability

Data is provided within the manuscript or supplementary information files.

### Declarations

### Ethics approval and consent to participate

The study followed the ethical standards of the Helsinki Declaration (World Medical Association's meeting of October 2013). The need for approval was waived by the Regional Committee on Health Research Ethics [file no: S-20162000-145, 20192000-154 and file no. 150/2016]. According to Danish law, approval was not required as no biomedical intervention was performed. The study was approved by the Danish Data Protection Agency [file no. 16/21313 – 2111, 16/35609 – 2263 and file no. 1-16-02-477-16] and the Danish Patient Safety Authority (file no. 3-3013-2513-1). All patients received information about the study beforehand, and consent for participation was obtained.

### **Consent for publication**

Not applicable.

### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Gynecology and Obstetrics, University Hospital of Southern Denmark, Hospital Sønderjylland, Aabenraa, Denmark <sup>2</sup>Department of Clinical Research, University Hospital of Southern Denmark, Aabenraa, Denmark

<sup>3</sup>DEFACTUM, Central Denmark Region, Aarhus, Denmark <sup>4</sup>IRS - Department of Regional Health Research, University of Southern Denmark, Odense, Denmark

# Received: 26 April 2024 / Accepted: 8 May 2025

Published online: 19 May 2025

### References

- Elwyn G. Shared decision making: what is the work? Patient Educ Couns. 2021;104(7):1591–5.
- Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Social science & medicine (1982). 1997;44(5):681–92.
- Coulter AC. Making shared decision-making a reality: no decision about me, without me. The King's Fund; 2011.
- 4. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns. 2006;60(3):301–12.
- Elwyn G, Frosch DL, Kobrin S. Implementing shared decision-making: consider all the consequences. Implement Science: IS. 2016;11:114.
- Scholl I, Koelewijn-van Loon M, Sepucha K, Elwyn G, Legare F, Harter M, et al. Measurement of shared decision making - a review of instruments. Z fur Evidenz Fortbild Und Qualitat Im Gesundheitswesen. 2011;105(4):313–24.
- Simon D, Schorr G, Wirtz M, Vodermaier A, Caspari C, Neuner B, et al. Development and first validation of the shared decision-making questionnaire (SDM-Q). Patient Educ Couns. 2006;63(3):319–27.
- Kriston L, Scholl I, Holzel L, Simon D, Loh A, Harter M. The 9-item shared decision making questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. Patient Educ Couns. 2010;80(1):94–9.
- De las Cuevas C, Perestelo-Perez L, Rivero-Santana A, Cebolla-Marti A, Scholl I, Harter M. Validation of the Spanish version of the 9-item shared Decision-Making questionnaire. Health Expectations: Int J Public Participation Health Care Health Policy. 2015;18(6):2143–53.
- Rodenburg-Vandenbussche S, Pieterse AH, Kroonenberg PM, Scholl I, van der Weijden T, Luyten GP, et al. Dutch translation and psychometric testing of the 9-Item shared decision making questionnaire (SDM-Q-9) and shared decision making questionnaire-Physician version (SDM-Q-Doc) in primary and secondary care. PLoS ONE. 2015;10(7):e0132158.

- Zisman-Ilani Y, Roe D, Scholl I, Harter M, Karnieli-Miller O. Shared decision making during active psychiatric hospitalization: assessment and psychometric properties. Health Commun. 2017;32(1):126–30.
- Alzubaidi H, Hussein A, Mc Namara K, Scholl I. Psychometric properties of the Arabic version of the 9-item shared Decision-Making questionnaire: the entire process from translation to validation. BMJ Open. 2019;9(4):e026672.
- Ballesteros J, Moral E, Brieva L, Ruiz-Beato E, Prefasi D, Maurino J. Psychometric properties of the SDM-Q-9 questionnaire for shared decision-making in multiple sclerosis: item response theory modelling and confirmatory factor analysis. Health Qual Life Outcomes. 2017;15(1):79.
- Wu TY, Chen CT, Huang YJ, Hou WH, Wang JD, Hsieh CL. Rasch analysis of the 9-Item shared decision making questionnaire in women with breast Cancer. Cancer Nurs. 2019;42(3):E34–42.
- Nejati B, Lin CC, Imani V, Browall M, Lin CY, Broström A, et al. Validating patient and physician versions of the shared decision making questionnaire in oncology setting. Health Promot Perspect. 2019;9(2):105–14.
- Rencz F, Tamasi B, Brodszky V, Gulacsi L, Weszl M, Pentek M. Validity and reliability of the 9-item shared decision making questionnaire (SDM-Q-9) in a National survey in Hungary. Eur J Health Economics: HEPAC: Health Econ Prev Care. 2019;20(Suppl 1):43–55.
- Goto Y, Miura H, Son D, Arai H, Kriston L, Scholl I, et al. Psychometric evaluation of the Japanese 9-Item shared decision-Making questionnaire and its association with decision conflict and patient factors in Japanese primary care. Jma J. 2020;3(3):208–15.
- Goto Y, Yamaguchi Y, Onishi J, Arai H, Härter M, Scholl I, et al. Adapting the patient and physician versions of the 9-item shared decision making questionnaire for other healthcare providers in Japan. BMC Med Inf Decis Mak. 2021;21(1):314.
- de Filippis R, Aloi M, Pilieci AM, Boniello F, Quirino D, Steardo L Jr, et al. Psychometric properties of the 9-Item shared Decision-Making questionnaire (SDM-Q-9): validation of the Italian version in a large psychiatric clinical sample. Clin Neuropsychiatry. 2022;19(4):264.
- Rosenlund M, Turja T, Saranto K, Kuusisto H, Jylhä V. Shared decision-making in healthcare: development and assessment of the translated Finnish version of the SDM-Q-9. Scand J Public Health. 2024;14034948241255181.
- Dahl Steffensen K, Hjelholt Baker V, Vinter MM. Implementing shared decision making in Denmark: first steps and future focus areas. Z fur Evidenz Fortbild Und Qualitat Im Gesundheitswesen. 2017;123–124:36–40.
- Steffensen KD, Knudsen BM, Finderup J, Würgler MW, Olling K. Implementation of patient-centred care in Denmark: the way forward with shared decision-making. Z fur Evidenz Fortbild Und Qualitat Im Gesundheitswesen. 2022;171:36–41.
- Hulbaek M, Jorgensen MJ, Mainz H, Birkelund R, Nielsen JB, Debrabant B, et al. Danish translation, cultural adaptation and validation of the shared decision making Questionnaire - Patient version (SDM-Q-9-Pat). Eur J Person Centered Healthc. 2018;6(3):438–46.
- Ibsen C, Maribo T, Nielsen CV, Hørder M, Schiøttz-Christensen B. ICF-Based assessment of functioning in daily clinical practice. A promising direction toward Patient-Centred care in patients with low back pain. Front Rehabil Sci. 2021;2:732594.
- Scholl I, Kriston L, Härter M. PEF-FB-9 Fragebogen Zur partizipativen entscheidungsfindung (revidierte 9-Item-Fassung). Klin Diagnostik U Evaluation. 2011;4:46–9.
- 26. Hulbaek M, Keudel P. Cultural adaption of the Danish SDM-Q-9 for teamconsultations (T-SDM-Q-9-Pat). Abstract Book ISDM. 2022;191:176.
- 27. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monogr Supplement. 1969;34(4):100.
- 28. Polit DF, Yang FM. Measurement and the Measurement of Change: A Primer for the Health Professions. (No title). 2014.
- Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for Patient-reported outcome measurement. Patient. 2014;7(1):23–35.
- Hambleton RK, Swaminathan H. Item response theory: principles and applications. Boston: Kluwer Academic; 1985.
- 31. Baker FB, Kim S-H. The basics of item response theory using R. Springer; 2017. p. 26.
- Edwards MC, Houts CR, Cai L. A diagnostic procedure to detect departures from local independence in item response theory models. Psychol Methods. 2018;23(1):138–49.
- Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. BMC Med Res Methodol. 2008;8(1):33.

- Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. J Stat Softw. 2011;39(8):1–30.
- Jiang S, Wang C, Weiss DJ. Sample size requirements for Estimation of item parameters in the multidimensional graded response model. Front Psychol. 2016;7.
- Team RC. A language and environment for statistical computing R Foundation for Statistical Computing. Vienna, Austria. 2021 [Available from: www.Rproject.org].
- Dai S. Handling missing responses in psychometrics: methods and software. Psych. 2021;3:673–93.
- Joseph-Williams N, Elwyn G, Edwards A. Knowledge is not power for patients: a systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. Patient Educ Couns. 2014;94(3):291–309.
- Elwyn G, Durand MA, Song J, Aarts J, Barr PJ, Berger Z, et al. A three-talk model for shared decision making: multistage consultation process. BMJ (Clinical Res ed). 2017;359:j4891.
- Alvarez K, Wang Y, Alegria M, Ault-Brutus A, Ramanayake N, Yeh YH, et al. Psychometrics of shared decision making and communication as patient centered measures for two Language groups. Psychol Assess. 2016;28(9):1074–86.
- Doherr H, Christalle E, Kriston L, Harter M, Scholl I. Use of the 9-item shared decision making questionnaire (SDM-Q-9 and SDM-Q-Doc) in intervention studies-A systematic review. PLoS ONE. 2017;12(3):e0173904.
- 42. Arora NK, McHorney CA. Patient preferences for medical decision making: who really wants to participate? Med Care. 2000;38(3):335–41.
- Degner LF, Kristjanson LJ, Bowman D, Sloan JA, Carriere KC, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. JAMA. 1997;277(18):1485–92.
- 44. Wallberg B, Michelson H, Nystedt M, Bolund C, Degner LF, Wilking N. Information needs and preferences for participation in treatment decisions

among Swedish breast cancer patients. Acta Oncol (Stockholm Sweden). 2000;39(4):467–76.

- Muscat DM, Gessler D, Ayre J, Norgaard O, Heuck IR, Haar S, et al. Seeking a deeper Understanding of 'distributed health literacy': A systematic review. Health Expectations: Int J Public Participation Health Care Health Policy. 2022;25(3):856–68.
- 46. Jansen T, Rademakers J, Waverijn G, Verheij R, Osborne R, Heijmans M. The role of health literacy in explaining the association between educational attainment and the use of out-of-hours primary care services in chronically ill people: a survey study. BMC Health Serv Res. 2018;18(1):394.
- Alameddine M, Otaki F, Bou-Karroum K, Du Preez L, Loubser P, AlGurg R, et al. Patients' and physicians' gender and perspective on shared decision-making: A cross-sectional study from Dubai. PLoS ONE. 2022;17(9):e0270700.
- 48. Jimenez-Fonseca P, Calderon C, Carmona-Bayonas A, Muñoz MM, Hernández R, Mut Lloret M, et al. The relationship between physician and cancer patient when initiating adjuvant treatment and its association with sociodemographic and clinical variables. Clin Transl Oncol. 2018;20(11):1392–9.
- Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paperand-pencil administration of patient-reported outcome measures: a metaanalytic review. Value Health. 2008;11(2):322–33.
- Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes. 2015;13:167.
- Bortolotti SLV, Tezza R, de Andrade DF, Bornia AC, de Sousa Júnior AF. Relevance and advantages of using the item response theory. Qual Quant. 2013;47(4):2341–6.

# Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.