Content validity, construct validity and magnitude of change for the eight-item HAKIR questionnaire - a patient reported outcome in the Swedish national healthcare quality registry for hand surgery.

**CURRENT STATUS:** POSTED

Ingela K Carlsson  
Lunds Universitet

Elisabeth Ekstrand  
Lunds Universitet

Mikael Åström  
Dept of Data Analytics Lund

Kerstin Stihl  
Södersjukhuset, Stockholm

Marianne Arner  
Karolinska Institutet

✉ marianne.arner@ki.se  
**Corresponding Author**  
**ORCiD:** https://orcid.org/0000-0002-6744-9465

**DOI:**  
10.21203/rs.2.18115/v1

**SUBJECT AREAS**  
Orthopedic Surgery, Health Economics & Outcomes Research

**KEYWORDS**  
Patient Reported Outcome Measures, Hand, Quality of Healthcare, Surveys and Questionnaires
Abstract
Background The Swedish National Healthcare Quality registry for hand surgery, called HAKIR, includes collection of patient-reported outcome measures (PROMs) for all operations performed at the specialist departments of hand surgery. A prerequisite for PROMs is that the collection of data is based on psychometrically sound outcome instruments. This study therefore aims to evaluate content and construct validity, floor and ceiling effects, data completeness and magnitude of change over time for the eight-item HAKIR questionnaire.

Methods Content validity was investigated by patients, an expert group and literature review. Construct validity was investigated through predefined hypotheses and correlation statistics between single-item questions and QuickDASH. Floor and ceiling effect and data completeness was analysed on 13197 preoperative, 10702 three months and 9986 twelve months postoperative questionnaires. Changes in scores for single-items between pre- and 3 and 12 months follow-up for elective hand-related diagnosis and between 3 and 12 months for acute injuries was quantified by effect size calculations.

Results The HAKIR Questionnaire (HQ-8) included items concerning pain on load, pain on motion without load, pain at rest, stiffness, weakness, numbness, cold sensitivity and ability to perform daily activities. Correlation coefficients between single-items and total QuickDASH score at pre, 3 and 12 months follow-up ranged from 0.44-0.79. No ceiling effect but a floor effect in the total group was seen in all items at all follow-ups. Percentage of missing item responses was < 2.6% except for the cold sensitivity question. The magnitude of change for individual items varied between small to large effects size in elective hand-related diagnosis. For acute injuries a small effect size was seen between 3 and 12 months follow-up.

Conclusions This study provides evidence of content and construct validity of the HQ-8 including lack of ceiling effect, expected floor effect, good data completeness and ability to detect changes in scores over time. Associations between single-items and QuickDASH indicate that HQ-8 measures unique disability aspects important in hand surgery. Further studies are needed to evaluate test-retest reliability, responsiveness including clinically important change of the HQ-8 questions in subgroups of
hand-related diagnoses.

**Background**

The Swedish National Healthcare Quality Registries (NQRs) have brought about considerable improvements in different fields of medicine [1, 2]. The first NQRs in the 1970s were orthopaedic arthroplasty registries, registering all joint prostheses that had been implanted and removed. As one result, Sweden now has one of the lowest revision rates for knee and hip implants in the world [3, 4]. The first NQR for hand surgery in Sweden, called HAKIR, was started in 2010 and includes all operations performed at the seven specialist departments of hand surgery in the country. To our knowledge there are no other healthcare quality registries evaluating outcome of hand surgery with national coverage. The broad ambition of HAKIR in collecting patient-reported outcome measures (PROMs) for all operated patients can be regarded as quite unique internationally.

Evaluating outcomes after hand surgery is complex since the distinction between a good or a bad result cannot be made simply by measuring joint motion or grip strength or following the rate of revision of joint implants. Common indications for hand surgical procedures are often perceived symptoms like ache, pain and paraesthesias. PROMs should therefore be included in a NQR for hand surgery to give a complete picture of treatment results. The short version of the Disabilities of the Arm, Shoulder and Hand (Quick DASH) [5] as well as a single-item questionnaire with seven questions concerning perceived symptoms in the affected/injured hand and one question about the ability to perform activities of daily living have been included in HAKIR since the start. This eight-item questionnaire is in future text referred to as HAKIR questionnaire (HQ–8) [6] and can be found in the online Additional file 1, HAKIR questionnaire, HQ–8.

A prerequisite for PROMs in a quality registry such as HAKIR is that the systematic collection of data is based on psychometrically sound outcome instruments. The QuickDASH has retained equal measurement properties as the original DASH with strong evidence for reliability and validity [7]. These aspects, however, have not been evaluated for the HQ–8. Content validity which is the extent to which the content of the instrument is an adequate reflection of the construct to be measured therefore needs to be evaluated [8]. Construct validity refers to the degree in which scores in an
instrument relate to other measures that are consistent with predefined hypotheses concerning the concept being measured [9]. The outcome measures should also be valid for the purpose and population and responsive enough to identify true and clinically meaningful changes in function, not just a change due to random error [10, 11]. Floor and ceiling effect occurs when a considerable proportion of subjects score the best or worst score [12]. The measure is then unable to discriminate between subjects at either extreme of the scale. This may indicate limited content validity and responsiveness as well as reduced reliability because changes cannot be measured [9]. Data completeness (item-response) is another feature important for the content validity of questionnaires. The aim of the present study was to evaluate content and construct validity as well as floor and ceiling effect, data completeness and the ability to detect magnitude of change for the single-item questions (HQ–8) included in HAKIR.

Methods
The HAKIR registry is web-based with a secure logon function. The legal requirements for healthcare registries in Sweden do not demand active consent from the participants. There are however strict demands on information to all participants before registration, for instance about the option of not being registered (opt-out) and the possibility of having all personal data erased at any time. When patients are scheduled for hand surgery, they are informed orally about HAKIR by staff, as well as being informed through an information brochure which is available in six languages (https://hakir.se/wp-content/uploads/2018/06/Eng_Patientinformation_2018.pdf). On the back of the brochure, a tag with username and password is attached and the patients log in themselves at a computer at the clinic. At three and twelve months after surgery, postoperative web questionnaires are automatically sent out by the registry to the e-mail address that the patient had registered. After 2 days, an sms is sent as a reminder. Some patients request a paper form instead, which they then get at their visit to the clinic. These patients are sent the postoperative questionnaires by surface mail by staff with a payed return envelope. The data in the questionnaires is then registered by staff. Paper questionnaires are very work demanding and costly, and eight out of nine participating departments have completely transitioned to web questionnaires [13].
All performed operations at each participating department are included in the basic registration in HAKIR, where surgical codes and reoperations are registered by hospital staff. Patients are asked to complete the HQ-8 and QuickDASH before, as well as three and twelve months after their surgery [13].

Exclusion criteria for questionnaires are age below 16 years, or cognitive problems, reoperation within one year, lack of Swedish social security number and patients who have declined participation in the registry. The seven departments for specialized hand surgery in Sweden have joined HAKIR sequentially since 2010 and the registry gained national coverage in 2014. Two private units also participate with mainly out-patient surgeries. Specialized hand surgery in Sweden is mainly performed at the seven hospital departments, located at the university hospitals around the country. Private units for hand surgery are usually small, single surgeon units mostly doing short out-patient surgeries. Emergency hand surgery, microsurgery and pediatric hand surgery is only performed at the seven hospital departments. Coverage is routinely followed in the registry and more than 80% of all surgeries performed at the participating units are included, to date more than 110 000 operations. Response rates for the questionnaires have varied since the start in 2010 from around 30% to about 60% at present. The response rates are similar between web and paper forms.

The data collection for the present study took place between February 1\textsuperscript{st} 2010 and December 31\textsuperscript{st} 2015, including 13197 preoperative questionnaires, 10702 three months and 9986 twelve months after surgery, giving a response rate of 45% for 3 months and 47% for 12 months questionnaires. A selection of common diagnostic subgroups were formed based on ICD10 coding (International Statistical Classification of Diseases and Related Health Problems - Tenth Revision) and surgical codes and analysed separately. The elective subgroups were; thumb osteoarthritis, Dupuytren´s contracture, Morbus de Quervain, ganglion, trigger finger, carpal tunnel syndrome, ulnar nerve entrapment and the acute subgroups were; digital nerve injuries, extensor and flexor tendon injuries, thumb ligament injury and finger fractures, see the online Additional files 2–4, Supplementary Table S1-S20.

HQ-8 - Content validation process
The choice of questions concerning different hand symptoms was initially made by the registry manager and a senior hand surgeon at the hand surgery department in Stockholm. The selection of symptoms was based on clinical experience and previous literature [14–21]. The symptoms were; pain on load, pain on motion without load, pain at rest, reduced range of motion, reduced strength, reduced sensibility or numbness and hand function in daily activities. The content validity of these questions was assessed by cognitive interviewing (think aloud) [22, 23] of seven patients (3 men and 4 women, age 25–68 years) while they were responding. Patients´ views on the relevance of each item, as well as any ambiguities of formulations were noted during the interviews which were performed by an experienced hand therapist. An oral summary was made together with each patient. The field notes were then transcribed and analyzed with content analyses by the interviewer (fourth author, KS) [24]. A review of the included questions and formulations was then performed by an inter-professional expert group consisting of the registry manager, hand surgeons, occupational and physiotherapists as well as nurses from all hand surgery departments. The final version of the HQ–8 was reached through consensus and unanimous decision and included eight questions [6]. As in the initial process mentioned above the review of literature and following discussion included a discussion about the clinical importance and relevance of the included symptoms. The final number of symptoms was limited not to overload the patients. (See the online Additional file 1, HAKIR questionnaire, HQ–8).

**HQ–8 –scale design**

During the period 2010 to 2013 the design used was as a horizontal visual analogue scale (VAS, 0–100) with verbal anchors at the beginning and end of the scale (0 representing no problem and 100 worst problem imaginable). This design was in 2013 changed to a numeric 11 point box scale (NRS–11) [25] supported with numerical descriptors (0, 10, 20 …100) upon the box. The verbal anchors remained unchanged [6]. Patients were asked to respond to their experienced problems during the last week *in the hand/arm relevant for surgery*. In case of surgery for an acute injury, patients were asked to estimate perceived problems prior to the injury.

**QuickDASH**

Furthermore, patients were asked to respond to the Swedish version of the 11-item QuickDASH [5].
The QuickDASH is a region-specific self-report outcome instrument quantifying physical function and symptoms in persons with any or multiple musculoskeletal conditions of the upper limb [26, 27]. Each item in QuickDASH scores on a 5-point scale (1 = No difficulty to 5 = Unable). Patients were requested to answer the questions based on their condition in the last week and regardless of which hand they used to perform the task. If they did not have the opportunity to perform an activity in the past week, they were asked to make their best estimate of which response would be the most accurate. A total QuickDASH score ranging from 0 (no disability) to 100 (most severe disability) was calculated from the item score [5].

Data analysis

Two aspects of validity (content and construct validity) was investigated. Content validity was assessed by patients, an inter-professional group of experts in the field of hand surgery and via literature review. Construct validity was investigated through predefined hypothesis and correlation statistics (Spearman’s rank correlation coefficient, $r_s$). The strength of the correlations was interpreted as: $r_s < 0.5$ low; 0.5 to $< 0.7$ moderate; $> 0.7$ high [28]. The analyses were made for the total group and in a selection of hand surgery diagnoses.

It was hypothesized that;

scores from question 8 in HQ–8 (ability to perform ADL) would correlate positively with the total QuickDASH score for the total group and in a selection of hand surgery diagnoses. A point estimate of the correlation of 0.70 or greater is considered a high correlation [8].

scores from questions 1–3 in HQ–8 (pain on load, pain on motion without load, pain at rest) would correlate positively with the total QuickDASH score for the total group and that scores from pain on load would show the largest correlation coefficient with a point estimate of the correlation of 0.70 or greater.

scores from question 5 in HQ–8 (weakness) would show a point estimate of the correlation of 0.70 or greater and a stronger positive correlation with the total QuickDASH score compared to questions 4, 6 and 7 (stiffness, numbness and cold sensitivity) for the total group.

Analyses were based on responded questionnaires for each follow-up. A high correspondence between the visual analogue scale has previously been established [29]. We therefore aggregated the data from the VAS and NRS responses in the present study.

Floor and ceiling effect was calculated as the percent of responded questionnaires with a score $<5$ (floor) and a score $>95$ (ceiling), respectively, for each of the HQ–8 questions at pre, 3 and 12 months follow-up. A threshold $>15\%$ was defined as a floor or ceiling effect [12]. Data completeness (item-
response) was calculated as the number of missing item responses for each HQ–8 question in relation to responded questionnaires. A threshold of >15% was defined as unacceptable [30].

The magnitude of change was quantified by effect size calculations. For elective hand-related diagnosis the mean paired change between pre- and 3 months follow-up as well as pre- and 12 months follow-up was divided by the dispersion measure (SD) of the preoperative (baseline) score in each HQ–8 question. For acute hand-related diagnosis the mean paired change between 3- and 12 months follow-up was divided by the dispersion measure (SD) of the 3-month score [31, 32]. According to Cohen’s criteria, an effect size of 0.20 is considered small, 0.50 is medium, and 0.80 is large [33]. Data was analysed using Microsoft Excel 2013 (Microsoft Cooperation, Redmond, Washington, United States) and the SPSS software package version 23 (IBM Cooperation, Amonk, New York, United States).

Results
Sample characteristics
In total, data from 33 885 questionnaires was analysed; 13 197 before operation, 10 702 at three months, and 9 986 at 12 months after hand surgery. Mean age was 52 (14-99), and composed of 50.0% men and 50.0% women. Thirty-six percent of the questionnaires were completed in a web format and 64% in a pencil and paper format.

Content validity
Seven questions were initially included in the HQ-8 questionnaire. The inter-professional expert group added one question concerning discomfort/problems when exposed to cold, giving a total of eight included questions. Formulations of four symptoms were changed from the initial version in order to achieve unambiguity and consistency with formulation of verbal anchors for all questions (0 representing no problem and 100 the worst problem imaginable). Reduced range of motion was therefore changed to stiffness, reduced strength to weakness, reduced sensibility or numbness to numbness/tingling in fingers and hand function in daily activities to ability to perform daily activities. The final version was achieved through consensus discussion in the expert group taking into consideration patients’ suggestions for clarification of symptoms and anchors, clinical experiences of the spectrum of symptoms and literature review [6]. See the online Additional file 1, HAKIR questionnaire, HQ-8. In order to reduce respondent burden and increase response rates, the number of included items was restricted, excluding questions concerning fine-motor skills, grip function and aesthetics.
Construct validity

The correlation coefficients between the single-item questions in HQ-8 and the total score in QuickDASH for the total group at pre, 3 and 12 months follow-up ranged from 0.44-0.79. The strongest correlation ($r_s=0.79$) in the total group was noted for question 8 (ability to perform daily activities) and the total QuickDASH score at the 3 month follow-up. The weakest correlation ($r_s=0.44$) was seen for question 7 (cold sensitivity) and total QuickDASH score in the pre-operative analysis. Detailed information for total group and a selection of diagnostic subgroups is available in Fig 1, Fig 2, Fig 3 and Fig 4 and in the online Additional file 2, Supplementary Table S1- S8.

Hypothesis 1: A high positive correlation (>0.70) was seen in the total group for the question concerning the ability to perform daily activities (#8) and the total QuickDASH score (0.72, 0.79, 0.76) for pre, 3 and 12 months follow up, respectively. For the subgroups thumb osteoarthritis, Morbus de Quervain, trigger finger and carpal tunnel syndrome the expected level > 0.70 was reached at the 3 and 12 months but not preoperatively. The $r_s$ values for Dupuytren´s contracture did not reach a coefficient over 0.70 in the analyses at either of the occasions. Fig 1, Fig 2, Fig 3 and Fig 4 and see the online Additional file 2, Supplementary Table S8.

Hypothesis 2: High positive correlations were noted in the total group for pain on load, pain on motion without load, pain at rest (#1-3) and the total QuickDASH score in the pre, 3 and 12 months analyses. Pain on load (#1) showed the strongest correlation at 3 and 12 months follow-up ($r_s = 0.76, 0.74$ respectively). However, a weaker $r_s (0.43-0.47)$ was noted preoperatively for all pain related questions and total QuickDASH scores for the subgroup thumb osteoarthritis. Fig 1, Fig 4 and see the online Additional file 2, Supplementary Table S1- S3.

Hypothesis 3: A stronger positive correlation in the total group was seen for weakness (#5) and the total QuickDASH score for pre, 3 and 12 months follow up respectively ($r_s = 0.67, 0.73, 0.74$), compared to $r_s$ values for stiffness (#4) ($r_s = 0.50, 0.58, 0.58$), numbness (#6) ($r_s = 0.52, 0.47, 0.54$) and cold sensitivity (#7) (range $r_s = 0.44, 0.46, 0.52$). Fig 1 and in the online Additional file 2, Supplementary Table S4 - S7. For the thumb osteoarthritis group an $r_s$ value of (0.36, 0.63, 0.70) was seen for weakness (#5) and total QuickDASH score preoperatively and at 3 and 12 month follow-up. Fig 4 and see the online Additional file 2, Supplementary Table S5.

Floor and ceiling effect

No ceiling effect (score >95) was noted in any of the HQ-8 questions or at pre, 3 and 12 months
follow-up in the total group or in the selection of diagnostic subgroups. A floor effect (score < 5) was seen in all HQ-8 questions and follow-ups in the total group. Table 1. Differences related to HQ-8 questions in a selection of subgroups are presented in the online Additional file 2, Supplementary Table S1-S8.

Table 1. Floor and ceiling effect. Percent of responded questionnaires, score <5 (floor) and score >95 (ceiling) for the HQ-8 questions at pre, 3 and 12 months follow-up (0 representing no problem and 100 worst problem imaginable)

| Category                  | #1     | #2     | #3     | #4     | #5     | #6     | #7     | #8  |
|---------------------------|--------|--------|--------|--------|--------|--------|--------|-----|
| <5                        | 21%    | 31%    | 38%    | 27%    | 24%    | 43%    | 45%    | 19% |
| >95                       | 5%     | 1%     | 1%     | 4%     | 4%     | 4%     | 3%     | 5%  |

| Category                  | #1     | #2     | #3     | #4     | #5     | #6     | #7     | #8  |
|---------------------------|--------|--------|--------|--------|--------|--------|--------|-----|
| <5                        | 23%    | 42%    | 57%    | 24%    | 23%    | 51%    | 48%    | 33% |
| >95                       | 1%     | 0%     | 0%     | 2%     | 2%     | 2%     | 3%     | 1%  |

| Category                  | #1     | #2     | #3     | #4     | #5     | #6     | #7     | #8  |
|---------------------------|--------|--------|--------|--------|--------|--------|--------|-----|
| <5                        | 34%    | 53%    | 62%    | 34%    | 32%    | 53%    | 46%    | 43% |
| >95                       | 1%     | 0%     | 0%     | 2%     | 2%     | 2%     | 4%     | 1%  |

Data completeness
The percentage of missing item responses in the responded questionnaires for all HQ-8 questions was < 2.6%, except for the question concerning cold sensitivity (#7). Table 2. Detailed information for the selection of diagnostic subgroups is available in the online Additional file 2, Supplementary Table S1-S8.

Table 2. Descriptive statistics for item-response in individual HQ-8 questions (0 representing no problem and 100 worst problem imaginable). Number and percent of missing item responses for each HQ-8 question (%) in total group at pre, 3 and 12 month follow up.

| Category      | #1     | #2     | #3     | #4     | #5     | #6     | #7     | #8  |
|---------------|--------|--------|--------|--------|--------|--------|--------|-----|
| Number*       | 13197  | 13197  | 13197  | 13197  | 13197  | 13197  | 13197  | 13197|
| Missing       | 244    | 235    | 247    | 328    | 342    | 312    | 813    | 206 |
| Missing %     | 1.8%   | 1.8%   | 1.9%   | 2.5%   | 2.6%   | 2.4%   | 6.2%   | 1.6% |

| Category      | #1     | #2     | #3     | #4     | #5     | #6     | #7     | #8  |
|---------------|--------|--------|--------|--------|--------|--------|--------|-----|
| Number*       | 10702  | 10702  | 10702  | 10702  | 10702  | 10702  | 10702  | 10702|
| Missing       | 234    | 181    | 175    | 174    | 202    | 210    | 811    | 207 |
| Missing %     | 2.2%   | 1.7%   | 1.6%   | 1.6%   | 1.9%   | 2.0%   | 7.6%   | 1.9% |

| Category      | #1     | #2     | #3     | #4     | #5     | #6     | #7     | #8  |
|---------------|--------|--------|--------|--------|--------|--------|--------|-----|
| Number*       | 9986   | 9986   | 9986   | 9986   | 9986   | 9986   | 9986   | 9986|
| Missing       | 188    | 151    | 158    | 158    | 179    | 180    | 285    | 170 |
| Missing %     | 1.9%   | 1.5%   | 1.6%   | 1.6%   | 1.8%   | 1.8%   | 2.9%   | 1.7% |

* = Total sample.

Magnitude of change
A varied pattern in effect size (ES, small-large) for the HQ-8 questions 1-8 was seen for subgroups of
elective hand-related diagnosis. The text below focus mainly on medium-large ES. Detailed information is given in the online Additional file 3, Supplementary table S9-15 and an overview of ES (small-large) in Table 3.

Thumb osteoarthritis: A medium ES was seen in all pain questions (#1-3) and ability to perform daily activities (#8) at 3 months follow-up, which increased to a large ES at the 12 months follow-up. Stiffness (#4) and weakness (#5) reached a medium ES at 12 months follow-up.

Dupuytren’s contracture: A medium ES was seen for stiffness (#5) and ability to perform daily activities (#8) at both follow-ups.

Morbus de Quervain: A medium ES was seen for pain on load (#1) and ability to perform daily activities (#8) at 3 months follow-up, which increased to a large ES for pain on load (#1) at 12 months follow-up. At 12 months follow-up, pain on motion without load, weakness and ability to perform daily activities (#2, 5, 8) showed a medium ES.

Ganglion: A medium ES was seen for pain on load (#1) at 12 month follow-up.

Trigger finger: A medium ES was seen for ability to perform daily activities (#8) at 3 months follow up, which increased to a large ES at 12 months follow up. A medium ES at 12 months follow-up was also seen for pain on motion without load (#2), pain at rest (#3), stiffness (#4) and weakness (#5) and a large ES for pain on load (#1).

Carpal tunnel syndrome: A medium ES was seen for pain at rest (#3) and numbness/tingling in fingers (#6) at 3 months follow-up, the latter increased to large ES at 12 months follow-up. A medium ES was seen for all other questions (# 1, 2, 3, 4, 5, 8) except cold sensitivity (#7) at 12 months follow-up.

Ulnar nerve entrapment: A small ES was seen for all questions at both follow-ups.

Table 3. Effect size at 3 and 12 months follow-up in subgroups of elective hand-related diagnosis. Detailed information is given in the online Additional file 3, Supplementary table S9-15.
| HQ-8                  | Thumb osteoarthitis | Dupuytr en’s contract ure | Morbus de Quervain | Ganglion | Trigger finger | Carpal tunnel syndrome |
|----------------------|---------------------|---------------------------|--------------------|----------|----------------|------------------------|
| 3 / 12 months follow up | 3 / 12 months follow up | 3 / 12 months follow up | 3 / 12 months follow up | 3 / 12 months follow up | 3 / 12 months follow up |
| 1 Pain on load       | M / L               | S / S                     | M / L              | S / M    | S / L          | S / M                  |
| 2 Pain on motion without load | M / L               | S / S                     | S / M              | S / S    | S / M          | S / M                  |
| 3 Pain at rest       | M / L               | S / S                     | S / S              | S / S    | S / M          | M / M                  |
| 4 Stiffness          | S / M               | M / M                     | S / S              | S / S    | S / M          | M / M                  |
| 5 Weakness           | S / M               | S / S                     | S / M              | S / S    | S / M          | S / M                  |
| 6 Numbness/tingling in fingers | S / S               | S / S                     | S / S              | S / S    | S / S          | M / L                  |
| 7 Cold sensitivity   | S / S               | S / S                     | S / S              | S / S    | S / S          | S / S                  |
| 8 Ability to perform daily activities | M / L               | M / M                     | M / M              | S / S    | M / L          | S / M                  |

Mean paired change between pre- and 3 months follow-up/SD on pre-months scores and mean change between pre- and 12 months follow-up/SD on the pre-months scores expressed as small (S), medium (M) or large (L) effect size (ES). According to Cohen’s criteria, an effect size of 0.20 is considered small, 0.50 is medium, and 0.80 is large [33].

For acute hand related diagnoses (digital nerve injuries, extensor and flexor tendon injuries, thumb ligament injuries and finger fractures), the mean paired change between 3 and 12 months follow-up was divided by SD on the 3 months scores. The effect size for all diagnoses and HQ-8 questions was small. See the online Additional file 4, Supplementary table S16-20.

**Discussion**

This study provides evidence of the validity of the HQ–8 questionnaire for the evaluation of patient reported outcomes in a registry-based setting. To include both single-item questions (HQ–8) and the QuickDASH provides the opportunity to evaluate both separate symptoms and overall self-reported disability. The HQ–8 questions mainly focus on impairments on the level of body function in the affected/operated hand, while most items in the QuickDASH evaluate the overall ability to perform
certain activities, regardless of which hand/arm that is used [34]. The additional questions about stiffness, weakness and cold sensitivity, as well as the specified questions about different aspects of pain, which all are included in the HQ-8, provide important information to describe a more complete picture of patients’ symptoms and the indications for surgery. Since the occurrence and importance of different symptoms varies between different hand related diagnoses e.g. a nerve injury, fracture or Dupuytren’s contracture, a conscious choice was to present the included HQ-8 items individually and not as a summed score. The collected information about individual symptoms can also give better and valid guidance to patients on the expected results after a surgical procedure. The main goal for a quality registry is to improve the quality of care by describing treatment results for a large number of patients. This data can be used for local improvement work and for clinical research based on large patient materials. By investigating patient reported outcomes, more targeted interventions can be offered to our patients. Patients with diagnoses such as Dupuytren’s contracture, in which pain is an uncommon symptom, usually score within normal population values for QuickDASH [18] which makes this instrument less valuable for evaluating treatment results in these patients. Question #4 in the HQ-8 regarding stiffness and #8 (ability to perform daily activities) may here serve as more accurate and sensitive outcome measures. The results of the effect size calculations (medium ES) also points toward the relevance of stiffness (#4) and ability to perform daily activities (#8) with the affected hand in HQ-8.

The content validity of the HQ-8 was established by taking into consideration patients’ comments and via literature review. The choice of items and verbal description of anchors was finally reached through consensus by experts who represented different health-care professionals including hand surgeons. Ratings of fine-motor skills, grip function and aesthetics were excluded to reduce the risk of respondent burden, despite being considered as valid aspects of hand function.

The use of an eleven increment NRS question has been recommended as being a responsive scale with good compliance rate, ease of use and similar psychometric properties compared to a visual analogue [29, 35–38]. Issues to consider in use of single-item questions are the definition of anchors. This has been commented on for pain ratings [39] but may be true even for ratings of other
impairments related to hand function. “Worst pain ever experienced” is another upper anchor alternative but is only interpretable when having knowledge about patient’s pain history [39]. The idea of accuracy when measuring pain has also been pointed out since an unequivocal reference standard does not exist or cannot be obtained [40]. Self-reports of pain can be influenced by previous experiences, behavioural, affective or cognitive factors and vary depending on the context in which the pain is experienced. When measuring change over time it is therefore of importance to use mean paired comparisons to limit the variability between person to person. Multidimensional pain scales have been recommended offering a broader understanding of pain experience beyond the simple factor of pain intensity as measured with single questions such as a NRS [41] but are not applicable in a total population of hand surgery patients. To capture different levels of pain intensity, ranking of both pain on load, on motion without load and at rest was included in HQ–8. Furthermore, it is also possible to include other additional self-report instruments specific for different hand-related diagnosis in research projects.

Construct validity was good showing a range of correlation coefficients (r_s 0.29—0.85) when analysed in the selected subgroups of hand-related diagnosis. A weaker r_s for e.g. stiffness, weakness and cold sensitivity in relation to QuickDASH may indicate clinical outcome aspects that are not captured by the QuickDASH instrument. The addition of a symptom questionnaire such as the HQ–8 may therefore provide additional value.

A high expected level of positive correlation (>0.70) between #8 (ability to perform daily activities) and the total QuickDASH score for the total group and in all follow-ups was found (hypothesis 1). This was expected since the main focus on items included in QuickDASH reflects the ability to perform certain activities. Question # 8 reflects the ability to perform ADL with the affected/operated hand and QuickDASH regardless of which hand that is affected. This may be part of an explanation why a “perfect” correlation close to 1 is not present and confirms the relevance of both outcome measures. Strong positive correlations were also seen in the total group for all HQ–8 questions concerning pain and the total QuickDASH score (hypothesis 2). In our experience pain is one of the most limiting
factors for satisfactory performance on activity level. This is also confirmed in previous studies [42–44]. A weaker $r_s$ was noted preoperatively for all pain related questions and total QuickDASH scores for the subgroup thumb osteoarthritis. A possible explanation may be that patients with thumb osteoarthritis rank their problems with pain as more problematic compared to performance in activities preoperatively. Access to compensatory strategies may be reflected in a proportional lower QuickDASH score and the correlation may therefore be weaker. Postoperatively, the perceived pain improves and the scores on pain and QuickDASH are therefore reflected in stronger $r_s$ values.

A stronger positive correlation in the total group was seen for weakness and the total QuickDASH score compared to $r_s$ values for stiffness, numbness and cold sensitivity (hypothesis 3). Hand strength is previously demonstrated to correlate strongly to DASH score in a variety of hand related diagnosis [42, 44–46]. For the thumb osteoarthritis group an $r_s$ value of (0.36, 0.63, 0.70) was seen for weakness and total QuickDASH score at pre, 3 and 12 month follow-up, which may indicate that an increased strength after surgery is important for performance on activity level.

A limited content validity may be indicated if a floor and ceiling effect occurs since the ability to discriminate between subjects at the extremes of the scale is lost [9]. A floor effect, defined as a 15% threshold [12] was seen in the total group and in all questions (HQ–8) with the largest floor effect seen at 12-months follow-up. This however was not observed in all of the selected subgroups since certain symptoms are neither relevant before, nor after surgery for some diagnoses. It is also expected that treatments should decrease or completely diminish symptoms that are relevant for different subgroups.

Data completeness was very good with 1.5 - 2.6% of patients with missing item responses depending on item and follow-up. This surpassed the rough guidelines on the defined threshold at >15% [30]. One exception was cold sensitivity (#7) with a slightly higher percentage of missing item responses at pre and 3 months follow-up (6.2 and 7.6%). This may be explained by the timepoint when the question was answered since the relevance of responding may vary during different seasons. Although cold sensitivity has been described as a frequent problem in a mixed group of hand injuries
[47], it is not present for all hand surgery patients. This may influence the motivation to reply. Interesting to note is that the effect sizes in various elective subgroups was consistent with clinical experience. For instance, pain is the symptom causing disability for patients with thumb osteoarthritis and is therefore the main indication for surgery. A large ES at 12 months follow up for all included pain questions in HQ-8, as well as the question concerning ability to perform daily activities, was therefore as expected. Another example of expected ES is the improvement of numbness/tingling in fingers (#6) (ES –0.79) following a carpal tunnel release. The small ES following surgery for ulnar nerve entrapment indicates less good treatment results in this patient group. The small ES between 3 and 12 months follow-up for the selected acute hand-related injuries is also as expected. The main change in scores probably occurs within 3 months after the injury and when the rehabilitation usually is completed. The preoperative scores are retrospectively filled in and recall bias especially in an acute injury situation may exist. At this point we therefore choose to prioritize ES calculations on the mean paired change between 3 and 12 months. Although, the results of the ES calculations were consistent with clinical experience following surgery in the selected subgroups of hand related diagnosis, one has to remember that ES only measures the magnitude of change and not the clinical importance [48]. According to the COSMIN guidelines responsiveness refers to the validity of a change score [48] and it has been defined as the ability to detect minimal important change (MIC) over time, even if these changes are small. Furthermore, instruments should be able to distinguish MIC from measurement error [9]. We therefore plan to study these aspects in a future study when test-retest data is collected.

**Methodological considerations**

The response rate was 45% for 3 months and 47% for 12 months postoperative questionnaires. It has been reported that on-line surveys in general are less likely than paper surveys to receive high response rates [49]. This was initially also the case for the HAKIR surveys, but after improving web-functionality and sending out a reminder sms the response rate is now similar between web and paper. Response rates around 30-40% have commonly been reported for large on-line surveys [49].
The response rate for the HAKIR questionnaire could therefore be considered as acceptable, even though efforts should be made to increase it further.

To determine the accuracy of HQ-8 in another culture or country or when different language versions are used, a reassessment of the validity and reliability is advocated. Guidelines for cultural-adaptation process in other languages are then to be followed [50, 51].

**Conclusions**

This study provides evidence of the content and construct validity of the HQ-8, including good data completeness, expected floor effect, lack of ceiling effect and an ability to detect changes in scores over time. The different associations between HQ-8 and QuickDASH in different hand diagnoses indicate that the HQ-8 measures unique aspects of disability. Additional strengths of the HQ-8 are the specific focus on the affected hand and more a detailed reporting of symptoms that patients usually regard as important, such as pain, stiffness and cold sensitivity. In our opinion, the HQ-8 may serve as an important complement to the “broader” content in QuickDASH with its main focus on overall performance on activity level. Further studies are needed to evaluate test-retest reliability, responsiveness including clinical meaningful change of the HQ-8 questions in subgroups of hand diagnoses.

**List Of Abbreviations**

PROM—Patient Reported Outcome measures

QuickDASH—Short version of Disabilities of the Arm, Shoulder and Hand

HQ-8—Eight item questionnaire in HAKIR

NQR—National Quality Registry

ICD 10—International Statistical Classification of Diseases and Related Health Problems - Tenth Revision

VAS—Visual Analogue Scale

NRS—Numeric Rating Scale

ES—Effect Size

SD—Standard Deviation

Declarations
Ethics approval and consent to participate
The study was approved by the Regional Ethics Committee in Stockholm, Sweden (Reference number: 2017/2023–31).

Consent for publication
Not applicable.

Availability of data and materials
Public access to data on individual patients is restricted by the Swedish Authorities (Public Access to Information and Secrecy Act; http://www.government.se/information-material/2009/09/public-access-to-information-and-secrecy-act/) but data can be made available for researchers after a special review that includes approval of the research project by both an Ethics Committee and the authorities’ data safety committees. For HAKIR, the Central Personal Data Controller is the South General Hospital in Stockholm and the Data Protection Officer can be reached at GDPR.sodersjukhuset@sll.se. We have supplied all aggregated data in supplementary tables.

Competing interests
The authors report no conflicts of interest.

Funding
The HAKIR registry is funded by the Swedish Association of Local Authorities and Regions and the Swedish Ministry of Health and Social Affairs.

Authors contribution
MA and IKC were responsible for the design and planning of the study. KS performed the initial content validation process. EE, MÅ, IKC performed the statistical analysis and interpreted the data together with MA. IKC did the drafting of the manuscript with contributions from the other authors. All authors read and approved the final manuscript.

Acknowledgements
The authors want to thank Lars Hagberg, MD, PhD, Stockholm for invaluable professional contribution during the early development of HQ–8.

References
1. Hailer NP, Garellick G, Karrholm J. Uncemented and cemented primary total hip arthroplasty in the Swedish Hip Arthroplasty Register. Acta Orthop. 2010;81(1):34–41.

2. Robertsson O, Ranstam J, Sundberg M, A WD, Lidgren L. The Swedish Knee Arthroplasty Register: a review. Bone Joint Res. 2014;3(7):217–22.

3. Makela KT, Matilainen M, Pulkkinen P, Fenstad AM, Havelin LI, Engesaeter L, et al. Countrywise results of total hip replacement. An analysis of 438,733 hips based on the Nordic Arthroplasty Register Association database. Acta Orthop. 2014;85(2):107–16.

4. Robertsson O, Bizjajeva S, Fenstad AM, Furnes O, Lidgren L, Mehnert F, et al. Knee arthroplasty in Denmark, Norway and Sweden. A pilot study from the Nordic Arthroplasty Register Association. Acta Orthop. 2010;81(1):82–9.

5. Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. BMC Musculoskelet Disord. 2006;7:44.

6. Swedish national healthcare quality registry for hand surgery-HAKIR. Patient questionnaire HQ–8. https://hakir.se/wp-content/uploads/2018/11/HQ–8-questionnaire-English–181003-.pdf Accessed November 30, 2018.

7. Kennedy CA, Beaton DE, Smith P, Van Eerd D, Tang K, Inrig T, et al. Measurement properties of the QuickDASH (disabilities of the arm, shoulder and hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. Qual Life Res. 2013;22(9):2509–47.

8. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–45.

9. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.

10. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand
outcome measure in different regions of the upper extremity. J Hand Ther. 2001;14(2):128–46.
11.Streiner DL, G.R N. Health measurements scales: a practical guide to their development and use. 3 ed. Oxford: Oxford University Press; 2003.
12.McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293–307.
13.Arner M. Developing a national quality registry for hand surgery: challenges and opportunities. EFORT Open Rev. 2016;1(4):100–6.
14.Barbier O, Penta M, Thonnard JL. Outcome evaluation of the hand and wrist according to the International Classification of Functioning, Disability, and Health. Hand Clin. 2003;19(3):371–8, vii.
15.Bialocerkowski AE, Grimmer KA, Bain GI. Validity of the patient-focused wrist outcome instrument: do impairments represent functional ability? Hand Clin. 2003;19(3):449–55, ix.
16.Chemnitz A, Dahlin LB, Carlsson IK. Consequences and adaptation in daily life - patients’ experiences three decades after a nerve injury sustained in adolescence. BMC Musculoskelet Disord. 2013;14(1):252.
17.Elliot D, Harris SB. The assessment of flexor tendon function after primary tendon repair. Hand Clin. 2003;19(3):495–503.
18.Engstrand C, Krevers B, Nylander G, Kvist J. Hand function and quality of life before and after fasciectomy for Dupuytren contracture. J Hand Surg Am. 2014;39(7):1333–43 e2.
19.Hobby JL, Watts C, Elliot D. Validity and responsiveness of the patient evaluation measure as an outcome measure for carpal tunnel syndrome. J Hand Surg Br. 2005;30(4):350–4.
20.Jerosch-Herold C, Mason R, Chojnowski AJ. A qualitative study of the experiences and expectations of surgery in patients with carpal tunnel syndrome. J Hand Ther. 2008;21(1):54–61; quiz 2.
21.Schuind FA, Mouraux D, Robert C, Brassinne E, Remy P, Salvia P, et al. Functional and outcome evaluation of the hand and wrist. Hand Clin. 2003;19(3):361–9.
22.Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. J Adv Nurs. 2003;42(1):57–63.
23.Willis GB, Artino AR, Jr. What Do Our Respondents Think We’re Asking? Using Cognitive
Interviewing to Improve Medical Education Surveys. J Grad Med Educ. 2013;5(3):353–6.

24. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today. 2004;24(2):105–12.

25. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs. 2005;14(7):798–804.

26. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative G. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87(5):1038–46.

27. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med. 1996;29(6):602–8.

28. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69–71.

29. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage. 2011;41(6):1073–93.

30. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; 2011.

31. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989;27(3 Suppl):S178–89.

32. Stratford PW, Binkley FM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. Phys Ther. 1996;76(10):1109–23.

33. Cohen J. Statistical power analysis for the behaviour sciences. New York: Academic Press1977.

34. Coenen M, Kus S, Rudolf KD, Muller G, Berno S, Dereskewitz C, et al. Do patient-reported outcome measures capture functioning aspects and environmental factors important to individuals with injuries or disorders of the hand? J Hand Ther. 2013;26(4):332–42; quiz 42.

35. Aasheim T, Finsen V. The DASH and the QuickDASH instruments. Normative values in the general population in Norway. J Hand Surg Eur Vol. 2014;39(2):140–4.
36. Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Validity of four pain intensity rating scales. Pain. 2011;152(10):2399–404.

37. Flaherty SA. Pain measurement tools for clinical practice and research. AANA J. 1996;64(2):133–40.

38. Gagliese L, Weizblit N, Ellis W, Chan VW. The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. Pain. 2005;117(3):412–20.

39. Walton DM, Elliott JM, Salim S, Al-Nasri I. A reconceptualization of the pain numeric rating scale: Anchors and clinically important differences. J Hand Ther. 2018;31(2):179–83.

40. Feinstein A. Clinimetrics. New Haven: Yale University Press 1987.

41. Badalamente M, Coffelt L, Elfar J, Gaston G, Hammert W, Huang J, et al. Measurement scales in clinical research of the upper extremity, part 2: outcome measures in studies of the hand/wrist and shoulder/elbow. J Hand Surg Am. 2013;38(2):407–12.

42. Ali M, Brogren E, Wagner P, Atroshi I. Association Between Distal Radial Fracture Malunion and Patient-Reported Activity Limitations: A Long-Term Follow-up. J Bone Joint Surg Am. 2018;100(8):633–9.

43. Montero A, Mulero JF, Tornero C, Guitart J, Serrano M. Pain, disability and health-related quality of life in osteoarthritis-joint matters: an observational, multi-specialty trans-national follow-up study. Clin Rheumatol. 2016;35(9):2293–305.

44. Swart E, Nellans K, Rosenwasser M. The effects of pain, supination, and grip strength on patient-rated disability after operatively treated distal radius fractures. J Hand Surg Am. 2012;37(5):957–62.

45. MacDermid JC, Wessel J, Humphrey R, Ross D, Roth JH. Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. Osteoarthritis Cartilage. 2007;15(5):524–30.

46. Wilcke MK, Abbaszadegan H, Adolphson PY. Patient-perceived outcome after displaced distal radius fractures. A comparison between radiological parameters, objective physical variables, and the DASH score. J Hand Ther. 2007;20(4):290–8; quiz 9.

47. Craigen M, Kleinert JM, Crain GM, McCabe SJ. Patient and injury characteristics in the development of cold sensitivity of the hand: a prospective cohort study. J Hand Surg [Am]. 1999;24(1):8–15.
48. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22.

49. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? Assessment & Evaluation in Higher Education. 2008;33(3):301-14.

50. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25(24):3186-91.

51. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993;46(12):1417-32.

Figures
Figure 1
Spearman’s correlation coefficient between HQ-8 questions and total QuickDASH score for the total group, preoperatively and at 3 and 12 months follow-up. The main axis represents the correlation coefficient (rs 0-1).
Carpal tunnel syndrome

Figure 2

Spearman’s rank correlation coefficient between questions in HQ-8 and total QuickDASH score for the subgroup carpal tunnel syndrome, preoperatively and at 3 and 12 months follow-up. The main axis represents the correlation coefficient (rs 0-1).
Figure 3

Spearman’s rank correlation coefficient between questions in HQ-8 and total QuickDASH score for the subgroup Dupuytren’s contracture, preoperatively and at 3 and 12 months follow-up. The main axis represents the correlation coefficient (rs 0-1).
Figure 4

Spearman’s rank correlation coefficient between questions in HQ-8 and total QuickDASH score for the subgroup thumb osteoarthritis, preoperatively and at 3 and 12 months follow-up. The main axis represents the correlation coefficient (rs 0-1).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

Additional file 2. Supplementary Table S1-S8.docx
Additional file 3. Supplementary Table S9-S15 - Copy.docx
Additional file 4. Supplementary Table S16-S20.docx
Additional file 1. HAKIR questionnaire, HQ-8.pdf