

Persons with Rheumatoid Arthritis Challenge the Relevance of the Health Assessment Questionnaire: A Qualitative Study of Patient Perception

DAVID EBBEVI,^{1,*} ANNA ESSÉN,² HELENA HVITFELDT FORSBERG¹

¹ Department of Learning, Informatics, Management and Ethics, Karolinska Institute, Solna, Sweden

² Center for Human Resource Management and Knowledge Work, Stockholm School of Economics, Sweden

*Corresponding author: david.ebbevi@ki.se

Background. The Stanford Health Assessment Questionnaire-Disability Index (HAQ) is widely used to measure functional ability in persons with Rheumatoid Arthritis (RA). The instrument was developed with limited involvement from persons with RA, and their perception of the instrument has not been studied in depth. The aim of this study was to explore how persons with RA experience the use of the HAQ in care.

Methods. Forty persons with RA were purposefully recruited to participate in semi-structured interviews. The interviews were then analyzed qualitatively using thematic analysis.

Results. The participants questioned the relevance of the HAQ but nevertheless experienced that the instrument had a profound effect on their understanding of health and how care is delivered. The analysis resulted in three themes: Problems with individual items, meaning of the summative score, and effects on care and health perceptions.

Conclusions. To make the HAQ relevant to persons with RA, it needs to be revised or to include an option to select items most meaningful to the respondent. To ensure relevance, the HAQ update should preferably be co-created by researchers, clinicians and persons with RA.

Keywords: *Rheumatoid Arthritis, thematic analysis, Outcomes assessment, Qualitative Research, Patient reported outcome measures*

<http://dx.doi.org/>

1. Introduction

Health outcome measurement has long been the focus of quality assessment in rheumatology[1]. The Stanford Health Assessment Questionnaire-Disability Index (HAQ) is a commonly used instrument to assess functional ability in Rheumatoid Arthritis (RA). It has been extensively validated[2], recommended by the American College of Rheumatology[3] and measured in the majority of studies of RA treatments[4]. Patients' perceptions of other instruments used in rheumatology have been qualitatively evaluated in depth[5, 6], but perceptions of the HAQ need further investigation.

During development of the HAQ, patient feedback identified imprecisions and ambiguity in the instrument, but patients did not take part in the initial stages of its development[1]. Excluding patients from initial development stages is common[7] but problematic because patient feedback is limited to the items researchers hypothesize as important. In the same manner, patient preferences among the HAQ items have been evaluated quantitatively[8, 9]. These studies mainly illustrated patients' perceptions of existing items rather than investigating which functional aspects patients would find

relevant to measure. Several qualitative studies mentioned limitations of the HAQ from the patient's perspective[10-13]. However, these studies did not have the specific aim to explore the HAQ and, consequently, did not show how patients experienced the instrument or what consequences the use of the HAQ may generate. Therefore, the aim of this study was to explore how persons with RA experience the use of the HAQ in care.

2. Methods

Empirical setting: Swedish translation of the HAQ and its use in RA care

The HAQ[14] consists of 20 items grouped into eight categories (Table 1). Each category concerns a functional area represented by two to three items. The items aim to capture specific functional abilities and cover all major joints of the body. For each item, the patient is asked to rate the difficulty on a six-point scale. Each item is assigned a score from 0-3: 0 points ("without any difficulty"), 1 point ("with some difficulty"), 2 points ("with much difficulty," "with a special device" or

"with help from another person"), 3 points ("unable to do") [15]. The HAQ score is calculated as the average of the highest score from each category. Hence, the HAQ score ranges from zero to three with a high score representing substantial difficulties.

Table 1. Back-translation of the Swedish HAQ

Category	Are you able to...
Dressing and grooming	shampoo your hair? dress yourself, including tying shoelaces and doing buttons?
Arising	stand up from an armless straight chair? get in and out of bed?
Eating	cut meat?* lift a full cup or glass to your mouth? prepare your own meal?**
Walking	walk outdoors on flat ground? climb down five steps?***
Hygiene	wash and dry your entire body? take a tub bath? get on and off the toilet?
Reach	reach and get down a two-kilogram bag of, for example, sugar from just above your head? bend down to pick up clothing from the floor?
Grip	open car doors? open jars which have been previously opened? turn faucets on and off?
Other activities	run errands and shop? get in and out of a car? vacuum?†

Notes: Based on translation by Ekdahl et al. [15]. Four items deviate from the original [14] as follows: * "cut your meat?" ** "open a new milk carton?" *** "climb up five steps?" † "do chores such as vacuuming or yardwork?"

In Sweden, RA is managed through publicly funded interprofessional specialist care. A majority of patients participate in the Swedish rheumatology quality registry, which stores data on clinical procedures and outcomes, as well as patient-reported outcome measures (e.g., the HAQ). Prior to appointments, patients fill in the HAQ electronically, either from home or in the waiting room. Most patients complete the HAQ once or twice per year. Only one answer per item is possible and at least one item in each category must be answered. Typically, the patient discusses the self-reported HAQ with the physician during the appointment. A graphical interface shows longitudinal trends.

Design

A qualitative design was chosen for this study to capture experiences of HAQ use that were as true to patients' experiences as possible [16]. It has been argued that this design is necessary to assess the patient value of a patient-reported outcome measure [17]. The study was based on 40 semi-structured interviews with persons with RA collected in two studies [18, 19] by the first and second authors. This specific material was chosen because the data contained rich patient accounts of the HAQ. The studies were approved by the

Regional Ethics Board in Stockholm (reg.nr: 2009.895-31.5 and 2012.1911-31.5).

Data

The first study [18] included 22 patient interviews and concerned patients' overall experiences of RA, including the structure, process and outcome of RA care. The second study [19] included 18 patient interviews and concerned patients' experiences of the process of care, including the usability and deployment of the electronic system patients use to enter self-reported HAQ data. Participants of both studies were purposefully sampled to create diversity in age, gender, disease severity, disease duration and satisfaction with care (i.e., maximum variation sampling [16]). The first and second authors conducted the interviews, which ranged from 60 to 115 minutes each, between 2009 and 2013. The interview location was chosen by each participant to promote a safe atmosphere. Interview guides were used in both studies to allow flexibility in conversation and allow patients to talk about unexpected issues [16]. Questions included, for example: What is your experience of self-registering measurements? How do you assess the different dimensions? How does your physician use the registry when you meet? Tell me about your last visit, what was it like? How can aspects of care be measured and for whom? How does the care influence your ability to live a normal life? Does the physician focus on the right areas during an appointment?

Analysis

The respective authors transcribed their interviews, and the first author thematically analyzed the data to capture manifest and latent content using the method described by Braun and Clarke [20]. The process is outlined with examples in Table 2. The first author conducted the analysis using NVivo v.10. When faced with ambiguities, interpretation was facilitated by Watzlawick et al.'s communication theory [21]. For example, participant dissatisfaction followed by expressions such as "I might think, kind of..." was not interpreted as the participants being unsure of their dissatisfaction, but rather as maintaining a relationship with the interviewer.

3. Results

The data suggested patients had not only heterogeneous experiences of the use of HAQ, but also commonalities. The various experiences salient in the data were grouped into the three themes outlined in Table 3.

Many aspects of the themes below are illustrated with quotes and, if contextually relevant, a short description of the participant behind the quote. Parenthesis after quotes show: Woman (W) / Man (M); age in years; and years since RA debut.

Problems with individual items

The individual items of the HAQ created much frustration for participants, leading them to describe the HAQ by many labels: "old-fashioned," "childish," "silly," "narrow," "retired-like," "too general," "unnecessary" or "dead boring." Despite positive experiences, negative experiences dominated the data. The results below are therefore presented as subthemes related to experienced negative properties of items: missing, unclear, unnecessary or static.

Missing items

Participants were asked about activities important to them before being primed by the HAQ. Their responses included factors present in

the HAQ: opening jars, cutting meat, using the toilet, showering, taking baths, getting dressed, tying shoelaces, vacuuming and cooking.

Responses also included factors not covered by the HAQ: wringing out

Table 2. Thematic analysis

Step	Description	Example from analysis in the present study
Familiarizing	Reading the data multiple times and noting ideas	Ideas: <ul style="list-style-type: none"> The instrument affects what is spontaneously seen as functional ability. Is this reflexivity or validity? Many items seem to be missing from HAQ HAQ complements other measurements
Coding	Applying open codes to data relevant to the research question	Coding: <ul style="list-style-type: none"> "I wish I could dress in what I wanted by myself. Just that is important. I would like to be able to shower myself, wash my hair, these kinds of practical things. Caring for myself, that would be most important" was coded as "wishing to dress, shower and wash hair by oneself"
Searching for themes	Grouping codes into initial subthemes and themes	Forming tentative subthemes (here with example of codes): <ul style="list-style-type: none"> <i>Function affects and is affected by the HAQ</i>† <ul style="list-style-type: none"> cleaning and cutting meat wishing to dress, shower and wash hair by oneself difficulty biking, drinking coffee and wiping the stove <i>Missing items</i> <ul style="list-style-type: none"> Questions missing can be discussed at the appointment HAQ is so narrow, I'm declared cured
Reviewing	Checking if the themes represent their codes and all relevant data	Merging subthemes and moving codes: <ul style="list-style-type: none"> The subtheme <i>Function affects and is affected by the HAQ</i>† was incorporated into <i>Missing items</i> because the data as a whole did not clearly support reflexivity of items The code "HAQ is so narrow, I'm declared cured" was moved back and forth between <i>Static items</i> and <i>Missing items</i> and ended up at <i>Missing items</i>
Defining	Analysis for renaming themes and formulating an explicit definition	Renaming: <ul style="list-style-type: none"> Subtheme <i>Effect on behavior</i> clarified as <i>Effects on physician behavior</i> A theme, <i>A wish to change the HAQ</i>, concerned only experiences of the individual items and not the participants' wishes, so instead was renamed <i>Problems with individual items</i>

Notes. Inductive analysis[20] in the present study, subthemes in italics. Even though described as a linear process, the themes were created iteratively, moving back and forth between the different steps. The themes were defined to maximize inner homogeneity and outer heterogeneity[16]. However, as themes rather than categories, they are not necessarily mutually exclusive. Hence, some data demonstrate more than one theme (e.g, the code "HAQ is so narrow, I'm declared cured"). †This theme concerned an early hypothesis that regular use of the HAQ would affect participants' perceptions of function, i.e., drawing attention to abilities in HAQ would make respondents more attentive to the abilities even though they were not important prior to exposure to HAQ (reflexivity).

Table 3. Summary of findings

Themes and subthemes	Defining the experience of...
Problems with individual items	frustration when responding to HAQ items
Missing items	wanting to include items in the HAQ that are not currently there
Unclear items	not knowing how to interpret a specific item
Unnecessary items	answering items perceived as not meaningful
Static items	the same items being scored every time the instrument is used
Meaning of the summative score	frustration and insights when trying to understand the summative score
Capturing holistic perspectives	the HAQ capturing a complete picture
Reflecting a temporary state	the HAQ as a snapshot of health
Requiring strategies for interpretation	understanding the score in relation to other measurements or the physician's interpretation
Effects on care and health perceptions	positive and negative effects from use of the HAQ in care
Effects on physician behavior	how use of the HAQ would cause some physician behavior

Effects on understanding of RA	how use of the HAQ would cause understanding of health
a soaked washcloth, holding a cup of coffee, dressing a child, opening locked doors, sitting for a long time, brushing hair, filling up the car, doing yardwork (present in original HAQ but not in the Swedish translation, see Table 1), bicycling, writing, painting and kneading dough. Typically, participants experienced limitations of the instrument but had difficulty presenting alternative items:	doctor], but he says this [HAQ] is made and the entire world uses it. (W, 66, 48)
I cannot come up with some concrete proposal of how it should be instead. Not right now anyway. Erhm, but some issues I think are very coarse....I would think that the wording could perhaps be changed on some things. (W, 49, 29)	In addition, patients criticized even the alternatives. For example, participants did not know how to answer when they could do something either with a special device or with the help of someone else, since the system only allowed for one answer.
When asked directly about what was missing, respondents often suggested measuring phenomena that are not functional, such as tiredness or factors relating to social life. However, physical-function items experienced as missing included sexual function, individualized items (exemplified below under Static items) or items that captured nuances of higher function. A person with low disease activity said:	<i>Unnecessary items</i> In contrast to unclear items, respondents regarded as unnecessary and skipped HAQ items concerning situations the respondent would never encounter:
You'll almost be declared cured [by the HAQ] when you write there, but there is so much other stuff I cannot do....If I would walk around in [this shopping mall] for three hours, then I would have to go in here [the café] to sit and rest. Then I would be in pain....Therefore I avoid certain things. Buy in bulk? Yes, of course I can do it. But I cannot go around [all the shops in the center], for then it will be, then it is over. (M, 37, 4)	Are you able to take down a pack of sugar? It is completely irrelevant to me. Can you bathe in a tub? I don't know, I don't have one. It appears to be designed for a specific type of rheumatism, age-rheumatism. Not for us who are young and have a completely different situation. (W, 40, 9)
<i>Unclear items</i> Some HAQ items raised questions among participants about the intention of the HAQ developer. This items included cutting, bathing, cooking, driving cars, picking up stuff from the floor, reaching for and getting down sugar from a shelf and walking on even ground. When encountering these items, participants blamed either themselves or the instrument developer for not understanding the items. In trying to make sense of them, participants related the items to highly individual experiences or stories. For example, participants understood the ability to take a tub bath differently: It depended on whether or not they had someone helping them, they filled the bathtub first or not or the side portion of the bathtub was removable.	Sometimes, participants saw unnecessary items as the result of standardization (detailed in the next section). Typically, participants with low disease activity also felt as though the items were too similar. One even found the repetitive process of answering to be childish:
In general, participants made one of two kinds of interpretations of unclear items. In one case, the item would be interpreted as though it reflected a particular part of their health status:	I might think that it's a bit childish...but there are certainly those who are sicker than me....But there is a bit of repetition....For me, there will be "no," "no," "no" all the way. But I understand that others will respond differently. (W, 68, 10)
There is a question about how to grab things from the floor, I think. And then I know that I'm always thinking, "But God, how will I answer that?" ...When I have pain in that wrist, then the problem is that I have difficulty grabbing small things. That is, if a sock fell on the floor or so. It's not that I have difficulty bending down and picking up something from the floor...[but] I find it difficult to get ahold of that little thing that might lie on the floor. That's really what I would like to tell you, not that I find it difficult to pick up something from the floor. (W, 51, 10)	<i>Static items</i> Participants commonly experienced having the same items in every HAQ measurement as standardization, which gave rise to positive and negative experiences. On a positive note, static items made participants think of areas of function they did not normally consider and thereby understood that their functional level had decreased. Negative experiences arose either from the questions or from the interaction with the physician after answering the HAQ. Participants reasoned that items irrelevant to them were there to capture divergence of the population. Despite this insight, the static items caused lack of patient engagement. This in turn made participants less attentive when answering the questions. For example, a woman who was generally satisfied with her care and perceived her disease activity and functional ability as very variable said:
In the other case, the item would be interpreted as though it reflected a specific type of activity:	Honestly speaking [answering] is a bit repetitive. Had they made some different wording and questions, one would probably get more involved and read the questions....Of course, I can walk on flat ground because I have problems with my hands....A bit too general questions. (W, 27, 9)
Can you cook? Yes, depending on what kind of food. [Instead,] it's, you can cut meat? Yes, if it's tender meat. What kind of knife am I using? I get so mad every time I answer those. I have told [my	Similarly, she would not answer general questions from her physician candidly. For example, she would tell the physician that she was fine even when she was not and acknowledged it only when her physician saw through it. Another participant suggested that static items should be personalized and go beyond functional ability:

It is the same questions all the time. It's a shame. It should vary. Do you cook? How is cooking? Are you eating? How long can you walk today? Have you been in town? Have you met acquaintances? Do you read? What are your interests? Because I mean, I have my son so I do not sit around on my own. But people who are alone cannot get out on their own. (W, 57; 2.5)

By contrast, if the results were supplemented with patient-physician communication after entering the data into the registry, static items were not necessarily a problem:

I do not feel like the registry standardizes care. I mean, it's for me, it's my values we're talking about. I do not have the same numbers as [other patients]. So that's good. If they fill out the same field as I do, I don't care. Even if I have symptoms that do not fit in the registry, the physician can enter it somewhere in some system. (W, 68, 10)

Meaning of the score

The HAQ score captured a holistic, multidimensional and continuous view of health status but was limited by the temporality and interpretation of the summative score. Understanding of the results was deepened by relating the HAQ score to other numbers or having earlier experience with the measurements.

Capturing holistic perspectives

When thinking about the items of the HAQ, participants related them to a much larger story. They understood the HAQ as a way of bringing the intricacies together into a complete picture:

I think [the HAQ] is relevant because it is good to know. Can this person get down [something that is high up], because it's pretty elementary. I had a mother who always put things [high] up because she felt it was exquisite exercise. So I usually take after her. But some things I have abandoned: no iron, no pans of iron and things like that because it is so much work....So therefore, Teflon. But on the other hand, the iron pots are so good to cook in, and it is so infrequently that one can use them. But right then, you adapt. But on the other hand, they are important, therefore, because then you can get a complete picture. (W, 56, 32)

Sometimes a holistic evaluation influenced how individual items were answered and results interpreted. A woman diagnosed two years ago and now working as a research nurse described how she would fill out the HAQ without thinking it through whenever she felt her disease activity had been low since her last HAQ response. By contrast, another woman called the summative score "wrong" when it did not agree with her holistic picture of her health status:

It's also difficult because we all have different diagnoses, and it [the HAQ] is supposed to pinpoint all of them. I get that, sort of. But for me, sometimes they are really stupid, sort of. So it [the HAQ score] will even give the wrong result. (W, 49, 29)

Reflecting a temporary state

Participants noted that the HAQ only captured a temporary health state and thus gave a snapshot rather than a longitudinal understanding. This made them wish for either rephrasing the

questions or higher measurement frequency. A person with generally low disease activity described:

You take the previous week [into account]. I do not meet my doctor for a year, so then I would like to [answer] maybe three months ago. Then I could not get my arms up here, but I can today, get those up there. So when I register today, it is great. But why was it like this three months ago? Why was it like that six months ago? Somehow [it would be good if] one registered all the time in between, before going to the doctor. So you can see that there have been some periods. A small diary or something, because it's so easy to forget. (W, 61, 19)

Participants also viewed measurements spaced through time as a way to assess regularly the same aspects of one's health, which gave a sense of continuity of care.

Requiring strategies for interpretation

The summative number of the HAQ score did not instantly make sense. Participants therefore struggled with the meaning of the score relative to their general feeling of function, as mentioned above under Capturing holistic perspectives. They also struggled with it relative to other measurements:

There are instructions. But they do not explain really what the system is for. Some are probably wondering, "Why do I enter this?" - The obvious benefit to patients. What does the figure show? 2.94 - what does it mean? There are better systems in other diseases that relate an index to a relevant measure of disease progression. (M, 57, 35)

Participants noted that it is easy to feel like they under- or over-scored in a self-assessment, for example, if they were not feeling well but were fully functional in terms of the questions contained in the HAQ. This became evident as participants discussed the results with their doctors. In particular, it was obvious for persons with low disease activity but high functional demands who scored low in the HAQ but felt their function was insufficient for their needs. If the HAQ score correlated badly with self-perceived health, participants advocated using other measurements such as clinical assessment or laboratory tests that sometimes provided different pictures. One participant who felt the HAQ questions were not for her but for sicker patients stated:

I have no need to analyze myself. I want the doctor to do it....I do not think it reflects how I feel. The doctor should rely more on blood tests than those [HAQ] tests. (W, 64, 1)

Another strategy to overcome the struggle of interpreting the score was to relate to how the physician would interpret the numbers. That is, participants wanted to understand how the physician would extract meaning from the HAQ score.

Effects on care and health perceptions

Participants experienced that the HAQ affected care by changing physician behavior during the appointments and affected the perception of health through their own understanding of RA.

Effects on physician behavior

The experienced effect of registering the HAQ was either a perceived lack of effect or various degrees of influence on physician

behavior. Participants sometimes saw the HAQ as a way to save time because physicians would not need to ask questions related to function. Patients perceived this either as good for the physician's productivity solely—with no benefit for the participant—or as beneficial for both because the saved time could be used for other activities that benefited the participant.

In the case of perceived extensive effect on physician behavior, the effect depended on the summative score and the answers to individual items. For example, if the answers showed trouble opening a jar, the physician could prescribe tools to help or understand that the participant was not able to work in demanding environments such as a repair shop. Therefore, the physician's use of individual items could determine if a participant liked or disliked the HAQ in general. A high summative score would have corresponding effects. Similarly, the HAQ could be a tool for the participant to influence care and the physician's behavior. This even went as far as participants wanting to change the HAQ because the current items lacked an effect on physician behavior in some respect:

The questions are so fixed, and I might have other things. I have, for example, a damn hard time to lift one of my legs into the car. I need to use my hands and I would like to highlight that because I am lobbying a lot for an operation. I have so much pain that I cannot walk in certain periods. But they say that they don't operate on young [people]. (W, 40, 9)

In addition, participants worried the HAQ self-assessment could negatively affect the quality of the general assessment because either it made doctors less prone to assessing function themselves or hard for patients to communicate their health status through the questions. As one participant observed in the waiting room:

There was some older woman there and she had a little difficulty to try to respond to that [the HAQ], so she got help from a nurse. I can, well, I think it is a pity; they might need more personal contact to get it explained and to tell more about how it is. (M, 46, 10)

Effects on understanding of RA

While the summative score triggered questions as described in Requiring strategies for interpretation, the process of using the HAQ also affected participants' understanding of RA. Patient engagement with the HAQ varied widely: Some answered as quickly as possible whereas others saw answering as an opportunity to reflect upon their functional status. Participants who did not discuss the results with the physician felt as though responding to the HAQ wasted time or the opposite—that there was no need to discuss it. Participants with the latter view instead thought the questions were valuable for gaining awareness of their problem or for documentation purposes. A participant habituated to constant pain and who, despite having symptoms considered her disease activity low, verbalized this experience:

[The HAQ can] be good for me, "Devils, I cannot do anymore,"...sometimes you get so used to your limitations that you do not think about that or you adapt, maybe I should say. You put the stuff further down, right? So they may not be stupid questions, [rather they] get yourself to wake up. (W, 56, 32)

In contrast, the items also illustrated examples of how sick one could get, and some participants viewed that awareness as negative. This was captured by a person with low disease activity:

It also feels as if the questions do not apply to me. I mean, I feel so good after all. I can lift a bag of flour. I can walk on level ground. It just makes me afraid because that's where I will be. I feel sicker. It almost feels as if I should mark that something is wrong. (W, 64, 1)

4. Discussion

The findings illustrate that persons with RA mainly experience the HAQ as flawed but, despite this, as positively affecting understanding of RA and the care received. The items cause frustration, but understanding RA facilitates interpretation of one's health. This was captured in three themes: Problems with individual items, illustrating HAQ limitations experienced in terms of individual items; Meaning of the score, experiencing HAQ strengths and limitations in terms of the summative score; and Effects on care and health perceptions, the positive and negative experiences of how HAQ use affects care delivery and understanding of health for persons living with RA.

The findings show how regular use of HAQ not only describes function, but also can affect care by facilitating communication between patient and physician. However, more interestingly, these findings suggest that patient conception of functional ability differs markedly from what the HAQ measures. For example, the participants experienced that the HAQ does not capture nuances of low disability (i.e., a perceived floor effect). Capturing such nuances was of great importance to participants with low disease activity but might be less obvious to instrument developers for whom such nuances are lost in comparison to patients with substantial functional impairment. That dissimilarity implies that current versions of the HAQ (in use in clinical practice) fail to generate data about functions persons with RA consider highly significant. This inadequacy could partly be traced to limited patient involvement in the HAQ development process: Patients were included only at the end of the process. Rather, the development of the HAQ departed from a professional research interest, with no consideration given to the effects on physician behaviors when the HAQ is regularly used. In addition, developers were arguably influenced by the expected outcome (in 1979) being different from the one seen today with early efficient treatment and tight monitoring.

The literature provides a similar picture of HAQ deficiencies

Earlier research extensively used qualitative methods to study the functional needs of persons with RA (e.g. [22, 23]). A few studies have, in addition, related those findings to the HAQ, such as highlighting how context affects HAQ relevance[11], as discussed in Unnecessary items; noting the floor effect[12] as presented in Missing items; or seeing the HAQ as a mere representation of more complex activities[13], as shown in Capturing holistic perspectives. The last of the above studies[13] made the point that tasks detailed in individual HAQ items sometimes are not themselves important but rather pose the challenge of working around them. For example, to run errands and shop is not the person's end goal but one of many ways in which that person can get groceries home for cooking[13]. Other subthemes in the present study were not previously described.

This study shows that persons with RA have different priorities among the HAQ items and functions. In the same manner, quantitative studies have investigated patient priorities among the HAQ items and demonstrated that priorities differed[9] with rather low agreement between persons with RA (weighted $\kappa = 0.241$)[8]. Interestingly, no HAQ item seems to be unimportant to all persons with RA[15]. If the instrument were individualized by selecting the five most important

items or weighted by adding additional questions about importance, then construct validity could be preserved[9]. Persons with RA freely named activities they considered important before being exposed to the HAQ and 31% of those important activities are not in the HAQ. Most of the named actions were leisure activities such as playing the guitar[8]. Leisure activities were also seen in Missing items of the present study.

Conflicts between scientific value and patient value

Measuring function with the well-spread HAQ can be defended in that it enables international and historical comparisons necessary for quantitative scientific evaluations of RA interventions. However, achieving such comparability comes with the risk of poorly correlating true functional ability with the measurement (i.e. HAQ validity). For example, participants sometimes skipped the question concerning the tub bath because they did not have a bathtub. For persons who can wash their bodies and use the toilet but might be unable to take a tub bath, having or not having a bathtub would affect the score. (Because these activities are scored together, see Hygiene in Table 1). This specific case could be tackled by explaining to respondents that they should imagine a response rather than skip such questions.

In general, societal development ensures a constant change of human behavior and activities. Therefore, regular revision of activity-based functional scales is necessary. Such an update of the HAQ could measure the same construct as the contemporary HAQ by closely correlating them and thus still make historical comparison possible. In a sense, this means “translating” the HAQ into a new time, in the same way it has been translated into many languages.

So how would this be done? To measure function in terms of activities, one must handle divergence. This challenge is twofold: The instrument needs to capture the divergence of both the population and the activities. Answering both challenges would make it difficult to close the gap between individualized information relevant for persons with RA and general information that can be communicated in an instrument. Despite this difficulty, it is obvious from the present study that there is room for improvement. Future studies might investigate whether the same functional demand could, for example, be captured in activities patients consider more meaningful, such as social[24] or leisure activities[25]. In this way, the importance of social and leisure activities seen in this and other RA studies would be better acknowledged.

Limitations

As seen in the present study, the character of specific items influenced the participants' experiences. Therefore, Ekdal et al.'s context adjustments of the HAQ[15] during translation might affect transferability of the findings. Transferring the findings is especially difficult because the rationale for translating, for example, “climb up five steps” into “climb down five steps” is not explained in their publication. However, most findings in the present study are not tied to items “adjusted for context”.

Further, this study used data collected with other research questions in mind. Although the findings themselves are more important than the research question in exploratory qualitative research, additional aspects of participants' experiences might have been better captured with a directed research question. The authors countered this by using a large number of interviews, but such breadth cannot substitute for any lack of depth. In the same manner, cognitive interviews with participants while they responded to the HAQ would have provided their immediate experiences of the HAQ. Whereas the focus of this article was instead the participant's long-term lived experiences and creation of meaning, additional data from cognitive interviews would likely have provided supplementary information.

Finally, the selection of participants did not include persons using the HAQ as part of a research trial. It is reasonable to believe that those persons would have different experiences because the influence of the HAQ in a trial is not as tangible as it is in care, and transferability of findings to research settings

Conclusion

The aim of this study was to explore how persons with RA experience the use of the HAQ in RA care. The findings mainly illustrate three types of experiences: wrestling with the limitations of the individual items, searching for meaning in a partly faulty summative score and handling the effects of using the HAQ. These findings suggest there is value in using a functional scale in clinical care, but the HAQ needs either revising by patients or individualized prioritization among items. Future research should investigate the relevance of items in different contexts and, to ensure relevance to persons with RA, co-produce an alternative instrument with relevant items and routines for use.

Funding Information. The study was funded by Vetenskapsrådet and Forte (the Swedish Research Council for Health, Work Life and Welfare, grant number 2012-1688, 2014-4238). Funders played no part in the design or execution of the study.

Acknowledgment. We thank all participants for open-heartedly sharing their experiences; Professor Staffan Lindblad, KVALNET and Carolina Wannheden for valuable support and feedback; and finally, Donna Magnani for excellent language editing.

5. References

1. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheumatol*. 1980;23(2):137-45.
2. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: Dimensions and practical applications. *Health Qual Life Outcomes*. 2003;1(1):1.
3. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheumatol*. 1993;36(6):729-40.
4. Kalyoncu U, Dougados M, Daures J, Gossec L. Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: A systematic literature review. *Ann Rheum Dis*. 2009;68(2):183-90.
5. Paterson C. Seeking the patient's perspective: A qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Qual Life Res*. 2004;13(5):871-81.
6. Ong BN, Hooper H, Jinks C, Dunn K, Croft P. 'I suppose that depends on how I was feeling at the time': Perspectives on questionnaires measuring quality of life and musculoskeletal pain. *J Health Serv Res Policy*. 2006;11(2):81-88.
7. Wiering B, Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: A scoping review. *Health Expect*. 2016;doi:10.1111/hex.12442.
8. Hewlett S, Smith AP, Kirwan JR. Values for function in rheumatoid arthritis: Patients, professionals, and public. *Ann Rheum Dis*. 2001;60(10):928-33.
9. Seror R, Tubach F, Baron G, Guillemin F, Ravaud P. Measure of function in rheumatoid arthritis: Individualised or classical scales? *Ann Rheum Dis*. 2010;69(01):97-101.
10. Dür M, Coenen M, Stoffer MA, Fialka-Moser V, Kautzky-Willer A, Kjeken I, et al. Do patient-reported outcome measures cover personal factors important to people with rheumatoid arthritis? A mixed methods design using the International Classification of Functioning, Disability and Health as frame of reference. *Health Qual Life Outcomes*. 2015;13(1):1.

11. Sanderson T, Hewlett S, Calnan M, Morris M, Raza K, Kumar K. Exploring the cultural validity of rheumatology outcomes. *Br J Nurs.* 2012;21:1015-23.
12. Sverker A, Östlund G, Thyberg M, Thyberg I, Valtersson E, Björk. Dilemmas of participation in everyday life in early rheumatoid arthritis: A qualitative interview study (The Swedish TIRA Project). *Disabil Rehabil.* 2015;37:1251-59.
13. Prodinger B, Shaw L, Stam T, Laliberte Rudman D. Enacting occupation-based practice: Exploring the disjuncture between the daily lives of mothers with rheumatoid arthritis and institutional processes. *Br J Occup Ther.* 2014;77:491-98.
14. Fries J. The assessment of disability: From first to future principles. *Rheumatology.* 1983;22 Suppl 1:48-58.
15. Ekdahl C, Eberhardt K, Andersson S, Svensson B. Assessing disability in patients with rheumatoid arthritis: Use of a Swedish version of the Stanford Health Assessment Questionnaire. *Scand J Rheumatol.* 1988;17(4):263-71.
16. Patton MQ. Qualitative evaluation and research methods. 2nd ed. Thousand Oaks: SAGE; 1990.
17. Trujols J, Portella MJ, Iraurgi I, Campins MJ, Siñol N, de Los Cobos JP. Patient-reported outcome measures: Are they patient-generated, patient-centred or patient-valued? *J Ment Health.* 2013;22(6):555-62.
18. Ebbevi D, Forsberg HH, Essén A, Ernestam S. Value-based health care for chronic care: Aligning outcomes measurement with the patient perspective. *Qual Manag Health Care.* 2016;25(4):203-12.
19. Essén A, Lindblad S. Innovation as emergence in healthcare: Unpacking change from within. *Soc Sci Med.* 2013;93:203-11.
20. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* 2006;3(2):77-101.
21. Watzlawick P, Bavelas JB, Jackson DD, O'Hanlon B. *Pragmatics of human communication: A study of interactional patterns, pathologies and paradoxes.* New York: WW Norton; 2011.
22. Ahlmen M, Nordenskiold U, Archenholtz B, Thyberg I, Ronnqvist R, Linden L, et al. Rheumatology outcomes: The patient's perspective. A multicentre focus group interview study of Swedish rheumatoid arthritis patients. *Rheumatology (Oxford).* 2004;44:105-10.
23. Marshall NJ, Wilson G, Lapworth K, Kay LJ. Patients' perceptions of treatment with anti-TNF therapy for rheumatoid arthritis: A qualitative study. *Rheumatology (Oxford).* 2004;43:1034-38.
24. Baumeister RF, Vohs KD, Aaker JL, Garbinsky EN. Some key differences between a happy life and a meaningful life. *J Posit Psychol.* 2013;8(6):505-16.
25. Unger LS, Kernan JB. On the meaning of leisure: An investigation of some determinants of the subjective experience. *J Consum Res.* 1983;9(4):381-92.



© 2017 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).