

# The Otology Questionnaire Amsterdam: A generic patient-reported outcome measure about the severity and impact of ear complaints. Validation, reliability and responsiveness

Jeroen T. Kraak  | Tom F. van Dam | Lisette M. van Leeuwen  |  
Sophia E. Kramer  | Paul Merkus 

Otolaryngology-Head and Neck Surgery,  
Ear & Hearing and Amsterdam Public Health  
Research Institute, Amsterdam UMC,  
Vrije Universiteit Amsterdam, Amsterdam,  
Netherlands

## Correspondence

Jeroen Timo Kraak, Otolaryngology-Head  
and Neck Surgery, Ear & Hearing and  
Amsterdam Public Health Research Institute,  
Amsterdam UMC, Vrije Universiteit  
Amsterdam, de Boelelaan 1117, Amsterdam,  
Netherlands.  
Email: jt.kraak@amsterdamumc.nl

## Abstract

**Objective:** To examine the construct validity, reliability and responsiveness of the Otology Questionnaire Amsterdam (OQUA).

**Design:** Multicentre, longitudinal study in 2 separate cohorts of patients visiting an ENT surgeon via an online survey programme.

**Setting:** Tertiary ENT clinics.

**Participants:** Cohort 1 consisted of patients at their first visit at an ENT outpatient clinic with an ear complaint. Cohort 2 consisted of patients who underwent surgery, with a 3-month follow-up post-surgery.

**Main outcome measures:** *Construct validity:* Hypothesis testing, internal consistency and inter-item correlation. *Reliability:* Test-retest reliability. The construct approach was used for assessing responsiveness. Hypotheses were formulated based on the association between the OQUA and Glasgow Health Status Inventory (GHSI) or Global Rating Scale (GRS).

**Results:** *Construct validity:* The correlation between the individual items in the impact domain ranged from 0.424 to 0.737. Confirmatory factor analysis showed a good fit. As expected, the OQUA impact showed strong relationships with GHSI total and general scale.

*Reliability:* The test-retest reliability coefficient ranged from 0.541 to 0.838.

*Responsiveness:* All hypotheses were conformed. As expected, the change score of the OQUA showed good correlation between OQUA impact and GHSI and moderate correlation between the GRS and OQUA complaints.

**Conclusion:** The OQUA has 8 complaint domains (earache, pressure sensation, itching, tinnitus, hearing loss, ear discharge, loss of taste and dizziness) and 1 impact domain. Each domain results in one score of 0-100. The OQUA shows good results for construct validity, (test-retest) reliability and responsiveness, supporting the potential benefit for the patient with an ear complaint visiting the ENT surgeon. The extensive validation furthermore confirms a certified generic otology PROM with an impact and a complaints' part, to be used in different types of otologic interventions and patient groups.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Clinical Otolaryngology* published by John Wiley & Sons Ltd

## 1 | INTRODUCTION

Adult patients visiting an ENT surgeon can present with a variety of ear complaints ranging from hearing loss and tinnitus to dizziness. These complaints often occur simultaneously. A patient-reported outcome measure (PROM) captures the patient's perspective and can be useful to evaluate this mix of complaints. Recently, Viergever et al searched for all existing Otology questionnaires available in the literature through a systematic review. No less than 144 unique questionnaires were identified. Several PROMs for measuring ear complaints were found. However, these questionnaires were either symptom specific (like the Tinnitus Handicap Index or Dizziness Handicap Index)<sup>1,2</sup> or disease specific (like the Chronic-Otitis-Media Q-12 or Meniere's disease Patient-Oriented Severity Index).<sup>3,4</sup> No generic PROM is available for measuring the severity and impact of multiple ear complaints without diagnosis in adults visiting an ENT surgeon. Furthermore, since the use of multiple questionnaires seems unpractical in practice, a generic otology PROM that addresses all relevant types of ear complaints and their impact on quality of life is desirable in the outpatient clinic.

Therefore, the Otology Questionnaire Amsterdam (OQUA) was recently developed by Bruinewoud et al 2018,<sup>5</sup> to allow the assessment of the presence and severity of ear complaints and their impact on the quality of life of patients visiting an ENT surgeon. The OQUA was developed according to the COSMIN development guidelines and with the involvement of an expert panel, patients' interviews and extensive field-testing, including factor analysis.<sup>5</sup>

With a validated questionnaire capable to measuring complaints and impact for all different kinds of ear complaints in the daily ENT practice, it could potentially be used in future to assess benefit of surgery from a patient perspective.

In the present study, we further investigated the quality of the OQUA as an instrument to be used in the ENT practice. Our study focused on the following questions: Does the OQUA measure what it is intended to measure (validation)? Is it stable in the scores it provides (reliability)? And does it show changes when a change is expected (responsiveness) before and after an intervention?

Validity was defined as the degree to which an instrument truly measures the construct.<sup>6</sup> If no 'gold standard' is available, then construct validity can be tested instead of criterion validity.<sup>7</sup> The construct validity consists of two parts, namely examination of structural validity (factor analysis) and hypothesis testing. The ICC for construct validity was already calculated by Bruinewoud et al<sup>5</sup> and showed good correlation between the individual items. For multi-item measurement instruments based on a formative model that applies to the complaint items in our model, there are no well-known measurement theories (a framework for model of the instrument can either be reflective or formative).<sup>7</sup> Therefore, the development of multi-item instruments following a formative model is merely based on common sense of experts in the field rather than on statistical procedures.<sup>5,7</sup>

Reliability was defined as 'the degree to which the measurement is free from measurement error'.<sup>8</sup> Mokkink et al (2010)<sup>9</sup> provided

### Keypoints

- A validated generic otologic questionnaire covering all eight relevant ear complaints and their impact.
- An effective tool to evaluate the effectiveness of surgical and non-surgical otological interventions.
- Pre- and post-intervention scores can be displayed graphically.

an extended definition of reliability: the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions, for example using different sets of items from the same multi-item measurement instrument (internal consistency) and over time (test-retest). Internal consistency of the OQUA was already tested by Bruinewoud et al (2018).<sup>5</sup>

The aim of this study was to investigate construct validity, test-retest reliability and responsiveness of the generic Otology Questionnaire Amsterdam. The OQUA was expected to show good construct validity, good test-retest reliability and if compared to a similar questionnaire like the Glasgow Health Status Inventory (GHSI) or a Global Rating Scale (GRS) to show good responsiveness.

## 2 | MATERIALS AND METHODS

The evaluation of OQUA's psychometric properties was performed following 'a practical guide to biostatistics and epidemiology for measurements in medicine' of de Vet et al 2010<sup>6</sup> and the COSMIN checklist.<sup>10</sup> The translated final version of the OQUA can be found in the appendix as supplement A; the original language is Dutch. All analysis were done using IBM SPSS Statistics version 22.

### 2.1 | Ethical considerations

The study protocol, including the development of the OQUA, was assessed by the Medical Ethics Review Committee of Amsterdam UMC, VU University Medical Centre, which led to the decision that the Medical Research Involving Human Subjects Act did not apply to this study. Written informed consent was obtained prior to study participation. All data from submitted questionnaires was completely anonymised. Data collection of cohort one was carried out between October 2016 and December 2016 and for cohort two between February 2017 and August 2018.

### 2.2 | Participants

For the assessment of construct validity and test-retest reliability, a consecutive cohort of 194 patients visiting the ENT surgeons

were recruited at Amsterdam UMC, location VUmc, Amsterdam, and at the Amstelland Hospital, Amstelveen, the Netherlands. For the assessment of responsiveness, a second prospective cohort of 50 consecutive patients undergoing ear surgery was recruited one week before their scheduled surgery. Patients were eligible to participate if they were 18 years and over, had an ear complaint, were not known to have a learning disability or cognitive impairment and had a good written understanding of the Dutch language.

### 2.3 | Procedures and measurements

All patients were asked to complete the questionnaires online. Patients in the first cohort were asked to complete the OQUA and GHSI (see below). After 6-14 days, patients were again asked to complete both questionnaires, assuming that this time interval was sufficient to minimise recall bias, yet short enough for their ear complaints and impact to remain unchanged. Patients in the second cohort were asked to complete the OQUA and GSHI approximately 1 week before their surgery (T0) and again after 3 months after surgery, including two GRSs (T1).

The OQUA consists of eight types of ear complaints domains and one impact domain in a 34-item questionnaire. Every complaint domain consists of a minimal of two types of items: one item for severity of the complaint measured with a Visual Analogue Scale (VAS; range 0-100) and one or more item(s) about the frequency of a specific aspect of the complaint measured with a five-point Likert response scale, ranging from 'almost never' (0p) to 'almost always' (5p). For both domains, a higher score indicates more complaints or higher impact.

The GHSI contains 18 items and measures the effect of health problem on the quality of life of a person.<sup>11</sup> The response to each question is based on a five-point Likert scale ranging from high health status through to low health status. A total score and 3 different subscales can be calculated (general subscale, social support score and physical health score). All scores range from 0-100. A higher score means less impact and better health status.<sup>11</sup>

Two 'GRS questions' were administered at 3-month follow-up. Patients were asked about the change in ear complaints before and after surgery and the change of impact of the complaints on their quality of live before and after surgery. Response options were on a five-point Likert scale ranging from prominent increase of complaints to great improvement.

### 2.4 | Statistical analyses

We produced descriptive statistics for the scores of the measurements. No missing data were expected in both cohorts since the online questionnaire could only be submitted when completed.

### 2.5 | Construct validity

Part of construct validity is inter-item correlation (ICC). In this study, construct validity was furthermore assessed by confirmatory factor analysis and hypothesis testing. Fit parameters were used to determine whether the data fitted the hypothesised factor structure. To evaluate model fit, the comparative fit index (CFI), Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA) were calculated.<sup>12</sup> Guidelines, proposed by Hu and Bentler,<sup>12</sup> suggest that models with CFI and TLI close to 0.95 or higher and RMSEA close to 0.06 or lower are representative of good fitting models.

For hypotheses testing, correlation hypotheses were computed comparing the sumscore of the impact items of the OQUA with the sumscore GHSI and its subscales.<sup>13</sup> Spearman's rho correlations were used for assessing the hypothesised relations because scores were non-normally distributed. Correlation was considered as low <.30; moderate .30-.59; and high  $\geq$ .60.<sup>14</sup>

### 2.6 | Test-retest reliability

Patients with active ear infections or who had complaints which were expected to change within 1-2 weeks after their visit at the ENT surgeon were regarded as unstable patients and were excluded from reliability analyses. To investigate the test-retest reliability of the OQUA, the quadratic weighted kappa was calculated for items measured with ordinal scales, and the intra-class correlation coefficient (ICC) was calculated for items measured with continuous scales using two-way ANOVA random effect models for agreement. An ICC value of 0.70, in a sample of 50 patients, was recommended as a minimum standard for reliability.<sup>15</sup> The kappa scores were interpreted through the Landis and Koch classification system; this system divides the score into 5 different classes, and the higher the score means better the reliability.<sup>16</sup>

### 2.7 | Responsiveness

Responsiveness was based on testing predefined hypotheses regarding the relationship between the OQUA, GSHI and GRS questions. We hypothesised good correlation between the change in OQUA impact domain and GSHI and less, but significant, correlation between change in OQUA complaints and GRS because multiple complaints had to be compared with one GRS. Pearson's correlation coefficients ( $r$ ) were used to determine the associations between the OQUA and the GHSI and GRS.

### 2.8 | Scoring system

To interpret the OQUA results, a scoring model had to be formulated. Because of the absence of a comparable scoring system of comparable questionnaire, one had to be developed. The scoring system was developed by an expert panel and authors of this article.

**TABLE 1** Demographic characteristics. T0 cohort (n = 194) underwent hypothesis testing, internal consistency and inter-item correlation. T1 cohort (n = 124) underwent test-retest and responsiveness (n = 47)

| Characteristics                        | Validity     |                | Reliability [test-retest] |                | Responsiveness |                |
|--|--------------|----------------|---------------------------|----------------|----------------|----------------|
|  | T0 (N = 194) | Percentage (%) | T1 (N = 124)              | Percentage (%) | T2 (N = 47)    | Percentage (%) |
| Age, y                                 |              |                |                           |                |                |                |
| Mean ± SD                              | 53 ± 16      | —              | 55 ± 15                   | —              | 52 ± 15        | —              |
| Range (min.-max.)                      | 18-82        | —              | 18-82                     | —              | 18-78          | —              |
| Sex                                    |              |                |                           |                |                |                |
| Female                                 | 94           | 48.5           | 57                        | 46.0           | 26             | 55.3           |
| Male                                   | 100          | 51.5           | 67                        | 54.0           | 21             | 44.7           |
| Hospital                               |              |                |                           |                |                |                |
| VU Medical Center                      | 180          | 92.8           | 114                       | 91.9           | 44             | 93.6           |
| Amstelland Hospital                    | 14           | 7.2            | 10                        | 8.1            | 3              | 6.4            |
| Type of surgery performed              |              |                |                           |                |                |                |
| Stapedotomy                            | N/A          | —              | N/A                       | —              | 9              | 19.2           |
| Cochlear implant                       | N/A          | —              | N/A                       | —              | 7              | 14.9           |
| Ossicular chain reconstruction         | N/A          | —              | N/A                       | —              | 9              | 19.2           |
| Cholesteatoma surgery                  | N/A          | —              | N/A                       | —              | 12             | 25.5           |
| Myringoplasty                          | N/A          | —              | N/A                       | —              | 5              | 10.6           |
| Other                                  | N/A          | —              | N/A                       | —              | 5              | 10.6           |
| Diagnosis                              |              |                |                           |                |                |                |
| Diagnosis was known by the patient     | 143          | 73.3           | 91                        | 73.4           | N/A            | —              |
| Diagnosis was unknown for the patient  | 20           | 10.3           | 14                        | 11.3           | N/A            | —              |
| No diagnosis was made (yet)            | 31           | 16             | 19                        | 15.3           | N/A            | —              |
| Distribution of diagnosis <sup>a</sup> |              |                |                           |                |                |                |
| Cholesteatoma                          | 14           | 9.8            | 10                        | 11             | N/A            | —              |
| Middle ear infection                   | 20           | 13.9           | 14                        | 15.4           | N/A            | —              |
| Otosclerosis                           | 16           | 11.2           | 10                        | 11             | N/A            | —              |
| Open cavity                            | 19           | 13.2           | 12                        | 13.2           | N/A            | —              |
| Hearing impaired                       | 69           | 48.2           | 46                        | 50.5           | N/A            | N/A            |
| Tinnitus                               | 45           | 31.4           | 31                        | 34             | N/A            | N/A            |
| Ear canal infection                    | 35           | 24.4           | 25                        | 2.7            | N/A            | N/A            |
| Excessive ear wax (cerumen)            | 15           | 10.4           | 11                        | 12.1           | N/A            | N/A            |
| Benign positional vertigo (BPPV)       | 9            | 6.3            | 7                         | 7.6            | N/A            | N/A            |
| Meniere's disease                      | 2            | 1.4            | 1                         | 1.1            | N/A            | N/A            |
| Neurinoma                              | 4            | 2.8            | 3                         | 3.2            | N/A            | N/A            |
| Ear canal stenosis                     | 6            | 4.2            | 4                         | 4.4            | N/A            | N/A            |
| Other                                  | 29           | 20.2           | 17                        | 18.6           | N/A            | N/A            |

<sup>a</sup>A patient can have multiple diagnosis.

Abbreviations: AA: Amino acid; NT: Nucleotide.

### 3 | RESULTS

#### 3.1 | Patient characteristics

In Table 1, patient characteristics of the 2 consecutive cohorts are presented. In the first cohort, 194 patients participated. With regard

to the follow-up measurement, 124 patients gave reply on our request to fill in the follow-up questionnaire (64%). No differences in demographics between the entire group of 194 patients and the subgroup of 124 were calculated. In the second cohort, 50 patients participated, but due to missing data, 3 patients had to be excluded. In total, 47 patients were eligible for responsiveness analyses.

## 3.2 | Construct validity

### 3.2.1 | Inter-item correlation

The inter-item correlation (ICC) matrix for construct was already performed by Bruinewoud et al 2018<sup>5</sup> and showed good correlation between the individual items. No items had a too high correlation with another item (>.9; ie most likely identical items) nor a too low correlation (<.2; ie no correlation; Table 2).

### 3.2.2 | Confirmatory factor analysis (CFA)

The CFA was performed using collapsed data; that is, answer categories containing less than 5% of the answers filled in were combined with the adjacent category. In items 1, 2, 3, 6 and 7, the answer categories 4 and 5 were combined (4; 'Agree', 5; 'Strongly agree'). In Item 4, answer categories 3, 4 and 5 were combined (3; 'Neutral', 4; 'Agree', 5; 'Strongly agree'). There were no residual correlations above 0.2, which indicates that there were no local dependencies.

### 3.2.3 | Hypothesis testing

All hypotheses were confirmed. As expected, moderate to high correlation coefficients were found between the impact items of the OQUA and the total score of the GHSI (0.672) and the general subscale of the GHSI (0.721). In addition, in line with expectations, weak to no correlation between the OQUA and the social (0.005) and physical subscale (0.177) of the GHSI was confirmed,

## 3.3 | Reliability

### 3.3.1 | Internal consistency

Cronbach's alpha for the impact items of the OQUA was .916, indicating strong evidence for a good internal consistency.

### 3.3.2 | Test-retest reliability

The mean interval between T0 and T1 was 8.69 days (range, 6-14 days). The quadratic weighted kappa's with associated 95% confidence interval bounds in the parentheses and standard error of measurements of the 26 items, 9 scored a very good (ak > 0.80) test-retest reliability. Sixteen items scored good (ak = 0.6-0.8), and only one item (hearing loss Q1; 'Can you hear someone approaching from behind?') scored moderate (ak < 0.6). Four of the items had a very good test-retest reliability (ICC > 0.8), and the other five had a good test-retest reliability (ICC = 0.6-0.8).

### 3.3.3 | Responsiveness

Change scores on both the OQUA impact items and GHSI total score were normally distributed. Responsiveness showed good correlation between OQUA impact and GHSI using Pearson's coefficient (-.711). This result matches our hypothesis. Correlation between the GRS and the complaints of the OQUA was calculated using the Spearman  $\rho$  and showed (.623).

Tables 3 and 4 show the responsiveness results in a subgroup analysis of cholesteatoma and otosclerosis patients. No other subgroup analysis was made due to small sample size. In the both subgroups, the preoperative response of the OQUA was spread over multiple complaints. The overall impact preoperative response was higher than postoperative response, but the change in complaints varies from patient to patient.

### 3.3.4 | Sequence

A considerable number of patients made a comment about the sequence of the questions. In the first version of the OQUA,<sup>5</sup> the sequence of the items was more randomly distributed to avoid duplication of answer categories to items that appeared to be alike. Keeping these questions close together could have interfered with the explorative factor analysis. An example of an item which

**TABLE 2** Inter-item correlation matrix showing good correlation between the individual items

|          | Impact_1 | Impact_2 | Impact_3 | Impact_4 | Impact_5 | Impact_6 | Impact_7 | Impact_8 | Impact_9 |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Impact_1 | 1.000    | .647     | .547     | .542     | .628     | .481     | .424     | .608     | .551     |
| Impact_2 | .647     | 1.000    | .499     | .641     | .513     | .431     | .491     | .533     | .633     |
| Impact_3 | .547     | .499     | 1.000    | .540     | .574     | .498     | .426     | .559     | .463     |
| Impact_4 | .542     | .641     | .540     | 1.000    | .468     | .467     | .494     | .492     | .600     |
| Impact_5 | .628     | .513     | .574     | .468     | 1.000    | .613     | .610     | .737     | .548     |
| Impact_6 | .481     | .431     | .498     | .467     | .613     | 1.000    | .700     | .695     | .492     |
| Impact_7 | .424     | .491     | .426     | .494     | .610     | .700     | 1.000    | .625     | .519     |
| Impact_8 | .608     | .533     | .559     | .492     | .737     | .695     | .625     | 1.000    | .635     |
| Impact_9 | .551     | .633     | .463     | .600     | .548     | .492     | .519     | .635     | 1.000    |

Note: >.9 means a too high correlation (ie most likely identical items); <.2 means a too low correlation (ie no correlation).

**TABLE 3** Responsiveness subgroup cholesteatoma: Pre- and post-intervention score impact and complaints

| patient | pre operative otalgia | post operative otalgia | pre operative pressure | post operative pressure | pre operative itch | post operative itch | pre operative tinnitus | post operative tinnitus | pre operative hearing loss | post operative hearing loss | pre operative otorrhoe | post operative otorrhoe | pre operative loss of taste | post operative loss of taste | pre operative dizziness | post operative dizziness | pre operative impact | post operative impact |
|---------|-----------------------|------------------------|------------------------|-------------------------|--------------------|---------------------|------------------------|-------------------------|----------------------------|-----------------------------|------------------------|-------------------------|-----------------------------|------------------------------|-------------------------|--------------------------|----------------------|-----------------------|
| #1      | 0                     | 0                      | 0,6                    | 0                       | 0,8                | <i>1,2</i>          | 0,2                    | 0                       | 0,2                        | <i>0,4</i>                  | 0                      | 0                       | 0                           | <i>77</i>                    | 0                       | 0                        | 20                   | <i>20</i>             |
| #2      | 21,6                  | 0                      | 33,6                   | 0                       | 14,4               | <i>0,2</i>          | 100                    | 40                      | 66,5                       | <i>4,3</i>                  | 19,6                   | 0                       | 0,2                         | 0                            | 20,4                    | <i>0,3</i>               | 33                   | <i>24</i>             |
| #3      | 2,4                   | <i>1,2</i>             | 4,5                    | <i>4,4</i>              | 33,6               | <i>4,8</i>          | 1,6                    | <i>0,6</i>              | 37,5                       | <i>13,3</i>                 | 1                      | 1                       | 38,4                        | <i>0,8</i>                   | 22,4                    | <i>16,5</i>              | 57                   | <i>48</i>             |
| #4      | 0                     | 0                      | 3                      | <i>0,2</i>              | 28,2               | <i>1,2</i>          | 0                      | 0                       | 23,2                       | <i>35,5</i>                 | 24,8                   | 0                       | 0                           | 0                            | 0                       | 0                        | 20                   | <i>53</i>             |
| #5      | 2,4                   | <i>2,8</i>             | 16                     | <i>18,9</i>             | 0,2                | <i>9</i>            | 0                      | 0                       | 14,4                       | <i>13,9</i>                 | 60                     | 0                       | 0,2                         | 0                            | 42,4                    | <i>6,7</i>               | 91                   | <i>26</i>             |
| #6      | 19,2                  | <i>0,2</i>             | 11,7                   | <i>0,2</i>              | 27,6               | 0                   | 67                     | <i>41</i>               | 16,9                       | <i>13,3</i>                 | 4,8                    | <i>0,6</i>              | 0,8                         | <i>0,2</i>                   | 4,7                     | <i>0,2</i>               | 66                   | <i>40</i>             |
| #7      | 32                    | <i>2,4</i>             | 46,9                   | <i>4,2</i>              | 11,2               | <i>10</i>           | 8                      | <i>0,2</i>              | 42,4                       | <i>3</i>                    | 37,2                   | <i>4,5</i>              | 0                           | 0                            | 1,2                     | 0                        | 42                   | <i>24</i>             |
| #8      | 14                    | <i>19,2</i>            | 32,4                   | <i>47,4</i>             | 10,4               | <i>28,2</i>         | 96                     | <i>80</i>               | 54,4                       | <i>39,9</i>                 | 3                      | <i>12,6</i>             | 2,6                         | <i>19,6</i>                  | 61,6                    | <i>32,4</i>              | 77                   | <i>55</i>             |
| #9      | 2                     | <i>1</i>               | 4                      | <i>0,2</i>              | 2,8                | <i>12</i>           | 0,2                    | 0                       | 3,3                        | <i>2,3</i>                  | 1,5                    | 0                       | 0                           | <i>40</i>                    | 0                       | 0                        | 31                   | <i>20</i>             |
| #10     | 0                     | 0                      | 2,1                    | <i>6</i>                | 0,2                | 0,2                 | 19,2                   | <i>63</i>               | 41,3                       | <i>54,6</i>                 | 0                      | <i>17</i>               | 0                           | <i>71</i>                    | 0,3                     | <i>0,2</i>               | 55                   | <i>66</i>             |
| #11     | 8,8                   | <i>22,8</i>            | 47,6                   | <i>32,8</i>             | 0                  | <i>9,2</i>          | 85                     | <i>80</i>               | 7,9                        | <i>46,8</i>                 | 0,2                    | <i>0,8</i>              | 0                           | <i>0,6</i>                   | 5,6                     | <i>16,4</i>              | 75                   | <i>48</i>             |
| #12     | 0                     | <i>0,6</i>             | 6                      | <i>49</i>               | 1                  | <i>0,8</i>          | 97                     | <i>98</i>               | 49,5                       | <i>50,7</i>                 | 0,4                    | <i>19</i>               | 0,6                         | <i>19</i>                    | 8,9                     | <i>8</i>                 | 62                   | <i>40</i>             |

Note: Numbers in green and bold show improvement, in red and italic shows deterioration and it remains black if unaltered.

**TABLE 4** Responsiveness subgroup cholesteatoma: Pre- and post-intervention score impact and complaints

| patient | pre operative otalgia | post operative otalgia | pre operative pressure | post operative pressure | pre operative itch | post operative itch | pre operative tinnitus | post operative tinnitus | pre operative hearing loss | post operative hearing loss | pre operative otorrhoe | post operative otorrhoe | pre operative loss of taste | post operative loss of taste | pre operative dizziness | post operative dizziness | pre operative impact | post operative impact |
|---------|-----------------------|------------------------|------------------------|-------------------------|--------------------|---------------------|------------------------|-------------------------|----------------------------|-----------------------------|------------------------|-------------------------|-----------------------------|------------------------------|-------------------------|--------------------------|----------------------|-----------------------|
| #1      | 2                     | <i>0,2</i>             | 71,1                   | <i>4,8</i>              | 0,2                | 0                   | 0,2                    | 0                       | 69,8                       | <i>73,7</i>                 | 0                      | 0                       | 10,8                        | 0                            | 0,2                     | 0                        | 62                   | <i>57</i>             |
| #2      | 0                     | <i>1,2</i>             | 0                      | <i>6,8</i>              | 0                  | <i>0,8</i>          | 80                     | <i>41</i>               | 41,4                       | <i>41,4</i>                 | 0                      | 0                       | 0                           | 0                            | 0                       | <i>0,8</i>               | 53                   | <i>64</i>             |
| #3      | 21                    | <i>2,4</i>             | 11,5                   | 0                       | 40                 | <i>18,4</i>         | 56                     | <i>43</i>               | 35,7                       | <i>12,7</i>                 | 2,4                    | 0                       | 0,4                         | <i>1,2</i>                   | 33,3                    | <i>1,4</i>               | 82                   | <i>33</i>             |
| #4      | 0                     | <i>0,2</i>             | 8,7                    | <i>10</i>               | 0,2                | <i>0,4</i>          | 34,2                   | <i>0,6</i>              | 32,5                       | <i>1,2</i>                  | 0                      | 0                       | 0,2                         | 0                            | 4                       | 0                        | 35                   | <i>20</i>             |
| #5      | 20                    | 0                      | 0                      | <i>0,3</i>              | 25,8               | <i>5,4</i>          | 1,6                    | <i>2,4</i>              | 66                         | <i>33,0</i>                 | 0                      | 0                       | 0                           | <i>80</i>                    | 0                       | 0                        | 68                   | <i>60</i>             |
| #6      | 1                     | <i>2,6</i>             | 1                      | <i>3</i>                | 0,8                | 0                   | 0,6                    | <i>15,2</i>             | 67,2                       | <i>34,0</i>                 | 0,6                    | <i>0,4</i>              | 0,6                         | <i>1,4</i>                   | 8,3                     | <i>1</i>                 | 71                   | <i>40</i>             |
| #7      | 10,2                  | 0                      | 0                      | 0                       | 0                  | 0                   | 100                    | <i>25,2</i>             | 36,4                       | 0                           | 0                      | 0                       | 0                           | 0                            | 16,8                    | <i>8</i>                 | 66                   | <i>20</i>             |
| #8      | 0                     | 0                      | 0                      | 0                       | 0                  | 0                   | 2                      | <i>0,2</i>              | 12,1                       | <i>4,7</i>                  | 0                      | 0                       | 0                           | 0                            | 0,3                     | <i>0,4</i>               | 35                   | <i>40</i>             |

Note: Numbers in green and bold show improvement, in red and italic shows deterioration and it remains black if unaltered.

appeared to be alike is 'I have ear discharge' compared to 'Liquid comes out of my ear'. After finishing the validation and item reduction, we have changed the sequence of the items in order to be more logical and bringing the items about the same domain together. In Figure 1, the latest version of the OQUA is presented.

### 3.3.5 | Scoring system

In order to implement the OQUA in daily practice, a scoring system was developed (Figure 2). The expert panel decided that every ear complaint in the complaint domain should be equally weighted and also that the frequency (measured with VAS) and severity (measured with 5-point Likert scales) in the ear complaint domain should be equally weighted. Additionally, every ear complaint domain (eg tinnitus, hearing loss) should have one score and impact scale should have a separate score. It was decided that a total complaint score covering all domains is not desirable and should not be calculated as it hides differences that may occur in the scoring of the separate domain. For example, the tinnitus score could decrease, while at the same time, the hearing score could increase. This would have resulted in an unchanged total score. Whether it is practically correct to mathematically balance the items and complaints, as done now, is an issue that needs further investigation

in future. In Figure 3, a graphical representation of how the OQUA result could look in the (electronic) patient file is shown.

## 4 | DISCUSSION

PROMs are essential for the assessment of subjective complaints and disability and are increasingly utilised. Still, in a general ENT practice, the use of otologic questionnaires is limited and it varies strongly between countries and clinics to which extent disease- or complaint-specific questionnaires are used. The OQUA is a newly developed generic otologic questionnaire, designed to measure the severity of ear complaints, the impact of these complaints on daily life and to assess the effectiveness of an intervention from a patients' perspective. It could be useful to avoid all kinds of questionnaires in a single practise and to compare outcomes in a larger group.

In this study, we analysed the validity of the OQUA. Good validity, reliability and responsiveness were confirmed. The validated version of the OQUA consists of 34 items and was slightly changed in its original sequence for practical reasons. It now has a scoring manual and is ready for broader use. This study shows that the OQUA is set for implementation as standard care in ENT practise in an increasing number of countries.



Complaints:

(Almost) never    Sometimes    Regularly    Often    (Almost) always

1. I have an earache

2. Indicate the severity of your earache on the line below.

(Almost) never    Sometimes    Regularly    Often    (Almost) always

3. I feel pressure in my ear.

4. My ear pops.

5. Indicate the severity of pressure in your ear on the line below.

(Almost) never    Sometimes    Regularly    Often    (Almost) always

6. I have an itch in or on my ear.

7. Indicate the severity of itching in or on your ear to the line below.

(Almost) never    Sometimes    Regularly    Often    (Almost) always

8. I hear a hum, murmur, beeping noise or buzzing sound.

9. Indicate the severity of your tinnitus (this can be a hum, murmur, beeping noise or buzzing sound) on the line below.

(Almost) never    Sometimes    Regularly    Often    (Almost) always

10. Can you hear somebody approaching from behind?

11. Can you hear cars passing by?

12. Can you hear from what corner of a room someone is talking to you being in a quiet house?

13. Can you understand the presenter of the news on TV at a normal volume?

14. Can you follow a conversation between a few people during dinner?

15. I am sensitive to loud noises.

16. Indicate the severity of your hearing loss on the line below.

FIGURE 1 (Continues)

|  | (Almost)<br>never | Sometimes | Regularly | Often | (Almost)<br>always |
|--|-------------------|-----------|-----------|-------|--------------------|
| 17. Liquid comes out of my ear.                                    |                   |           |           |       |                    |
| 18. Pus comes out of my ear.                                       |                   |           |           |       |                    |
| 19. Indicate the severity of your ear discharge on the line below. |                   |           |           |       |                    |
|  | (Almost)<br>never | Sometimes | Regularly | Often | (Almost)<br>always |
| 20. I have a poor sense of taste                                   |                   |           |           |       |                    |
| 21. Indicate the severity of your loss of taste on the line below. |                   |           |           |       |                    |
|  | (Almost)<br>never | Sometimes | Regularly | Often | (Almost)<br>always |
| 22. I have balance problems.                                       |                   |           |           |       |                    |
| 23. I feel dizzy   |                   |           |           |       |                    |
| 24. When I move my head I get dizzy.                               |                   |           |           |       |                    |
| 25. Indicate the severity of your dizziness on the line below.     |                   |           |           |       |                    |

**Impact:**

|   | Strongly<br>disagree | Disagree | Neutral | Agree | Strongly<br>agree |
|---|----------------------|----------|---------|-------|-------------------|
| 26. I get irritated due to my ear problems.   |                      |          |         |       |                   |
| 27. I get upset due to my ear problems.   |                      |          |         |       |                   |
| 28. I have impaired concentration due to my ear problems.   |                      |          |         |       |                   |
| 29. I feel depressed due to my ear problems.  |                      |          |         |       |                   |
| 30. My ear problems are very tiring.  |                      |          |         |       |                   |
| 31. My ability to take part in social activities (hobbies, sport or leisure-time activities) is limited due to my ear problems. |                      |          |         |       |                   |
| 32. I have had to modify my daily activities and/or work due to my ear problems.  |                      |          |         |       |                   |
| 33. My ear problems make life difficult for me.   |                      |          |         |       |                   |
| 34. I am concerned about my ear problems.   |                      |          |         |       |                   |

**FIGURE 1** OQUA questionnaire (English translation, Original language Dutch). The questionnaire contains 34 items of which 9 are about impact and 25 about the 8 most prominent ear complaints (earache, pressure sensation, itching, tinnitus, hearing loss, ear discharge, loss of taste, dizziness). Each complaint has one question in VAS scale to score the severity of the complaint. 18 questions with a 5-likert scale answer category. Scoring system is explained in Figure 2.

#### 4.1 | Strength and weakness of the study

The design of the OQUA is based on the COSMIN guidelines.<sup>10</sup> All required steps for the development of new instruments were taken, and the OQUA was extensively field-tested with sufficient number of patients. The study population represented an heterogeneous

group by including patients of different ages, sexes, complaints and pathologies. Furthermore, the cohort in which the responsiveness of the OQUA was assessed had an additional heterogeneous distribution of surgical procedures.

Although the results confirm our hypotheses and show satisfactory results for construct validity, test-retest reliability and responsiveness,



### Scoring total OQUA

There are two separate scores in the OQUA, there is no total score for the OQUA. Still the Impact score can be used separately and the complaint domains can be used separately but have to be kept/ presented together to be valid. The developers advise not to calculate a total score, as this does not give a valid overview of complaints.

### Scoring Complaint domains

Calculations are made to present a final number from 0 to 100 for each domain. Each domain is separated in a frequency (5 point likert questions) and severity (VAS) except the impact domain.

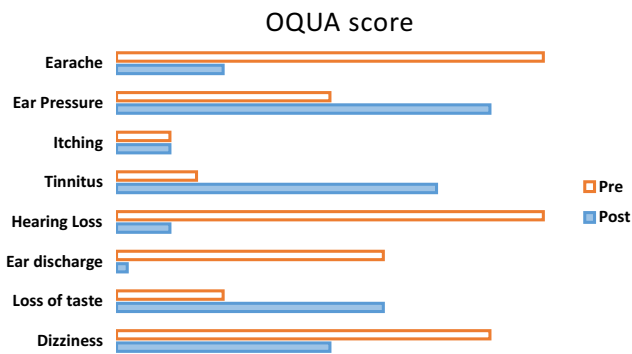
Depending on the number of questions of each domain the Likert questions will be calculated to a maximum total of 10 and the VAS divided to maximum total of 10. The product of these two (frequency x severity) being the final domain score (per complaint) ranging from 0-100.

1. Earache: divide the VAS score (Q2) by 10 and then multiply by the answer score of question 1 (Q1) multiplied with 2. Earache score =  $[(Q2/10)*(Q1*2)]$
2. Ear Pressure sensation: divide the VAS score (Q5) by 10 and multiply by the average score of Q3 -4 and then multiply with 2. Ear pressure=  $[(Q5/10)*((Q3+Q4)/2*2)]$
3. Itching: divide the VAS score (Q7) by 10 and then multiply by the answer score of Q6 multiplied with 2. Itching score =  $[(Q7/10)*(Q6*2)]$
4. Tinnitus: divide the VAS score (Q9) by 10 and then multiply by the answer score of Q8 multiplied with 2. Tinnitus score =  $[(Q9/10)*(Q8*2)]$
5. Hearing loss: divide the VAS score (Q16) by 10 and multiply by the average score of Q10-15 and then multiply with 2. Hearing score=  $[(Q16/10)*((Q10+Q11+Q12+Q13+Q14+Q15)/6*2)]$
6. Ear discharge: divide the VAS score (Q19) by 10 and multiply by the average score of Q17-18 and then multiply with 2. Ear discharge=  $[(Q19/10)*((Q17+Q18)/2*2)]$
7. Loss of Taste: divide the VAS score (Q21) by 10 and then multiply by the answer score of Q20 multiplied with 2. Loss of taste score =  $[(Q21/10)*(Q20*2)]$
8. Dizziness: divide the VAS score (Q25) by 10 and multiply by the average score of Q22-24 and then multiply with 2. Dizziness=  $[(Q25/10)*((Q22+Q23+Q24)/3*2)]$

### Scoring Impact domain

A maximum score of 100 can be achieved in the impact scale if answers are filled out in the most unfavourable way. Meaning if a patient has a low impact score, the complaints have low impact on the patients daily life.

Calculation: Sum the nine scores of Q26-34, divide by 45 and multiply by 100 to get a 0-100 score.



**FIGURE 2** Scoring system of the OQUA

**FIGURE 3** Example of graphical representation of the OQUA

some limitations of the study should be mentioned. Unfortunately, there is no consensus in literature about the group size to evaluate validity. Fifty patients is advised by de Vet et al.<sup>7</sup> Accordingly, 50 patients were included in the cohort for examining the OQUA's responsiveness, but due to missing data, three patients had to be excluded. However, the remaining sample is thought to be a good representation of a general ENT practice and the number big enough to examine responsiveness. On the other hand, one could argue that this amount is critical for confirmation of responsiveness in each domain.

The GRS used to examine OQUA's complaints scores did not distinguish in the pre-determined separate domains (eg tinnitus, hearing loss). Rather, the GRS asked for a change in complaints overall. Using this method, complaints might have cancelled each other out in the correlation with GRS or downsized the effect. For example, at baseline

(pre-surgery), hearing loss is the most prominent complaint of patient x which improves in after surgery. In addition, the same patient might suffer from vertigo after surgery which was not present pre-surgery. The positive and negative result may have downsized the correlation in the responsiveness statistics. Furthermore, the patient might relate the GRS to his/hers most prominent complaint and not to the broader spectrum of complaints. Another limitation could be a partial floor effect; when patients fill-out rather low scores at the preoperative moment, there will be little to no room for improvement 3 months after surgery.

The decision to calculate separate domain scores and not a total complaint score covering all domains was made based on the fact that the complaints are all separate domains in a formative model and can be seen as different (sub)constructs. Meaning that the items of the different domains not necessarily correlate with each other as can

Impact pre: 35

Impact post: 22

been seen in the development paper of the OQUA by Bruinewoud et al 2018.<sup>5</sup> By using separate scores, a more accurate insight in the patients' scale of complaints is given. This could, on the other hand, also be seen as a disadvantage, as it can be challenging to compare the results of the complaints with other cross-sectional studies in literature. Many questionnaires often use a single score for the total of complaints or a score for the complaints and impact as a whole. Up till now, studies often used a complaint- or disease-specific questionnaire. Complaint-specific questionnaires can be compared with the OQUA-related domain, but will lack information on other ear complaints. Disease-specific questionnaires are difficult to compare with the OQUA as the questions in a disease-specific questionnaire will be tailored towards the specific group of patients with their specific complaints. These studies might be designed to evaluate response to treatment, but most likely will only focus on the patients' ear complaints specific for that disease and discarding the possible other ear complaints which can occur after ear surgery (eg tinnitus or taste problems). Disease-specific questionnaires should preferably be combined with a separate QoL questionnaire to assess the relevance and impact of these complaints on the quality of life. Comparing cross-sectional ear-related quality-of-life questionnaires with the OQUA will be possible as the OQUA has a separate scoring system on impact.

#### 4.2 | Current position of the OQUA

In the systematic review that was mentioned in the introduction, one generic otologic questionnaire, the COQOL,<sup>17</sup> was identified. The COQOL was developed with the aim to assess quality of life and impact of ear complaints of patients visiting an ENT surgeon. The OQUA distinguishes itself from the COQOL by having two separate scales/instead of one, measuring not only the impact on quality of life, but also the severity of all major ear complaints as separate domains. In line with their objective to measure impact/quality of life, the COQOL validated its responsiveness compared with the SF36 questionnaire. The questionnaire does not include items about the severity of ear complaints. Despite this difference, the COQOL is mapped for economic evaluation,<sup>18</sup> which is still a future perspective of the OQUA. The OQUA is currently implemented in our electronic health record using EPIC version AUG 2019. After implementation, this construct file can be used across the country for all hospitals using Epic as well. Furthermore, the OQUA will also be build-in in other electronic health records (EHR) used across the country, for example Hix (Chipsoft).

#### 4.3 | Future research

The OQUA is designed to be easily applicable in daily practice, being able to be used on a computer, tablet/phone and on paper. With the scoring manual (Figure 2), a graphical representation can be developed according to the wishes of the ENT surgeon

and in line with the capacity of their facility (electronic patient file, ICT department or institution). Future studies might focus on enablers of and barriers to its implementation. By adapting the graphical representation to the preferences of the users and having the OQUA being filled out days before consultation, the implementation is more likely to be successful in the end. Furthermore, translation in foreign languages and cross-culturally validation is of additional value. Chinese validation and Swedish validation are currently under investigation. Optimisation of graphical representation might be needed in the near future.

The OQUA should give us a better insight in the patient complaints and impact in the large amount of different diagnoses in otology. A more evidence-based approach with the help of this PROM and broad evaluation of the current treatments might change our future treatment and hopefully our results for the best. Longitudinal analysis, before and after otologic interventions, using the OQUA and at different time intervals in large groups should give us these insights.

## 5 | CONCLUSION

A validated generic otologic questionnaire covering all relevant ear complaints, and their impact, is presented. The results of this study show good construct validity, test-retest reliability and responsiveness. The OQUA is ready for broad implementation for all patients visiting the ENT surgeon with an ear complaint. Its design furthermore makes the OQUA an effective tool to evaluate the effectiveness of surgical and non-surgical interventions by the ENT surgeon.

### CONFLICT OF INTEREST

None to declare.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

### ORCID

Jeroen T. Kraak  <https://orcid.org/0000-0002-7064-1853>

Lisette M. van Leeuwen  <https://orcid.org/0000-0003-0671-0161>

Sophia E. Kramer  <https://orcid.org/0000-0002-0451-8179>

Paul Merkus  <https://orcid.org/0000-0002-5859-7386>

### REFERENCES

1. Jacobson GP, Newman CW. The development of the Dizziness Handicap Inventory. *Arch Otolaryngol Head Neck Surg.* 1990;116(4):424-427.
2. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Head Neck Surg.* 1996;122(2):143-148.
3. Philips JS, Haggard M, Yung M. A new health-related quality of life measure for active chronic otitis media (COMQ-12): development and initial validation. *Otol Neurotol.* 2014;35:454-458.
4. Gates GA, Veral AM. Validation of the ménières disease patient-oriented symptom-Severity Index. *Arch Otolaryngol Head Neck Surg.* 2005;131:863-867.

5. Bruinewoud EM, Kraak JT, van Leeuwen LM, Kramer SE, Merkus P. The Otology Questionnaire Amsterdam: a generic patient reported outcome measure about the severity and impact of ear complaints. A cross-sectional study on the development of this questionnaire. *Clin Otolaryngol*. 2018;43(1):240-248.
6. Vet HCWd, Terwee CB, Mokkink LB, Knol DL. *Measurements in Medicine – A Practical Guide*. Cambridge, UK: Cambridge University Press; 2011.
7. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. Cambridge, UK: Cambridge University Press; 2011.
8. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745.
9. Mokkink LBTC, Patrik DL, et al. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *J Clin Epidemiol*. 2010;63:737-745.
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual 2012.
11. Robinson K, Gatehouse S, Browning GG. Measuring patient benefit from otorhinolaryngological surgery and therapy. *Ann Otol Rhinol Laryngol*. 1996;105(6):415-422.
12. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Model*. 1999;6(1):1-55.
13. Robinson K, Gatehouse S, Browning GG. *The Glasgow Health Status Questionnaires Manual*. Nottingham, UK: MRC Institute of Hearing Research, Glasgow Royal Infirmary; 1998.
14. Jacob C. *Statistical power analysis for the behavioral sciences*. (2nd ed). Department of Psychology, University New York, New York, NY: Lawrence Erlbaum Associates Publishers; 1988. 567 p.
15. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
16. Landis JR, Koch GC. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
17. Martin TPC, Moualed D, Paul A, et al. The Cambridge Otology Quality of Life questionnaire: an otologyspecific patient-reported outcome measure. A paper describing the instrument design and report of preliminarnt reliability and validity. *Clin Otolaryngol*. 2015;40:13-139.
18. Frew EJ, Harrison M, Rossello Roig M, Martin TP. Providing an extended use of an otological-specific outcome instrument to derive cost-effectiveness estimates of treatment. *Clin Otolaryngol*. 2015;40(6):593-599.

**How to cite this article:** Kraak JT, van Dam TF, van Leeuwen LM, Kramer SE, Merkus P. The Otology Questionnaire Amsterdam: A generic patient-reported outcome measure about the severity and impact of ear complaints. Validation, reliability and responsiveness. *Clin Otolaryngol*. 2020;45: 506–516. <https://doi.org/10.1111/coa.13545>