

# Comparing the nine-item Shared Decision-Making Questionnaire to the OPTION Scale – an attempt to establish convergent validity

Isabelle Scholl PhD Dipl.-Psych\*, Levente Kriston PhD Dipl.-Psych\*, Jörg Dirmaier PhD Dipl.-Psych\* and Martin Härter MD PhD Dipl.-Psych†

\*Research Associate, Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, and

†Professor, Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

## Abstract

### Correspondence

Isabelle Scholl PhD Dipl.-Psych  
Department of Medical Psychology  
University Medical Center  
Hamburg-Eppendorf  
Martinistr. 52  
D-20246 Hamburg  
Germany  
E-mail: i.scholl@uke.de

### Accepted for publication

10 October 2012

**Keywords:** convergent validity, measurement, psychometrics, shared decision-making

**Background** While there has been a clear move towards shared decision-making (SDM) in the last few years, the measurement of SDM-related constructs remains challenging. There has been a call for further psychometric testing of known scales, especially regarding validity aspects.

**Objective** To test convergent validity of the nine-item Shared Decision-Making Questionnaire (SDM-Q-9) by comparing it to the OPTION Scale.

**Design** Cross-sectional study.

**Setting and participants** Data were collected in outpatient care practices. Patients suffering from chronic diseases and facing a medical decision were included in the study.

**Methods** Consultations were evaluated using the OPTION Scale. Patients completed the SDM-Q-9 after the consultation. First, the internal consistency of both scales and the inter-rater reliability of the OPTION Scale were calculated. To analyse the convergent validity of the SDM-Q-9, correlation between the patient (SDM-Q-9) and expert ratings (OPTION Scale) was calculated.

**Results** A total of 21 physicians provided analysable data of consultations with 63 patients. Analyses revealed good internal consistency of the SDM-Q-9 and limited internal consistency of the OPTION Scale. Inter-rater reliability of the latter was less than optimal. Association between the total scores of both instruments was weak with a Spearman correlation of  $r = 0.19$  and did not reach statistical significance.

**Discussion** By the use of the OPTION Scale convergent validity of the SDM-Q-9 could not be established. Several possible explanations for this result are discussed.

**Conclusion** This study shows that the measurement of SDM remains challenging.

## Introduction

In the last few years, there has been a clear move towards increased patient involvement and shared decision-making (SDM) in many countries.<sup>1</sup> Research activity into SDM can be categorized into four different domains<sup>2</sup>: (i) *definition* of SDM and development of *frameworks*, (ii) development and psychometric testing of *measurement scales*, (iii) *development and evaluation of SDM interventions* and (iv) *implementation* in routine practice. Several frameworks have been developed and the concept of SDM has been defined by many authors, which has led to the development of an integrative model of SDM.<sup>3</sup> Furthermore, several measurement scales have been constructed.<sup>4-6</sup> Developments regarding frameworks and measurement scales have allowed researchers to conduct studies that assess the effectiveness of SDM.<sup>7,8</sup> This evidence has supported the development of initial projects aiming at implementing SDM in routine clinical practice, especially in the US,<sup>9</sup> the UK<sup>10</sup> and Canada.<sup>11</sup> When looking at the four domains postulated by O'Connor,<sup>2</sup> it might seem intuitively logical that these research activities have taken place in a chronological order: Using this logic, it can be postulated that the development and psychometric testing of measurement scales has been well researched, especially when looking at the vast amount of literature on the evaluation of interventions (for an overview see<sup>7,8</sup>).

However, when looking closely at published instruments, many problems regarding the assessment of SDM-related constructs remain. Whilst the reliability of scales measuring differing aspects of SDM has been evaluated, validity aspects have not or not extensively been investigated, and several reviews on measurement scales have called for further validation studies to be able to measure SDM constructs both in a *reliable and a valid* manner.<sup>4-6,12</sup> At the same time, even when results regarding validity of the scales are missing, many of these instruments are being used as outcome

measures. This might generate research results biased by poor measurement methods. Furthermore, psychometrically sound scales are needed if SDM is going to be implemented in routine practice.<sup>13</sup> To overcome this problem, further testing of psychometric properties of most scales is necessary, especially to generate results on validity aspects.<sup>5,6</sup> Existing instruments assess different SDM-related constructs. There are scales that assess *decision antecedents* (e.g. preference for participation in decision-making), scales that assess the *decision process* (e.g. clinicians' behaviour in the consultation) and tools that measure *decision outcomes* (e.g. decisional regret, satisfaction with the decision).<sup>5</sup> In recent years, the *decision process* has come into focus.<sup>14</sup> Measures that assess this construct are needed to evaluate studies on SDM (e.g. training programmes). There exist several scales that measure the *process* of SDM, mainly from an external observer's perspective or from the patient's view (for a more detailed overview on measurement scales, see<sup>5</sup>). However, only few of these instruments are available in German and those that are available have shown psychometric problems in German validation studies.<sup>15,16</sup>

One scale, which has been developed in the last few years, is the nine-item Shared Decision-Making Questionnaire (SDM-Q-9).<sup>17</sup> This instrument aims to measure the process of SDM in the medical consultation from the patients' perspective. As a self-report measure, the SDM-Q-9 has the advantage of taking little time to fill-in and thus is feasible to use in research settings, especially compared with the more time-consuming third party rating scales that need recording of consultations. It has been developed in several phases by a methods group on SDM research.<sup>16,18</sup> Its revised nine-item version has been tested in a large German sample. It has shown to be a reliable instrument with a Cronbach's  $\alpha$  of 0.94 and corrected-item-total correlations above 0.7. Furthermore, factorial validity (i.e. the extent to which the scores of a scale are an adequate reflection of the dimensionality of the construct

to be measured<sup>19</sup>), as one part of construct validity (i.e. the extent to which a scale measures the theoretically defined construct,<sup>20</sup>), has been tested and results indicate that the scale is valid regarding the factorial structure.<sup>17</sup> However, like many other scales in the field of SDM (cf.<sup>5,6</sup>), there has not been any further investigation regarding validity aspects other than factorial validity.

To overcome this gap, the aim of this study was to test for convergent validity (i.e. the extent to which a scale correlates with another scale that measures the same construct,<sup>20</sup>) of the 9-item Shared Decision-Making Questionnaire by comparing it to the OPTION Scale,<sup>21</sup> an observer rating scale measuring the same concept, that is, the process of SDM in the medical encounter. We chose the OPTION Scale as a measure to compare the SDM-Q-9 with, as both instruments assess behavioural aspects of the decision-making process and both have been developed based on defined steps of the SDM process.<sup>17,21,22</sup> Furthermore, at the time of planning our study (2007), the OPTION Scale was the only scale measuring the process of SDM that was (i) available in German language and that (ii) had undergone psychometric testing of that German version.<sup>23</sup> Our hypothesis was that if both scales measure the same construct they should correlate substantially ( $r \geq .5$ ).

## Methods

### Design and setting

Data were collected in a cross-sectional study. The study was part of a research programme on patient-centeredness and chronic diseases funded by the German Ministry of Education and Research ([www.forschung-patientenorientierung.de](http://www.forschung-patientenorientierung.de)).

### Participants and recruitment strategy

Due to the focus of the funding programme on chronic diseases (both somatic and mental illnesses), the sample consisted mainly of

outpatient consultations with patients suffering from one of three long-term diseases: type 2 diabetes, chronic back pain and depression. We focused on these diseases because they are all common chronic conditions that are treated both in primary and specialty outpatient care and entail preference-sensitive treatment decisions.

We aimed to recruit 30 physicians and asked each participating physician to collect data from three patients. We gradually informed randomly selected practice-based physicians in Hamburg (Germany) about the study (either by mail or telephone) and invited them to participate. Patients were recruited by the participating physicians. Patients had to meet the following inclusion criteria: (i) a diagnosis of type 2 diabetes, chronic back pain or depression, (ii) above 18 years of age, (iii) German-speaking and (iv) facing a treatment decision regarding one of the three diagnoses named above. Severe cognitive impairment was a criterion for exclusion. Few physicians had difficulties in recruiting enough patients with the above-mentioned diagnoses. To minimize missing data, these physicians were instructed to include patients with other chronic diseases (e.g. hypertension) facing a preference-sensitive treatment decision.

### Data collection

Before the beginning of the data collection in the practices, each physician was visited in his/her practice by a member of the research team (IS) and received a standardized instruction on how to administer the study. Data were collected by the physicians in their practices between August 2009 and September 2010. We asked participating physicians to review their appointment schedule for upcoming consultations, identify all patients meeting the above named eligibility criteria, consecutively inform them about the study and enrol them if they consented (consecutive sample). Patients that consented to participate in the study had to sign a consent form at the beginning of the consultation. During the data collection phase

physicians received regular support over the telephone by the research team. Furthermore, if required, we offered to support the data collection in the practice. After the data collection, participating physicians received financial compensation (150 Euro for the complete study participation).

### Measures

Physicians that agreed to participate completed a short questionnaire on demographic characteristics at the beginning of the study. Each consultation with a patient who agreed to participate was audio-recorded by the physician. After the consultation, patients filled in a questionnaire to assess their view on the SDM process with the SDM-Q-9.<sup>17</sup> The content of the SDM-Q-9 is displayed in Table 1. The questionnaire also contained demographic questions. Physicians were asked to fill in a questionnaire about medical data after the consultation with the patient. The complete questionnaires used in the study are available upon request from the corresponding author. After finishing the data collection in the practices, audio-recorded consultations were transcribed and evaluated with the OPTION Scale<sup>21</sup> by two raters. Similar to the original study on the OPTION Scale, audio recording only was used, for reasons of better practicability under routine clinical conditions. The content of the OPTION Scale is displayed in Table 2. The raters have a background of clinical psychology and medicine, respectively, and were both enrolled in a PhD programme on SDM. Prior to the use of the instrument, the

**Table 1** Item content of SDM-Q-9

1	Making clear that a decision is needed
2	Asking for preferred involvement
3	Informing that different options exist
4	Explaining advantages and disadvantages
5	Helping to understand information
6	Asking for preferred option
7	Weighing options
8	Selecting an option
9	Agreeing on how to proceed

rationers had attended a 2-days workshop on the German version of the OPTION Scale organized by an experienced OPTION Scale rater, who has played a major role in the four-stage translation process of the OPTION Scale into German<sup>21</sup> and who has been extensively trained by the original author. This rater training included calibration of ratings.

### Data analyses

Accounting for study discontinuation (an estimated dropout of 20% of physicians) and missing data (estimated 12.5% of consultations), we aimed for a final sample size of  $N = 63$ , which allows for the detection of correlations above 0.5 with a power of 80% and thus provides a solid basis for the planned analyses.

First, we conducted preparatory analyses to test the reliability of the SDM-Q-9 and the German version of the OPTION Scale in this particular sample, as reliability is a prerequisite for the analysis of validity. We calculated internal consistency (Cronbach's  $\alpha$ , corrected-item-total correlations) for both scales as well as the intraclass correlation coefficient (ICC) for single items and the sum score of the OPTION Scale to test inter-rater reliability.

**Table 2** Item content of the OPTION Scale

1	Drawing attention to an identified problem that requires a decision-making process
2	Disclosing that there is more than one way to deal with the identified problem
3	Assessing the patient's preferred approach to receiving information
4	Listing options
5	Explaining the pros and cons of the options
6	Exploring the patient's expectations
7	Exploring the patient's concerns
8	Checking that the patient's understanding of the information
9	Offering the patient explicit opportunities to ask questions
10	Eliciting the patient's preferred level of involvement in decision-making
11	Indicating the need for a decision-making stage
12	Indicating the need to review the decision

Second, we calculated Spearman's correlation coefficient to test the interrelation of the sum scores of both scales. Third, we performed explorative subgroup analyses regarding demographic and clinical patient characteristics (age, gender, education and medical condition of the patients), as previous results had shown differential item functioning.<sup>18</sup> Furthermore, additional *post hoc* analyses were performed for clarification of results. These included calculating pairwise correlations between items that measure corresponding parts of the decision-making process, calculating correlations between subscales that are built by these items, calculating the correlation between the subscale of items that reflect clinician behaviour (SDM-Q-9 items 1–6) with the OPTION Scale, investigating possible associations between the scales using log-transformed data as well as testing for a nonlinear relationship using second-order polynomials, a detailed examination of the OPTION Scale focussing on item-rater interactions, analysis whether the observed difference between SDM-Q-9 and OPTION scores is associated with patient or physician characteristics, and exploratory multilevel modelling of the data. For all correlations, the mean of the two OPTION Scale ratings was used.

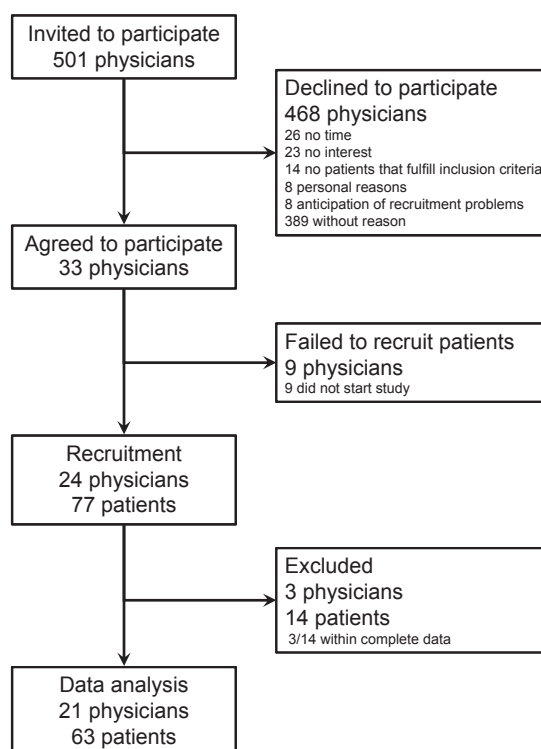
### Ethical considerations

The study was carried out in accordance with the Code of Ethics of the Declaration of Helsinki and was approved by the Ethics Committee of the State Chamber of Physicians in Hamburg (Germany).

## Results

### Sample characteristics

A total of 21 physicians provided complete data sets (SDM-Q-9 filled in and audio-taped consultation regarding a treatment decision) on consultations with 63 patients that could be included in our analyses. A participant flow chart is displayed in Fig. 1. Clinical and



**Figure 1** Participant flow chart.

demographic characteristics of the patients are presented in Table 3.

Approximately two-thirds of the patient sample were women (64%). Patients' mean age was 55 years, ranging from 23 to 93 years. Approximately one-third of the consultations were on type 2 diabetes (38%) or chronic back pain (35%), respectively. A quarter of the consultations were on depression (24%), and the remaining 3% were on other chronic diseases.

Characteristics of the 21 physicians contributing analysable data are reported in Table 4. Twelve of them were general practitioners, four were specialists in diabetology, three were orthopaedists, and two were psychiatrists. Their mean age was 49 years, ranging from 35 to 64 years. Slightly more male than female physicians participated in the study (57%).

Based on available information on sex and specialization from all invited physicians, we compared the observed distribution of these characteristics among the participating physicians with the expected distributions assuming a non-biased random sampling. Results of



**Table 3** Characteristics of the participating patients

	N = 63 <sup>1</sup>	In %
Sex		
Female	40	63.5
Male	23	36.5
Age, years		
Mean (SD, range)	54.8 (15.0, 23–93)	–
18–45 yrs.	18	29.0
45–65 yrs.	26	41.9
>65 yrs.	18	29.0
Education		
Low <sup>2</sup>	34	54.8
Medium <sup>3</sup>	20	32.3
High <sup>4</sup>	8	12.9
Occupation		
Employed	28	46.7
Retired	21	35.0
Homemaker	4	6.7
Student	1	1.7
Unemployed	6	10.0
Family status		
Never married	18	30.0
Married	28	46.7
Divorced	9	15.0
Widowed	5	8.3
Mother tongue		
German	61	100
Health problem in rated consultation		
Type 2 diabetes	24	38.1
Chronic back pain	22	34.9
Depression	15	23.8
Other	1	3.2

<sup>1</sup>Sample size varies between 60 and 63 due to missing values.

<sup>2</sup>Years of education completed  $\leq 9$ .

<sup>3</sup>Years of education completed 10–12.

<sup>4</sup>Years of education completed  $\geq 13$ .

**Table 4** Characteristics of the participating physicians

	N = 21	In %
Sex		
Female	9	42.9
Male	12	57.1
Age, years		
Mean (SD, range)	48.8 (8.4, 35–64)	–
Profession		
General practitioners	12	57.1
Orthopaedists	3	14.3
Psychiatrists	2	9.5
Diabetologists	4	19.1
Years of experience in general/specialist practice		
Mean (SD)	10.2 (8.4)	–
Range	1–31	–

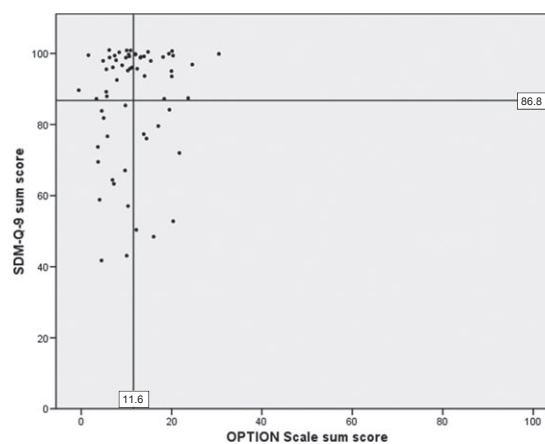
chi-squared tests showed no evidence of any statistically significant difference regarding specialization and sex between invited and participating physicians.

#### Results of preparatory analyses regarding reliability

Reliability analyses regarding the SDM-Q-9 yielded a Cronbach's  $\alpha$  of 0.92 and corrected item-total correlations ranging from 0.52 to 0.85. Regarding the OPTION Scale, analyses of internal consistency showed a Cronbach's  $\alpha$  of 0.68. Only seven of the 12 items showed corrected item-total correlations above the threshold of 0.4. The other five corrected item-total correlations ranged from  $-0.06$  to 0.38. Regarding inter-rater reliability, an ICC of 0.68 was found for the sum score of the OPTION Scale. ICCs of single items varied largely between  $-0.05$  and 0.78.

#### Main findings regarding the correlation of the scales

The mean sum score of the OPTION Scale was 11.6 on a scale ranging from 0 to 100 points (SD 6.3, range 0–32). The mean sum score of the SDM-Q-9 was 86.8 on a scale ranging from 0 to 100 points (SD 16.4, range 42–100). Fig. 2 shows a scatter plot of the sum scores of both instruments.

**Figure 2** Scatter plot of the sum scores of the SDM-Q-9 and the OPTION Scale.

The correlation between the two sum scores was weak with a Spearman's  $r$  of 0.19 and did not reach statistical significance ( $P = 0.069$ ). Subgroup analyses (regarding age, gender, education and medical condition of the patients) revealed no statistically significant correlations in subgroups after adjusting (Bonferroni correction) the level of significance for multiple testing (see Fig. 3).

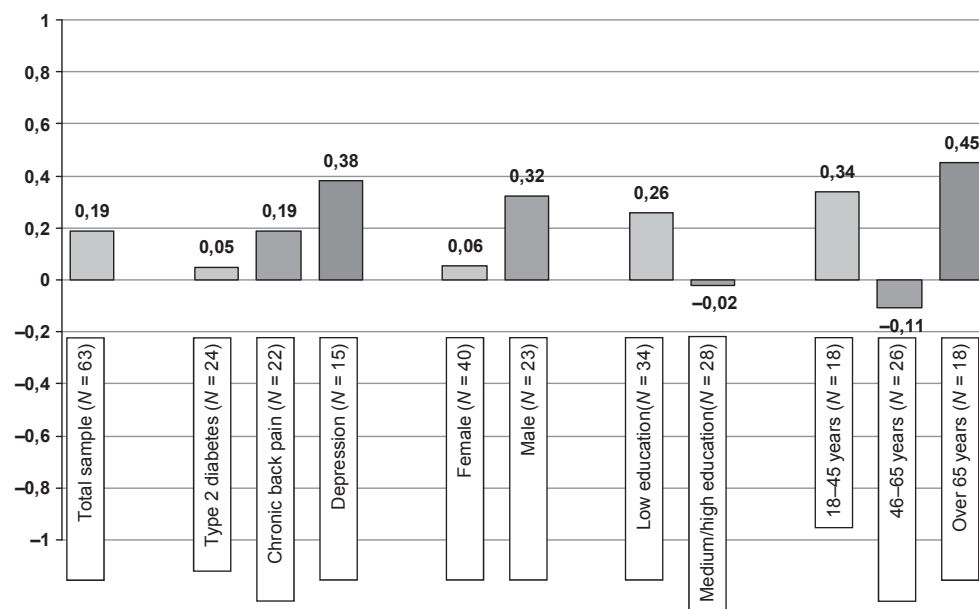
#### Results of additional *post hoc* analyses

Correlations of single items of both scales are displayed in Table 5. Neither the five pairs of items that assess the same part of the decision-making process nor the items that assess different parts of the decision-making process show a statistically significant correlation after adjusting (Bonferroni correction) the level of significance for multiple testing. Furthermore, the sum scores of the respective five items of each scale do not correlate statistically significantly ( $r = 0.14$ ,  $P = 0.141$ ). Similarly, the sum score of the items 1–6 of the SDM-Q-9 (clinician behaviour) does not correlate statistically significantly ( $r = 0.16$ ,  $P = 0.108$ ) with the OPTION Scale.

To test whether a nonlinear relationship between the measures may be present, we repeated the analyses with log-transformed data [ $\log(x)$  for the OPTION score and  $\log(101-x)$  for the SDM-Q-9 score due to corresponding concentration of the values at the higher and the lower end of the scales, respectively; log refers to the natural logarithm] and using first- and second-order polynomials. None of these analyses indicated a statistically significant association between the measures.

In a detailed examination of the ratings on the OPTION Scale using a two-way repeated-measures analysis of variance, we have identified both a statistically significant rater difference (mean ratings of one rater were higher than mean ratings of the other rater;  $P = 0.008$ ), a statistically significant variation among item difficulties ( $P < 0.001$ ), and a statistically significant rater-item interaction ( $P < 0.001$ ).

For further exploration of the findings, we investigated bivariate associations of patient and physician characteristics (as listed in Tables 3 and 4) with the observed difference between SDM-Q-9 and OPTION scores using Spearman correlation and analysis of variance.



**Figure 3** Spearman's correlation coefficients in subgroups.

**Table 5** Correlation of single items of the SDM-Q-9 and OPTION Scale

OPTION Scale	SDM-Q-9								
	Item 1 <sup>a</sup>	Item 2 <sup>c</sup>	Item 3 <sup>b</sup>	Item 4 <sup>c</sup>	Item 5 <sup>d</sup>	Item 6	Item 7	Item 8	Item 9
Item 1 <sup>a</sup>	0.23	0.23	0.03	0.11	-0.03	0.09	0.12	0.10	0.06
Item 2 <sup>b</sup>	-0.10	-0.05	0.09	-0.10	-0.14	0.12	0.09	0.14	0.01
Item 3	0.04	-0.01	0.04	-0.07	0.03	0.05	0.10	0.02	0.04
Item 4	-0.10	-0.10	0.15	0.01	-0.10	0.16	0.18	0.14	-0.02
Item 5 <sup>c</sup>	0.02	-0.02	0.22	0.14	-0.14	0.07	0.14	0.06	0.01
Item 6	-0.06	-0.17	0.14	-0.06	-0.04	0.16	-0.03	-0.02	-0.05
Item 7	0.01	0.09	0.21*	0.12	0.04	0.22	0.27	0.27	0.18
Item 8 <sup>d</sup>	0.13	0.20	-0.01	0.18	-0.03	0.03	0.17	0.09	0.00
Item 9	-0.05	-0.04	0.02	-0.08	-0.18	-0.02	0.06	-0.01	0.05
Item 10 <sup>e</sup>	-0.04	0.08	-0.12	0.02	0.11	0.05	0.18	0.14	0.11
Item 11	0.15	0.16	-0.13	0.04	-0.05	0.05	0.14	0.14	0.03
Item 12	0.10	0.17	0.00	0.07	0.14	0.10	0.32	0.27	0.28

Items with shadowed background are the ones that assess the same part of the decision-making process (see<sup>a-e</sup>)

<sup>a</sup>OPTION Item 1: The clinician draws attention to an identified problem as one that requires a decision-making process. SDM-Q-9 Item 1: My doctor made clear that a decision needs to be made.

<sup>b</sup>OPTION Item 2: The clinician states that there is more than one way to deal with the identified problem ('equipose'). SDM-Q-9 Item 3: My doctor told me that there are different options for treating my medical condition.

<sup>c</sup>OPTION Item 5: The clinician explains the pros and cons of options to the patient (taking 'no action' is an option). SDM-Q-9 Item 4: My doctor precisely explained the advantages and disadvantages of the treatment options.

<sup>d</sup>OPTION Item 8: The clinician checks that the patient has understood the information. SDM-Q-9 Item 5: My doctor helped me understand all the information.

<sup>e</sup>OPTION Item 10: The clinician elicits the patient's preferred level of involvement in decision-making. SDM-Q-9 Item 2: My doctor wanted to know exactly how I want to be involved in making the decision.

None of these tests provided a statistically significant result.

The main analyses were performed without considering the multilevel structure of the data (patients are nested within physicians), because previous empirical findings suggest that physician-level variation in the SDM-Q-9 is very low,<sup>24</sup> the number of patients per physician was rather low ( $63/21 = 3$ ), and frequently used maximum-likelihood estimators can be instable in small samples. We performed an exploratory two-level analysis to identify the amount of variation in the measures that can be attributed to differences between physicians rather than between consultations. In this context, the ICC can be interpreted as per cent ratio of the physician-level and the total variance. ICC was negligible and not statistically significant for the SDM-Q-9 (1.6%;  $P = 0.112$ ) but substantial for the OPTION scale (52.8%;  $P < 0.001$ ). The correlation between the two measures (corrected for the nested data structure) was  $r = 0.068$  ( $P < 0.001$ ), thus being statistically significant but very weak.

## Discussion

In this study, the convergent validity of SDM-Q-9 was tested by comparing it to the OPTION Scale. Only a weak and statistically not significant correlation of 0.19 was found between the two scales in the total sample of 63 consultations. No statistically significant correlations were found in the subgroup analyses. Correlations of single items of the two scales revealed that the items that assess exactly the same part of the decision-making process show no statistically significant correlation. Thus, the hypothesis of a substantial correlation between the two scales has to be rejected. Convergent validity of the SDM-Q-9 could not be established using the OPTION Scale. There are several alternative possible explanations for this result.

First, results might be influenced by response biases of the patients. By giving the instruction to fill in the questionnaires immediately after the consultation and give it in a sealed envelope to a member of the practice team, we tried



to avoid recall bias and reduce missing data. However, this might have increased *social desirability bias*.<sup>25</sup> This could be addressed in future research by installing project managers for recruitment, instead of asking the physicians to enrol patients themselves. Another cognitive process, the striving to *reduce cognitive dissonance*, could also explain the high a SDM-Q-9 scores. Cognitive dissonance is a feeling of discomfort that results from conflicting cognitions (e.g. beliefs, values or perceptions). To overcome that dissonance, people modify their cognitions and create consistency.<sup>26</sup> The results of the OPTION Scale suggest that patients in this study were involved to a low degree in the decision-making process. However, the literature suggests that the majority of patients wish to be involved,<sup>27–30</sup> and thus, the experience of not being involved may have clashed with their preference for involvement. To reduce the resulting cognitive dissonance, patients may have rated the decision-making as shared. Qualitative study designs could give insights into possible cognitive dissonance phenomena. A further confounding factor on patients' ratings may be that their physician asked them to *participate in a research project*. Possibly this special attention that patients received from their physicians biased their rating by making them feel more important and more involved in decisions regarding their treatment.

Moreover, there are two *methodological problems* that might have influenced the results. First, *low variance* due to ceiling effects of the SDM-Q-9 and floor effects of the OPTION Scale may account for the low correlation between the two scales. The OPTION Scale in particular demonstrated very low variance, which is likely to deteriorate measures of association (sometimes referred to as the problem of 'restriction of range'). Similarly, low OPTION scores have been reported by others, both in the original version<sup>31</sup> and in the German version.<sup>32</sup> Future studies could increase variance by testing the scales in clinically more heterogeneous samples. Second, preparatory analyses revealed *problems regarding*

*the reliability of the OPTION Scale*. This leads to an increased measurement error, which could have influenced the results. While variation among raters is frequently present (without influencing the main analysis of association testing) and differences in item-difficulties are psychometrically acceptable, the rater-item interaction suggests implementation problems with the OPTION Scale. Similar methodological problems have been reported in a recent review of studies using this instrument.<sup>33</sup> *Post hoc* exploratory multilevel analyses indicated that around half of the total variance in OPTION ratings is due to differences between physicians rather than between consultations, suggesting that the measured construct has a stable physician-related pattern across consultations.

Furthermore, the cross-sectional design may have led to an *underestimation of the amount of patient involvement in the decision-making process*, as in the treatment of the included chronic conditions any decision can be made in several rather than in a single consultation (cf.<sup>34</sup>). The audio-recorded consultations in this study all included the decision-making process itself, but it might be that some of the 'choice talk' or 'option talk' as described by Elwyn and colleagues<sup>35</sup> occurred in other consultations prior to the one recorded. To this end, the assessment of one single consultation by the use of the OPTION Scale may not reflect clinical practice where the decision-making process may be spread over several consultations. Even if patients were instructed to rate only the last consultation, they may have rated the whole process, resulting in higher scores on the SDM-Q-9.

Another possible explanation is that, experts and patients might have based their rating of the decision-making process on *different reference points*. While experts have an ideal model of SDM (divided into different steps; cf.<sup>36–38</sup>) in their mind while rating the consultation, patients are not familiar with the theoretical concept and might be used to the still predominant paternalistic model of decision-making. Thus, they might overestimate the degree of

involvement in the decision-making process (in the way defined by experts). For example, a physician that briefly states that there is a second possible treatment option might be rated as SDM behaviour by a patient that has rarely experienced such behaviour but rated much lower by an SDM expert rater. There is evidence that patients have a *different understanding of and perspective on SDM* than used in academic environments.<sup>39</sup> This hypothesis of different reference points could be further investigated by providing a description for each item of the SDM-Q-9 with an appropriate anchoring of the response steps. Another difference between external raters and patients that might influence the ratings is connected to *meta-cognitive processes*.<sup>40</sup> While an external rater's sole task is to rate *how* the patient and the physician engage in decision-making, the patient may be more concerned about presenting the problems and discussing the available treatment options, etc. Thus, a patient report scale such as the SDM-Q-9 requires the patient to engage in meta-cognitive activity during the consultation (i.e. processing his/her and the physician's behaviour and asking oneself how the decision-making process is going). This results in competing demands for the patient and might affect the recall of *how* the decision-making process went, which might explain the divergent findings.<sup>41</sup>

Some of these explanations are also discussed in the literature. Recent studies on physician–patient–communication in general and SDM in particular have found *divergent results between patients' and observers' ratings of consultations*. Our result is comparable to the result of Burton and colleagues,<sup>42</sup> who found no correlation between the patients' perceived involvement and the observer assessment with the OPTION Scale. Another current study also showed discrepancies between patients' and external raters' views regarding the degree of involvement in decision-making.<sup>34</sup> Similar results were reported from two studies in the primary care setting.<sup>43,44</sup> One of them<sup>43</sup> discussed that patients' perceptions of involvement in decision-making seem to be influenced

by the general communication skills of the physicians. This inconsistency of observer- and patient-reported assessments of SDM has also been described by Légaré and colleagues,<sup>7</sup> who postulate that no conclusions regarding congruence of assessments can be drawn at the moment. Results of the large-scale DECISIONS study on patients facing common medical decisions indicate that there is no relationship between patients' perceptions of being informed about a certain medical decision and their knowledge scores regarding that medical domain; they considered themselves very well informed, while they performed poorly in answering knowledge questions.<sup>45,46</sup> A similar problem with patient self-reports not corresponding to observer assessments can be seen in studies investigating communication skills on a broader level.<sup>47,48</sup> Discrepancies between self- and observer ratings have also been found in other medical research areas (e.g. pain intensity<sup>49</sup> and depression<sup>50,51</sup>).

These research findings suggest doubt regarding whether observer rating scales like the OPTION Scale are the most adequate tools to test for convergent validity of a self-report scale like the SDM-Q-9. However, 'among the blind, the one-eyed is king', and at the time of planning the study (2007), the German version of the OPTION Scale was to our knowledge the best available tool to use as a comparator. Even if the German version had not undergone extensive psychometric testing, there was at least some data supporting its use.<sup>23,31</sup> Another self-report instrument, the Dyadic OPTION Scale, has recently been published.<sup>52,53</sup> One might argue that this is a better tool for testing convergent validity of the SDM-Q-9, but so far no results regarding its psychometric properties have been published. The main limitation of the study is that the ratings were carried out by different parties. Therefore, it remains unclear whether or not the lack of correlation was because patients have different perspectives than the external raters, or whether the SDM-Q-9 and OPTION Scale are measuring different constructs. However, the fact that both scales have been developed on the basis of the

same model<sup>36</sup> and that other studies have found similar divergence in perspectives,<sup>34,42–44</sup> we hypothesize that the low correlations are rather connected to the different viewpoints than to the fact that the scales assess different constructs. This uncertainty could be overcome in the future by adapting the SDM-Q-9 to a version for third party raters and testing that version against the SDM-Q-9 patient version. The confounding of perspective and construct is likely to be one of the central issues in research on SDM in general.

Furthermore, generalizability of the findings may be limited by a self-selection bias as only a small proportion of the invited physicians consented to participate in the study. These physicians might be more in favour of adopting SDM than other physicians who declined participation. Another limitation is connected to the recruitment of the patients. Although we clearly instructed physicians to consecutively enrol *all* eligible and consenting patients, we cannot completely rule out non-adherence to this instruction. Some physicians may have chosen patients who they thought would best represent SDM or with whom they had a good relationship and who might have given 'better' ratings to their physicians. Furthermore, we treated the physician–patient dyads as equal units. This issue has so far been neglected in research on SDM and large scale trials are needed to investigate whether dyads vary depending on characteristics of both physician and patient (e.g. similarity or dissimilarity regarding gender and age). Large-scale trials would also allow taking the clustered nature of data more fully into account. Another limitation of the study is related to the cross-sectional design that did not allow for the audio-taping of the longitudinal aspect of decision-making process over several consultations. However, it is a general problem in research on SDM that longitudinal studies on the process of SDM are missing. Further work is needed, especially on how to measure this process if it is stretched over several consultations. Moreover, more research is needed regarding different aspects

of validity of most measurement scales.<sup>5</sup> While no psychometrically sound instruments that measure exactly the same construct are available, other scales measuring similar concepts could be used to test convergent validity of the SDM-Q-9 and other scales. Initial psychometric testing of the English version of the SDM-Q-9 have found a positive correlation with satisfaction with decision and a negative correlation with decisional conflict, which may be seen as first indicators of convergent and discriminant validity, respectively.<sup>54</sup> More studies using this kind of Campbell and Fiske approach<sup>55</sup> are needed. However, to truly understand the differences found between observer and patient ratings, both in our and in other studies,<sup>34,42–44</sup> quantitative assessment needs to be underpinned by qualitative studies.<sup>25</sup> For example, this could be achieved by studies using probing to gain deeper understanding of patients' active cognitive processes during completion of the questionnaire.<sup>56</sup> Finally, research on measurement of SDM could benefit from more conceptual and methodological work, for example, by investigating whether the predominantly used reflective measurement model is appropriate for different existing measures.<sup>57</sup>

## Conclusion

While preliminary analyses supported prior results regarding the high internal consistency of the SDM-Q-9,<sup>17</sup> convergent validity could not be established by the use of the OPTION Scale. Possibly, unsatisfactory psychometric properties of the OPTION Scale can partly explain this result. These findings are in line with recent studies on measurement scales in the field of SDM,<sup>5</sup> showing that there is still a lot to be done until we can claim to have psychometrically sound scales for measuring SDM-related concepts. Thus, the second of the four domains (measurement) postulated by O'Connor<sup>2</sup> is not resolved yet. This has implications for the following domains – both for efficacy and effectiveness trials and for implementation projects. Qualitative studies are needed to

examine the difference found depending on which stakeholder's view is assessed.

## Acknowledgements

We would like to thank all patients and physicians who participated in the study and the student research assistants Eva-Maria Müller, Stephanie Pahlke and Sarah Röttger for their work in the project. Furthermore, we would like to thank Ines Weßling for her work as a rater, Dr. Andreas Loh for providing training on the German version of the OPTION Scale and Gemma Cherry for copyediting the manuscript. This project was funded by the German Ministry of Education and Research (project number: 01GX0742).

## Role of funding

The sponsors were not involved in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

## Conflict of interest

None.

## References

- Härter M, van der Weijden T, Elwyn G. Policy and practice developments in the implementation of shared decision making: An international perspective. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2011; **105**: 229–233.
- O'Connor A. Progress and prospects in shared decision making (presentation). 6th International Shared Decision Making Conference. Maastricht: unpublished work; June 2011.
- Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. *Patient Education and Counseling*, 2006; **60**: 301–312.
- Légaré F, Moher D, Elwyn G, LeBlanc A, Gravel K. Instruments to assess the perception of physicians in the decision-making process of specific clinical encounters: A systematic review. *BMC Medical Informatics and Decision Making*, 2007; **7**: 16.
- Scholl I, Koelewijn-van Loon M, Sepucha K *et al.* Measurement of shared decision making – A review of instruments. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2011; **105**: 313–324.
- Simon D, Loh A, Härter M. Measuring (shared) decision-making – A review of psychometric instruments. *Zeitschrift für Ärztliche Fortbildung und Qualitätssicherung*, 2007; **101**: 259–267.
- Légaré F, Ratté S, Stacey D *et al.* Interventions for improving the adoption of shared decision making by healthcare professionals. *Cochrane Database of Systematic Reviews*, 2010; **5**: CD006732.
- O'Connor AM, Stacey D, Entwistle V *et al.* Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, 2009; **3**: CD001431.
- Frosch DL, Moulton BW, Wexler RM, Holmes-Rovner M, Volk RJ, Levin CA. Shared decision making in the United States: Policy and implementation activity on multiple fronts. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2011; **105**: 305–312.
- Coulter A, Edwards A, Elwyn G, Thomson R. Implementing shared decision making in the UK. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2011; **105**: 300–304.
- Légaré F, Stacey D, Forest P-G, Coutu M-F. Moving SDM forward in Canada: Milestones, public involvement, and barriers that remain. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2011; **105**: 245–253.
- Dy SM. Instruments for evaluating shared medical decision making: A structured literature review. *Medical Care Research and Review*, 2007; **64**: 623–649.
- Braddock CH. The emerging importance and relevance of shared decision making to clinical practice. *Medical Decision Making*, 2010; **30**: 5–7.
- Elwyn G, Miron-Shatz T. Deliberation before determination: The definition and evaluation of good decision making. *Health Expectations*, 2009; **13**: 139–147.
- Edwards A, Elwyn G, Hood K *et al.* The development of COMRADE- A patient-based outcome measure to evaluate the effectiveness of risk communication and treatment decision making in consultations. *Patient Education and Counseling*, 2003; **50**: 311–322.
- Giersdorf N, Loh A, Bieber C *et al.* [Development and validation of assessment instruments for shared decision making]. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 2004; **47**: 969–976.
- Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. *Patient Education and Counseling*, 2010; **80**: 94–99.



- 18 Simon D, Schorr G, Wirtz M *et al.* Development and first validation of the shared decision-making questionnaire (SDM-Q). *Patient Education and Counseling*, 2006; **63**: 319–327.
- 19 Mokkink LB, Terwee CB, Patrick DL *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 2010; **63**: 737–745.
- 20 Andresen EM. Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine and Rehabilitation*, 2000; **81**: S15–S20.
- 21 Elwyn G, Edwards A, Wensing M, Grol R. *Shared decision making measurement using the OPTION instrument*. Cardiff: Cardiff University, 2005.
- 22 Elwyn G, Edwards A, Wensing M, Hood K, Atwell C, Grol R. Shared decision making: Developing the OPTION scale for measuring patient involvement. *Quality and Safety in Health Care*, 2003; **12**: 93–99.
- 23 Loh A, Simon D, Hennig K, Hennig B, Härter M, Elwyn G. The assessment of depressive patients' involvement in decision making in audio-taped primary care consultations. *Patient Education and Counseling*, 2006; **63**: 314–318.
- 24 Kriston L, Härter M, Scholl I. A latent variable framework for modeling dyadic measures in research on shared decision-making. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2012; **106**: 253–263.
- 25 Mead N, Bower P. Patient-centredness: A conceptual framework and review of the empirical literature. *Social Science and Medicine*, 2000; **51**: 1087–1110.
- 26 Festinger L. *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press, 1957.
- 27 Bastiaens H, Van Royen P, Pavlic DR, Raposo V, Baker R. Older people's preferences for involvement in their own care: A qualitative study in primary health care in 11 European countries. *Patient Education and Counseling*, 2007; **68**: 33–42.
- 28 Chewning B, Bylund CL, Shah B, Arora NK, Gueguen JA, Makoul G. Patient preferences for shared decisions: A systematic review. *Patient Education and Counseling*, 2012; **86**: 9–18.
- 29 Coulter A, Magee H. *The European patient of the future*. Berkshire: Open University Press, 2003.
- 30 Hamann J, Neuner B, Kasper J *et al.* Participation preferences of patients with acute and chronic conditions. *Health Expectations*, 2007; **10**: 358–363.
- 31 Elwyn G, Hutchings H, Edwards A *et al.* The OPTION scale: Measuring the extent that clinicians involve patients in decision-making tasks. *Health Expectations*, 2005; **8**: 34–42.
- 32 Hirsch O, Keller H, Müller-Engelmann M, Gutenbrunner MH, Krones T, Donner-Banzhoff N. Reliability and validity of the German version of the OPTION scale. *Health Expectations*, 2011. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21521432>, accessed 21 September 2012. DOI 10.1111/j.369-7625.2011.00689.x.
- 33 Nicolai J, Moshagen M, Eich W, Bieber C. The OPTION scale for the assessment of shared decision making (SDM): methodological issues. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2012; **106**: 264–271.
- 34 Behrend L, Maymani H, Diehl M, Gizlice Z, Cai J, Sheridan SL. Patient-physician agreement on the content of CHD prevention discussions. *Health Expectations*, 2010; **14**: 58–72.
- 35 Elwyn G, Frosch D, Thomson R *et al.* Shared decision making: A model for clinical practice. *Journal of General Internal Medicine*, 2012; **27**: 1361–1367.
- 36 Elwyn G, Edwards A, Kinnerley P, Grol R. Shared decision making and the concept of equipoise: The competences of involving patients in healthcare choices. *British Journal of General Practice*, 2000; **50**: 892–897.
- 37 Härter M. [Shared decision making – from the point of view of patients, physicians and health politics is set in place]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2004; **98**: 89–92.
- 38 Whitney SN, Holmes-Rovner M, Brody H *et al.* Beyond shared decision making: An expanded typology of medical decisions. *Medical Decision Making*, 2008; **28**: 699–705.
- 39 Entwistle V, Prior M, Skea ZC, Francis JJ. Involvement in treatment decision-making: Its meaning to people with diabetes and implications for conceptualisation. *Social Science and Medicine*, 2008; **66**: 362–375.
- 40 Hall JA, Murphy NA, Mast MS. Nonverbal self-accuracy in interpersonal interaction. *Personality and Social Psychology Bulletin*, 2007; **33**: 1675–1685.
- 41 Kenny DA, Veldhuijzen W, van der Weijden T *et al.* Interpersonal perception in the context of doctor-patient relationships: A dyadic analysis of doctor-patient communication. *Social Science and Medicine*, 2010; **70**: 763–768.
- 42 Burton D, Blundell N, Jones M, Fraser A, Elwyn G. Shared decision making in cardiology: Do patients want it and do doctors provide it? *Patient Education and Counseling*, 2010; **80**: 173–179.
- 43 Ford S, Schofield T, Hope T. Observing decision-making in the general practice consultation: Who makes which decisions? *Health Expectations*, 2006; **9**: 130–137.
- 44 Saba G, Wong S, Schillinger D *et al.* Shared Decision Making and the experience of partnership in primary care. *Annals of Family Medicine*, 2006; **4**: 54–62.
- 45 Hoffman RM, Lewis CL, Pignone MP *et al.* Decision-making processes for breast, colorectal,

- and prostate cancer screening: the DECISIONS survey. *Medical Decision Making*, 2010; **30**: 53S–64S.
- 46 Sepucha KR, Fagerlin A, Couper MP, Levin CA, Singer E, Zikmund-Fisher BJ. How does feeling informed relate to being informed? The DECISIONS survey. *Medical Decision Making*, 2010; **30**: 77–84.
- 47 Epstein RM, Franks P, Fiscella K *et al.* Measuring patient-centered communication in patient-physician consultations: Theoretical and practical issues. *Social Science and Medicine*, 2005; **61**: 1516–1528.
- 48 McCormack LA, Treiman K, Rupert D *et al.* Measuring patient-centered communication in cancer care: A literature review and the development of a systematic approach. *Social Science and Medicine*, 2011; **72**: 1085–1095.
- 49 Mäntyselkä P, Kumpusalo E, Ahonen R, Takala J. Patients' versus general practitioners' assessments of pain intensity in primary care patients with non-cancer pain. *British Journal of General Practice*, 2001; **51**: 995–997.
- 50 Enns MW, Larsen DK, Cox BJ. Discrepancies between self and observer ratings of depression. The relationship to demographic, clinical and personality variables. *Journal of Affective Disorders*, 2000; **60**: 33–41.
- 51 Steer RE, Beck AT, Riskind JH, Brown G. Relationships between the Beck Depression Inventory and the Hamilton Psychiatric Rating Scale for Depression in depressed outpatients. *Journal of Psychopathology and Behavioral Assessment*, 1987; **9**: 327–339.
- 52 Melbourne E, Roberts S, Durand M-A, Newcombe R, Legare F, Elwyn G. Dyadic OPTION: Measuring perceptions of shared decision-making in practice. *Patient Education and Counseling*, 2011; **83**: 55–57.
- 53 Melbourne E, Sinclair K, Durand M-A, Légaré F, Elwyn G. Developing a dyadic OPTION scale to measure perceptions of shared decision making. *Patient Education and Counseling*, 2010; **78**: 177–183.
- 54 Wills CEW, Glass K, Holloman C *et al.* Validation of the SDM Questionnaire-9 (SDM-Q-9) in a stratified age-proportionate U.S. sample. 6th International Shared Decision Making Conference. Maastricht: unpublished work; June 2011.
- 55 Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959; **56**: 81–105.
- 56 Streiner DL, Norman GR. *Health measurement scales – A practical guide to their development and use*. Oxford: Oxford University Press, 2008.
- 57 Wollschläger D. Short communication: Where is SDM at home? putting theoretical constraints on the way shared decision making is measured. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 2012; **106**: 272–274.