Shared decision making: developing the OPTION scale for measuring patient involvement

G Elwyn, A Edwards, M Wensing, K Hood, C Atwell, R Grol

**Background:** A systematic review has shown that no measures of the extent to which healthcare professionals involve patients in decisions within clinical consultations exist, despite the increasing interest in the benefits or otherwise of patient participation in these decisions.

**Aims:** To describe the development of a new instrument designed to assess the extent to which practitioners involve patients in decision making processes.

**Design:** The OPTION (observing patient involvement) scale was developed and used by two independent raters to assess primary care consultations in order to evaluate its psychometric qualities, validity, and reliability.

**Study sample:** 186 audiotaped consultations collected from the routine clinics of 21 general practitioners in the UK.

**Method:** Item response rates, Cronbach’s alpha, and summed and scaled OPTION scores were calculated. Inter-item and item-total correlations were calculated and inter-rater agreements were calculated using Cohen’s kappa. Classical inter-rater intra-class correlation coefficients and generalisability theory statistics were used to calculate inter-rater reliability coefficients. Basing the tool development on literature reviews, qualitative studies and consultations with practitioner and patients ensured content validity. Construct validity hypothesis testing was conducted by assessing score variation with respect to patient age, clinical topic “equipoise”, sex of practitioner, and success of practitioners at a professional examination.

**Results:** The OPTION scale provided reliable scores for detecting differences between groups of consultations in the extent to which patients are involved in decision making processes in consultations. The results justify the use of the scale in further empirical studies. The inter-rater intraclass correlation coefficient (0.62), kappa scores for inter-rater agreement (0.71), and Cronbach’s alpha (0.79) were all above acceptable thresholds. Based on a balanced design of five consultations per clinician, the inter-rater reliability generalisability coefficient was 0.68 (two raters) and the intra-rater reliability generalisability coefficient was 0.66. On average, mean practitioner scores were very similar (and low on the overall scale of possible involvement); some practitioner scores had more variation around the mean, indicating that they varied their communication styles to a greater extent than others.

**Conclusions:** Involvement in decision making is a key facet of patient participation in health care and the OPTION scale provides a validated outcome measure for future empirical studies.

Patient involvement in shared decision making has been the subject of debate, with some claiming that it should be mandatory while others point out the problems, but it remains an area where few empirical studies have been conducted. A systematic review has shown that there is no existing measure of the extent to which healthcare professionals involve patients in decisions within clinical consultations. Although some instruments include some components of patient involvement, they were found to be insufficiently developed to measure accurately this facet of communication in patient-clinician interactions. The underlying ethical principles of patient autonomy and veracity underpin this development and, coupled with the interested of consumers, professionals and policy makers, drive a research need to ascertain whether achieving greater involvement in decision making is associated with improved patient outcomes.

The area is complex and the concept is not easy to measure. It is reported that, typically, less than 50% of patients wish to be involved in the decision making processes despite the possibility that “involvement” could have a positive effect on health outcomes. Recent qualitative research conducted with a wide range of consumer and patient groups revealed only minor reservations about participation in decision making processes, provided the process was sensitive to individual preferences at any given time points.

Patients stated that professionals should definitely provide information about treatment options, but should respect the extent to which patients wish to take on decision making responsibilities in clinical settings. The underlying principles of the shared decision making method have been described elsewhere and, following a literature review and a series of qualitative and quantitative studies, a skills framework has been proposed. This framework is composed of a set of competences that include the following steps:

- problem definition (and agreement);
- explaining that legitimate choices exist in many clinical situations, a concept defined as professional “equipoise”;
- portraying options and communicating risk about a wide range of issues—for example, entry to screening programmes or the acceptance of investigative procedures or treatment choices); and
- conducting the decision process or its deferment.

These are all aspects of consultations that need to be considered by an instrument designed to assess whether clinicians engage patients in decisions. It is the accomplishment of these competences that forms the conceptual basis for the OPTION scale.
OPTION (acronym for “observing patient involvement”) is
an item-based instrument completed by raters who assess
recordings of consultations (audio or video). It has been
developed to evaluate shared decision making specifically
in the context of general practice, but it is intended to be general
enough for use in all types of consultations in clinical practice.
The OPTION scale is designed to assess the overall shared
decision making process. In summary, it examines whether
problems are well defined, whether options are formulated,
information provided, patient understanding and role prefer-
ence evaluated, and decisions examined from both the profes-
sional and patient perspectives.

Some suggest that clinical practice should be categorised by
a taxonomy of policies—that is, whether the screening,
testing, or treatment under consideration is a “standard”, a
“guideline”, or an “option”—and that clinicians should vary
the degree of patient involvement on this basis. “Standards”
theoretically provide strong evidence of effectiveness and
strong agreement about best treatment. “Guidelines” are less
prescriptive and, where there are “options”, the evidence
regarding effectiveness or otherwise is unclear. It is then pro-
posed that patient involvement be reserved for situations
where clear “options” exist. This scale was designed, however,
from the standpoint that there are opportunities for patients
to be involved in decisions across the spectrum of evidence for
effectiveness or professional agreement about best practice.
Firstly, there are few situations where interventions are free
from harm, and so it is almost always appropriate to raise
awareness about such outcomes. Secondly, patients have
legitimate perspectives on many social and psychological
aspects of decisions whereas the evidence base almost
certainly restricts itself to providing data about the biomedical
aspects of decision making. The instrument developed was
therefore a more generic tool capable of assessing the extent
to which clinicians involve patients in decisions across a range of
situations, excluding emergencies or other compromised
circumstances.

The aim of the study was to enable accurate assessments of
the levels of involvement in shared decision making achieved
within consultations in order to provide research data for
empirical studies in this area. This paper describes the develop-
ment of the instrument and assesses its ability to discrimi-
nate involvement levels and the decision making methods
used in consultations within and between differing practition-
ers by reporting key aspects of the tool’s validity and reliabil-
ity using a sample of consultations recorded in a general prac-
tice setting.

METHODS
The psychometric characteristics of the OPTION scale were
applied to a sample of audiotaped consultations collected from
the routine clinics of 21 GPs and rated by two observers. Valid-
ity issues were considered at both theoretical (construct
emergence) and item formulation and design stages; con-
struct validity was also investigated. The reliability of the scale
was calculated by assessing response rates, inter-item and
item-total correlations, inter-rater agreement (kappa), and
inter- and intra-rater reliability coefficients using both classi-
cal and generalisability theory statistical methods.

Approval to conduct the work was obtained from the Gwent
local research ethics committee.

Overall design features
The content validity of the instrument was developed by
appraising existing research and undertaking qualitative
studies to define the clinical competences of patient involve-
ment in shared decision making in clinical consultations.1

Content validity and concept mapping
The development process followed established guidelines.2
The systematic review2 allowed existing scales—especially
measures of related concepts such as “patient centredness”
and “informed decision making”3—27—to be considered criti-
cally. Qualitative studies using key informants to clarify and
expand the competences revealed that clinicians have specific
perceptions about what constitutes “involvement in decision
making” which are matched in part, but not entirely, by
patient views23 and emphasised the importance of checking
patient role preference (item 10, table 2). The use of design
and piloting iterations involving both patient and clinician
groups ensured content validity and formulated items. In
addition, a sample of consultations in which clinicians were
intent and experienced at involving patients in discussions and
sharing decisions were purposively chosen and examined.23
Thus, the theoretical construct was refined by an assessment
of clinical practice.22 The synthesis of this body of work
enabled the development of a theoretical framework for
patient involvement in decision making and informed the
design of the OPTION instrument.

Instrument and scale development
An 18-item pilot instrument was used by five GP key
informants2 and one non-clinical rater to assess six simulated
audiotaped consultations; item refinement and scale develop-
ment involved three iterative cycles over a 12 month period.
These simulated consultations had been modelled to contain
differing levels of patient involvement and decision making
methods. This process reduced item ambiguity, removed value
laden wordings, and resulted in short and (where possible)
padronised measures.26 A 5-point scale, anchored at both
ends with the words “strongly agree” and “strongly disagree”,
was used to avoid the loss of scoring efficiency in dichot-
omised measures.28 Revisions included removing two duplica-
tive items, increasing the focus on observable “clinician
behaviour” rather than attempting to assess patient percep-
tions of the consultation, and modifying item sequence.

This version was subjected to further piloting using a
second calibration audiotape containing modelled consulta-
tions (two “paternalistic” consultations, three “shared deci-
sion making” and two “informed choice” examples). These
consultations were rated by two non-clinical raters using the
OPTION scale and two other scales—namely, the determina-
tion of “common ground” developed by Stewart et al in
Ontario32 and Braddock’s measure of “informed decision
making”22—which were selected as the most comparable
scales identified.1 The raters provided written feedback and
regarded the pilot 16-item OPTION instrument as a more
acceptable and feasible tool. For the assessment of the
simulated tapes the OPTION scale achieved an inter-rater reli-
bility correlation coefficient of 0.96 compared with a score of
0.76 for the Braddock scale and 0.4 for the Stewart “common
ground” scale. These initial results were therefore promising
and a stable version of the instrument (June 2000) was
described in a manual for raters. By participating in item revi-
sion and the development of the manual drafting, the raters
were integrated into a calibration process before applying the
instrument to a series of naturally occurring consultations.

Data collection: practitioner and patient samples
To test the instrument, recordings of consultations were taken
from the recruitment phase of a proposed trial of shared deci-
sion making and risk communication.24 As part of the recruit-
ment process to the study, GPs in Gwent, South Wales were
asked to audiotape consecutive consultations during a routine
consulting session in general practice. To be eligible for possi-
ble recruitment into the trial the GPs had to have been princi-
pals in a general practice for at least 1 year and less than 10
years. The potential sample pool of 104 GPs in 49 practices
(mean age 41 years, 62% men) was initially approached by
letter (followed by telephone contact) and asked to participate
in a research trial. As far as we are aware, these volunteer
practitioners were naïve to the concepts that we were measuring and had not been exposed to any training or educational interventions that could have influenced their proficiency in this area. Patients attending on the specified recording dates gave their consent using standard procedures, and their age and sex were recorded. Apart from these consent procedures, no other stipulations were imposed and the data collected contained recordings covering the range of conditions typically seen in routine general practice sessions.

Each consultation recording (Spring 2000) was rated in the autumn using the OPTION instrument by two calibrated raters who were non-clinical academics in social sciences and who remained independent of the main research team. Tapes are available for re-assessment. A random sample of 21 consultations (one per clinician) was selected for test-retest analysis and repeated ratings conducted by the two raters.

Data analysis
The data were analysed by taking the response to each item and calculating a summed OPTION score which was then scaled to lie between 0 (least involved) and 100 (most involved). Inter-item and item-total correlations were calculated and inter-rater agreements were calculated using Cohen’s kappa. As well as assessing a classical inter-rater intraclass correlation coefficient, the inter-rater and intra-rater reliability coefficients of the instrument were calculated using the statistical techniques described in generalisability theory.26 This theory uses modified analysis of variance techniques to generate “generalisability coefficients”.26 The methods enable multiple sources of error variance to be calculated and subsequent generalisations to be made about the degree to which these sources are contributing to the overall variability. This allows decisions to be made about the effect of changing the characteristics of the measurement process—for example, number of raters or number of consultations per practitioner26—in order to assess the instrument’s reliability. We also estimated whether consultation scores clustered within practitioners by calculating an intracluster correlation. We also estimated whether consultation scores clustered within practitioners by calculating an intracluster correlation coefficient and calculating a summed OPTION score which was then scaled to lie between 0 (least involved) and 100 (most involved). The data were analysed by taking the response to each item and calculating a summed OPTION score which was then scaled to lie between 0 (least involved) and 100 (most involved). Inter-item and item-total correlations were calculated and inter-rater agreements were calculated using Cohen’s kappa. As well as assessing a classical inter-rater intraclass correlation coefficient, the inter-rater and intra-rater reliability coefficients of the instrument were calculated using the statistical techniques described in generalisability theory.26 This theory uses modified analysis of variance techniques to generate “generalisability coefficients”.26 The methods enable multiple sources of error variance to be calculated and subsequent generalisations to be made about the degree to which these sources are contributing to the overall variability. This allows decisions to be made about the effect of changing the characteristics of the measurement process—for example, number of raters or number of consultations per practitioner26—in order to assess the instrument’s reliability.

Assessment of the construct validity of the OPTION instrument was conducted by examining four hypothetical constructs—namely, that the OPTION score level would be influenced by patient age (negative), sex of clinician (positive in favour of female), qualification of clinician (positive), and whether the clinical topic was one where clinical equipoise existed (positive). The existence of equipoise was determined by a clinical assessment of the audiotape sample content (GE). Studies have also examined the effect of sex of the physician on communication within consultations. Although an area of debate,27 Hall et al27 found that female physicians made more partnership statements than male physicians and Coates’ review29 reported a broad consensus that female language is generally more cooperative. Although there is no consistent evidence, we examined this by comparing the mean OPTION scores for the eight female clinicians with those of their 13 male colleagues (t test). In 1995 the examination for membership of the Royal College of General Practitioners, UK (MRCGP) introduced a video assessment and listed shared decision making as a merit criterion. Although there exists evidence that GPs in training do not involve patients in decision making,27 it was conjectured that success in the examination (at any time, before 1995, or after 1995) might be associated with higher scores (t test), although we did not expect strong correlations. It has been established in cross sectional studies that increasing patient age leads to less patient preference for involvement,28,29 and we assessed the correlation (Pearson) between OPTION scores and patient age. It was also hypothesised from previous qualitative work that decisions were more likely to be shared in consultations that contained clinical problems characterised by professional equipoise such as hormone replacement therapy.30 The consultations were differentiated (by GE) according to this characteristic and any significant differences between the mean OPTION scores were determined (weighted t test). No attempt was made to establish criterion (specifically concurrent) validity.

Results
Sample characteristics
Of the potential sample pool of 104 practitioners, 21 GPs in separate practices who showed interest in being recruited into the trial provided a tape of a routine clinic before receiving any detailed information about the proposed research. These GPs represented a slightly younger group than the sampling frame (mean age 38 years), identical M:F ratio (38% female), and 16 (76%) had been successful in the membership examination of the Royal College of General Practitioners compared with an overall membership level of 54% in the sampling frame. Of the 242 consecutive patients approached in all practices, 12 (5%) declined to have the consultation recorded (the maximum refusal in any one practice was three patients in a series of 15). The remaining 230 consultations were assessed and, after removing consultations where there were technical recording problems, 186 consultations were available for analysis (average of 8.8 consultations per practitioner). There was no age and sex difference between the consultations excluded because of poor recordings and those included for analysis. One practitioner recorded five consultations but most recorded eight or more. There were twice as many consultations with women in the sample and 66% of the patients seen were aged between 30 and 70 years. The demographic and clinical characteristics of the recorded consultations are summarised in table 1.

Scale refinement
The performance of the 16-item scale was analysed in detail. Four of the items had been formulated to try and discriminate between styles of clinician decision methods to distinguish between paternalism, on the one hand, and the transfer of decisional responsibility to the patient on the other. The other 12 items had been constructed to determine performance within a construct of a defined set of steps and skills. The reliability of items that attempted to differentiate between decision making styles was poor, and a decision was made to

| Table 1 Demographic and clinical characteristics of the recorded consultations (n=186) |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Male/female                    | 60 (32%)        | 126 (68%)       |                 |                 |                 |                 |                 |
| Age (years)                    | Mean 43.3, SD 20.6, range 4 months–83 years |                 |                 |                 |                 |                 |
| Duration of consultation (min) | Mean 8.2, SD 4.0, median 7.3, range 22.5 |                 |                 |                 |                 |                 |
| Clinical problems              |                 |                 |                 |                 |                 |                 |
| Respiratory                    | 28              |                 |                 |                 |                 |                 |
| Musculoskeletal                | 27              |                 |                 |                 |                 |                 |
| Dermatological                 | 21              |                 |                 |                 |                 |                 |
| Psychological                  | 13              |                 |                 |                 |                 |                 |
| Cardiovascular                 | 12              |                 |                 |                 |                 |                 |
| Hypertension                   | 11              |                 |                 |                 |                 |                 |
| HT                              | 11              |                 |                 |                 |                 |                 |
| Other                           | 63              |                 |                 |                 |                 |                 |
focus on a scale that was composed of the items that specifically evaluated the agreed competence framework. It is the reliability and construct validity of this 12-item scale that is reported.

**Response rates to OPTION items**

Items 1, 2, 3, 4, and 6 had a range of responses across the 5-point scale but with a predominance of low scores (see table 2 for summary of responses to items). Oversights in item completion led to an average of 0.9% missing values that were distributed evenly across all items (see table 2). The results indicate that the clinicians generally did not portray equipoise (71% strongly disagree); they did not usually list options (71.8% strongly disagree); they did not often explain the pros and cons of options (71.5% strongly disagree); and they did not explore patients’ expectations about how the problem(s) were to be managed (69.9% strongly disagree). Responses to items 7, 8, and 9 revealed most variation across scale points. Item 7 asked whether the clinician explored the patients’ concerns (fears) about how the problem(s) are to be managed. The clinician checks that the patient has understood the information (0.8 strongly disagree) and the clinician provides opportunities for the patient to ask questions (1.9 strongly disagree) but an arrangement to review the decision (or the deferment) was rarely observed (19.4 strongly disagree). The clinician explores the patients’ expectations (or ideas) about how the problem(s) are to be managed (0.5 strongly disagree).

**Reliability of the OPTION score (summed and scaled scores)**

For all 12 items the mean Cohen kappa score was 0.66, indicating acceptable inter-rater agreement for this type of instrument after correcting for chance. Exclusion of item 9 (which requires further attention because of its low kappa score) increased the mean kappa score to 0.71. For the kappa scores the scale was aggregated to three points (agree, neutral, disagree; see table 2). Five point kappa scores are shown in parentheses. Coefficient (Cronbach’s α) was 0.79, indicating little redundancy in the scale (using the mean of the two rater scores). The inter-rater intraclass correlation coefficient for the OPTION score was 0.62. Based on a balanced design of the first five consultations on each practitioner’s audiotape, the inter-rater reliability generalisability coefficient was 0.68 (two raters) and, using the test-retest data, the intra-rater reliability generalisability coefficient was 0.66. The corrected item-total correlations lay between 0.35 and 0.66 except for items 1 and 5 which had correlations of 0.05 and 0.07, respectively. Kaiser-Meyer-Olkin measure of sampling adequacy was 0.82, indicating a very compact pattern of item correlation and justifying the use of factor analysis. Confirmatory factor analysis using principal components revealed that variable loading scores in a forced single factor solution resulted in scores that were above 0.36 (the recommended thresholds for sample sizes of approximately 200) for all except items 1 and 5 (–0.10 and 0.09). Item 1 asked whether a “problem” is identified by the clinician and perhaps should be regarded as a gateway item to the scale—that is, if a problem is not identified then it is difficult to see how the other items can be scored effectively. Item 5 had a low endorsement rate which was anticipated given current practice. Items 2–4 and 6–12 had a mean factor loading of 0.64. A total of 35.2% of the variance was explained by one latent component. Of a total of 66 possible inter-item correlations, 49 were above 0.25 (mean r = 0.40).

Given these reliability indicators, the overall mean (SD) OPTION score for all clinicians on a scale of 0–100, averaged across both rater scores, was 16.9 (7.7), 95% confidence interval 15.8 to 18.0, with a minimum score of 3.3 and a maximum of 44.2 across the sample. The scores are skewed towards low values (see fig 1). At the individual clinician level the mean OPTION scores lay between 8.8 and 23.8 with an intracluster correlation coefficient of 0.22 (across individual means), indicating significant clustering of consultation scores within clinicians. These scores and the quartiles for each practitioner are shown in fig 2. Note that some clinicians have a much wider range of involvement score, indicating a more variable consulting style. The results show that the general level of patient involvement achieved in these consultations was low.

### Table 2 Option item response, missing value rates (%), and Cohen’s kappa

<table>
<thead>
<tr>
<th>OPTION scale item</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
<th>Missing values (%)</th>
<th>Kappa score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) The clinician identifies a problem(s) needing a decision making process</td>
<td>49.5</td>
<td>33.1</td>
<td>11.0</td>
<td>4.3</td>
<td>1.3</td>
<td>0.8</td>
<td>0.61 (0.31)</td>
</tr>
<tr>
<td>(2) The clinician states that there is more than one way to deal with an identified problem (“ equipoise”)</td>
<td>6.2</td>
<td>3.2</td>
<td>5.4</td>
<td>13.4</td>
<td>71.0</td>
<td>0.8</td>
<td>0.82 (0.50)</td>
</tr>
<tr>
<td>(3) The clinician lists “options” including the choice of “no action” if feasible</td>
<td>6.7</td>
<td>4.0</td>
<td>7.0</td>
<td>9.7</td>
<td>71.8</td>
<td>0.8</td>
<td>0.75 (0.51)</td>
</tr>
<tr>
<td>(4) The clinician explains the pros and cons of options to the patient taking “no action” is an option</td>
<td>3.5</td>
<td>3.2</td>
<td>9.4</td>
<td>11.6</td>
<td>71.5</td>
<td>0.8</td>
<td>0.68 (0.43)</td>
</tr>
<tr>
<td>(5) The clinician checks the patient’s preferred information format (words/numbers/visual display)</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.5</td>
<td>98.4</td>
<td>0.8</td>
<td>0.98 (0.98)</td>
</tr>
<tr>
<td>(6) The clinician explores the patient’s expectations (or ideas) about how the problem(s) are to be managed</td>
<td>0.5</td>
<td>1.1</td>
<td>8.6</td>
<td>18.8</td>
<td>69.9</td>
<td>1.1</td>
<td>0.75 (0.34)</td>
</tr>
<tr>
<td>(7) The clinician explores the patient’s concerns (fears) about how the problem(s) are to be managed</td>
<td>1.3</td>
<td>4.6</td>
<td>12.1</td>
<td>22.0</td>
<td>59.1</td>
<td>0.8</td>
<td>0.53 (0.42)</td>
</tr>
<tr>
<td>(8) The clinician checks that the patient has understood the information</td>
<td>0.8</td>
<td>1.1</td>
<td>35.2</td>
<td>26.9</td>
<td>34.9</td>
<td>1.1</td>
<td>0.38 (0.10)</td>
</tr>
<tr>
<td>(9) The clinician provides opportunities for the patient to ask questions</td>
<td>1.9</td>
<td>3.2</td>
<td>40.1</td>
<td>17.2</td>
<td>36.0</td>
<td>1.6</td>
<td>0.20 (-0.08)</td>
</tr>
<tr>
<td>(10) The clinician asks for the patient’s preferred level of involvement in decision making</td>
<td>0.8</td>
<td>1.3</td>
<td>4.0</td>
<td>8.1</td>
<td>84.9</td>
<td>0.8</td>
<td>0.86 (0.66)</td>
</tr>
<tr>
<td>(11) An opportunity for deferring a decision is provided</td>
<td>1.1</td>
<td>2.4</td>
<td>4.8</td>
<td>7.5</td>
<td>83.3</td>
<td>0.8</td>
<td>0.83 (0.66)</td>
</tr>
<tr>
<td>(12) Arrangements are made to review the decision (or the deferment)</td>
<td>19.4</td>
<td>7.8</td>
<td>35.2</td>
<td>5.4</td>
<td>30.9</td>
<td>0.8</td>
<td>0.58 (0.44)</td>
</tr>
</tbody>
</table>

*Kappa scores are for agreement across sum of “agree, neutral and disagree” scale points; scores in parentheses are kappa scores for 5-point scale agreement.
between practitioners, there is considerable variability within practitioners, as shown by the differing quartile ranges around their mean scores (fig 2). Some clinicians have a narrower range of scores than others. This may indicate that these clinicians are able to modify their involvement levels across different consultations and to adapt it to the preferred roles of patients in these interactions. This is, however, a conjecture that needs further investigation.

The content validity of the instrument was based on formulating the items from the existing literature, using the results of a series of studies designed to understand how patient involvement can best be achieved in professional practice, followed by subsequent development using an iterative design and assessment cycle. The results with the instrument in this sample of consultations indicate that low levels of involvement in shared decision making are achieved by GPs and that paternalism is the typical “modus operandi” in routine consultations. These practitioners volunteered to take part in a research study on communication skills, and represent those with a high level of confidence in their skills who were aware that we were recording their consultations. Results from other practitioners are likely to be at least on a par or, most likely, lower.

The results indicate that the OPTION instrument achieves acceptable levels of measurement reliability for use in research settings. By focusing on a specific dimension this scale seems to have acceptable levels of reliability compared with similar measures.‡§† Construct validity was supported by a correlation between involvement scores and patient age and the existence of clinical equipoise in the consultation (although the sample was limited); both hypotheses are supported by previous findings. The lack of correlation between involvement scores and sex of the practitioner or success at the sample was therefore required to agree an index problem. Guidance is given for this issue in a revised manual. In summary, the problem is chosen for which the prime attention is given during the consultation or for which the clinician achieves the greatest involvement score, as the aim is to score demonstrated ability not to calculate involvement across all possible decisions. Secondly, parent and child consultations required additional guidelines (advising that the interaction between the clinician and the adult was assessed), and the raters had to judge which was the main patient participant where teenagers were being consulted. It was not possible to estimate concurrent validity (correlation of the measure with some other scale of the concept or trait to be assessed) as there was neither a “gold standard” nor a comparable instrument available. Correlation with patient opinions about their preferred and achieved involvement levels will be reported in further studies from trials conducted in parallel with this validation study.

Psychometric assessment also revealed areas where further instrument refinement is necessary. Item 1 may need to be conceptualised as a “gateway” item in which the assessment of involvement in decision making cannot continue if no agreed problem can be identified. Although item 5 has a relatively high kappa score, the response rate was skewed and the factor loading is low. The item is retained, however, as it asks about a feature (use of risk communications tools) that is known not to occur in current service settings. As interventions to change this situation are being introduced, however,

**DISCUSSION**

**Principal findings**

The results of this study show that the OPTION scale provides a method of scoring the extent to which clinicians involve patients in the decision making process at the consultation level. Based on the psychometric characteristics reported, we were satisfied that the scale could be used to provide a score for the competence framework we had defined as “shared decision making”. Although there is little overall variance

![Figure 1](https://example.com/figure1.png) **Distribution of OPTION scores.**

![Figure 2](https://example.com/figure2.png) **Mean OPTION scores for clinicians (box plots).**

**Construct validity**

Two constructs were found to be correlated with levels of involvement in decision making—namely, patient age and the existence of a clinical topic where professional equipoise could be expected. The correlation coefficient between the mean OPTION score and patient age (adult age range) was -0.144 (p<0.01) and confirmed the hypothesis that involvement levels reduced as patient age increased. Although this was a small sample, it was found that consultations that contained clinical problems characterised by having a greater likelihood of professionals exhibiting equipoise about treatment choice (n=15 consultations, 8.1%), such as discussion of HRT or depression, had a mean OPTION score of 21.6 which was significantly higher than the mean scores achieved in consultations where equipoise topics did not occur (16.4, p<0.01, weighted t test). Confirming the hypothesis that involvement increases where this characteristic exists. Sex of the clinician and success otherwise in the MRCGP examination were not associated with differences in OPTION scores.

![www.qshc.com](https://example.com)
the results are likely to change with time as decision aids are introduced into clinical settings. Item 9 questions whether clinicians “provide opportunities for the patient to ask questions” but it has low kappa scores and a factor loading score below 0.2. This item needs modification and further testing to overcome the variation in scoring judgement. There is also a need to consider changing the scale from one that measures magnitude rather than attitude.

Implications for research and formative skill development

OPTION scores for these routine consultations taken from general practice in a UK setting are low. For some items almost no responses were registered—for example, there was 99.7% disagreement with item 5 which asked if the clinician “checks the patient’s preferred information format”. Further research work in this area will involve presenting information in different formats and it is known that, when practitioners develop the skills of involving patients, there is a tendency for a pendulum effect. Retaining these items and others that reveal skewed or “floor” scores should enhance the ability of the instrument to register change.

The OPTION scale can therefore be used to determine the extent to which clinicians involve patients in clinical decisions. It should be noted that the results show that some practitioners have a wider scatter of scores than others. This result is congruent with the theoretical stance that practitioners should be flexible in their consulting style and adapt to the nature of the problem and the patient preference for participation in clinical decisions, although we cannot be certain that this has occurred. It is noteworthy, however, that these OPTION scores are low and it is anticipated that higher scores will be evident after periods of skill development. The instrument should be used to determine scores at a group level (mean scores) or at consultation levels and not to provide a definitive OPTION score that is taken to be characteristic of that practitioner’s ability, unless attention is given to case mix, sample size, and confidence interval estimation. The responsiveness of the instrument to change (increased levels of patient involvement in decision making after skill development) will be validated in further evaluations. It should be emphasised that this tool is designed as an evaluation of a consultation process. It does not measure patient’s preferred information format: the revised Maastricht communication and decision-making process at the consultation level.

In the face of the widespread acceptance that patient centrerness is a fundamental goal in clinical practice, and that shared decisions is one of the key components of this approach, the result of this study confirms that the practice of GPs, as represented by this sample (who are an “above average” sample in terms of MRCGP membership and willingness to participate in this type of research), lies far away from espoused models in books and communication skills courses and, indeed, the wishes of certain patients. Do data from service contexts challenge these espoused models? Are the ideals of patient centrerness and involvement in decision making completely unrealistic for day to day service contexts? Given that clinicians are consistently positive about the principles of patient centrerness and patient participation in decision making processes, perhaps the issue of skill development is only a small obstacle and the structural constraints, particularly the lack of time and readily accessible and relevant information about the harms and benefits of health-care interventions, are the true limiting factors. These practitioners volunteered to have their consultations studied, even so, the results reveal a very limited degree of patient participation. This study, among many others, provides additional evidence for the assertion that successful patient participation demands more time than is currently allocated. Perhaps these results also lend support to others for the need to harness technologies such as decision aids so that consultations have firmer foundations for partnerships.

**Key messages**
- The OPTION scale provides a method of scoring the extent to which clinicians involve patients in the decision making process at the consultation level.
- Content validity was based on formulating the items from existing literature, using the results of a series of studies designed to understand how patient involvement can be best achieved in professional practice, followed by an iterative design and assessment cycle.
- Construct validity was supported by the finding of a correlation between involvement scores and patient age.
- Psychometric assessment also revealed areas where further instrument refinement is necessary.
- OPTION scores for a sample of routine consultations taken from general practice in a UK setting are low.

**REFERENCES**

Shared decision making


Shared decision making: developing the OPTION scale for measuring patient involvement

G Elwyn, A Edwards, M Wensing, K Hood, C Atwell and R Grol

doi: 10.1136/qhc.12.2.93

Updated information and services can be found at:
[http://qualitysafety.bmj.com/content/12/2/93](http://qualitysafety.bmj.com/content/12/2/93)

These include:

**References**
This article cites 24 articles, 9 of which you can access for free at:
[http://qualitysafety.bmj.com/content/12/2/93#BIBL](http://qualitysafety.bmj.com/content/12/2/93#BIBL)

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
[http://group.bmj.com/group/rights-licensing/permissions](http://group.bmj.com/group/rights-licensing/permissions)

To order reprints go to:
[http://journals.bmj.com/cgi/reprintform](http://journals.bmj.com/cgi/reprintform)

To subscribe to BMJ go to:
[http://group.bmj.com/subscribe/](http://group.bmj.com/subscribe/)