Readability Assessment of Commonly Used German Urological Questionnaires

Pavel Lyatoshinsky\textsuperscript{a}  Manolis Pratsinis\textsuperscript{b}  Dominik Abt\textsuperscript{a}  Hans-Peter Schmid\textsuperscript{a}  Valentin Zumstein\textsuperscript{a,b}  Patrick Betschart\textsuperscript{a}

\textsuperscript{a}Department of Urology, Cantonal Hospital of St. Gallen, St. Gallen, Switzerland; \textsuperscript{b}Department of Urology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Key Words
Readability assessment • Urology • Questionnaire • German language

Abstract

\textbf{Purpose:} We intended to assess the readability of the German versions of commonly used urological questionnaires and identify questions that are potentially demanding for patients. \textbf{Materials and Methods:} The Guidelines of the European Association of Urology were analyzed for recommended questionnaires. Readability of the German versions of these questionnaires including their respective single-items was analyzed using established readability assessment tools. \textbf{Results:} A total of 13 questionnaires were analyzed. The calculated readability scores ranged between the 4.3th and 10.3th grade level. Easiest readability as calculated by median grade levels was found for the short and long forms of the International Consultation on Incontinence Questionnaires-Female and -Male Lower Urinary Tract Symptoms and the SF-Qualiveen (all median grade level 5.0). The short form of the International Index of Erectile Function showed the hardest readability (median grade level 10.0). Readability of the single-items varied widely between the assessed questionnaires with up to 80% (the International Index of Erectile Function ) of their single-items being written above recommended grade levels. \textbf{Conclusions:} The majority of commonly used German urological questionnaires comply with recommended readability levels. Some questionnaires as well as single-items of most of the questionnaires clearly exceed recommended readability levels. This should be considered for interpretation of their results and when revising questionnaires.

Introduction

Medical questionnaires are not only routinely used in clinical practice, but also for the assessment of patient reported outcomes in clinical trials. The use of validated questionnaires allows a reproducible evaluation of recent complaints and the assessment of changes during a longer-term course. Therefore, most of the urological guidelines encourage clinicians to implement questionnaires into clinical practice and diagnostic procedures.

Development processes of medical questionnaires vary substantially, which has been shown to affect their quality [1]. Especially, readability represents a key component of comprehensibility and has been shown to be often neglected [2]. Therefore, the American Medical Association and the USA National Institutes of Health (NIH) recommend a reading level of medical content for non-professionals from 5th to 6th grade [3] and 7th to 8th grade [4] respectively to improve the readability of medical health materials.

Readability of commonly used medical questionnaires has been assessed for different medical disciplines [5, 6]. However, these analysis exclusively assessed the mean...
level of comprehensibility of complete questionnaires, and did not evaluate the readability of each survey item. Thus, very complicated single items may be concealed by less demanding ones and result in an overall good readability of the whole questionnaire [7].

A recent readability analysis of English urological questionnaires found an overall good readability for only 5 out of 13 questionnaires. Moreover, all of the surveys showed very demanding single items [8]. Therefore, the aim of our study was to assess the readability of frequently used urological questionnaires in German language including an analysis of their single-items to identify questions that might be particularly demanding for patients and may require a special consideration when interpreting the results.

Materials and Methods

The 2018 guidelines of the European Association of Urology were screened for recommended questionnaires. Validated versions of the questionnaires were used whenever available. The 36-Item Short Form Health Survey version 2 (SF-36v2) [9, 10] was chosen as a benchmark questionnaire as it is one of the most commonly used questionnaires, and has recently been subjected to a comprehensibility improvement [10].

The readability assessment was performed by extracting each question or single-item into a separate Microsoft Word document (Microsoft Corp., Redmond, Washington, WA, USA). Sub-questions or items within the questionnaires which were not recognized and analyzed by the software as a complete sentence or question were combined with potential question choices to form and test only complete sentences as recommended elsewhere [11]. Single word answers, as well as copyright notices, disclaimers, acknowledgments, author information, citations and references were excluded from analysis. All analysis were performed using the Readability Studio Professional Edition version 2015 software for Mac (Oleander Software, Ltd., Vandalia, OH, USA). A complete list of tools used for the analysis is provided in table 1.

Descriptive statistics using median and range were performed for all tests corresponding to a grade level, i.e. German Simple Measure of Gobbledygook (G-SMOG) [12, 13], Wiener Sachtextformel (WS) [12], Lesbarkeitsindex [14]. The Amstad Test [15] is charted and results in a score between 0 and 100.

Results

A total of 13 questionnaires were analyzed for their readability (table 2), only 6 of these surveys have been validated for the German language. The other 7 questionnaires were obtained from the corresponding questionnaire specific homepage SF-Qualiveen [16], International Consultation on Incontinence questionnaires...
### Table 1. Description of applied readability assessments (based on Betschart et al. [25])

<table>
<thead>
<tr>
<th>Readability tool</th>
<th>Tool details</th>
<th>Parameters analyzed</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-SMOG (Bamberger - Vanecek) [12, 13]</td>
<td>German variation of the original SMOG-score; calculates the grade level of a document</td>
<td>number of words with 3 or more syllables</td>
<td>(no. of words with 3 or more syllables) – 2</td>
</tr>
<tr>
<td>Wiener Sachtextformel [12]</td>
<td>for evaluation of nonfiction literature; average sentence length, word length and complexity are analyzed; results in a grade level score</td>
<td>word length/complexity and sentence length; complex words have 3+ syllables, long words 7+ characters.</td>
<td>( 0.1935 \times [(\text{number of complex words}) / \text{number of words}) \times 100] + 0.1672 \times (\text{number of words}/\text{number of sentences}) + 0.1297 \times (\text{number of long words}/\text{number of words}) \times 100] – 0.0327 \times (\text{number of monosyllabic words}/\text{number of words}) \times 100] – 0.875 )</td>
</tr>
<tr>
<td>Lesbarkeitsindex [14]</td>
<td>formula that can be applied to documents of any Western European language; possible scores are 0 to 70. Higher scores represent harder readability; possible grade level adjustment for use with German literature ranging from 4th to 15th grade recalculation of the Flesch Reading Ease [26] for German text; typical use of longer words in German language compared to English is considered; uses also a lower weighting for the sentence-length factor; possible scores are 0 to 100, the higher the easier to read</td>
<td>sentence length and number of long words (7+ characters)</td>
<td>number of words/number of sentences ( \times 100 ) (number of long words/number of words)</td>
</tr>
<tr>
<td>Amstad Test [15]</td>
<td></td>
<td>average number of syllables per word and average number of words per sentence</td>
<td>( 180 – (\text{number of syllables/number of words}) – [58.5 \times (\text{number of words/number of sentences})] )</td>
</tr>
</tbody>
</table>

### Table 2. Median grade levels of the different surveys

<table>
<thead>
<tr>
<th>Survey</th>
<th>No. of items</th>
<th>SMOG</th>
<th>WS</th>
<th>LIX</th>
<th>Median (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICIQ-FLUTS-SF [27]</td>
<td>24</td>
<td>4.3</td>
<td>5.6</td>
<td>5.0</td>
<td>5.0 (4.3–5.6)</td>
</tr>
<tr>
<td>ICIQ-FLUTS-LF [27]</td>
<td>36</td>
<td>4.4</td>
<td>5.6</td>
<td>5.0</td>
<td>5.0 (4.3–5.6)</td>
</tr>
<tr>
<td>ICIQ-MLUTS-SF [28]</td>
<td>26</td>
<td>4.4</td>
<td>5.3</td>
<td>5.0</td>
<td>5.0 (4.4–5.3)</td>
</tr>
<tr>
<td>SF-Qualiveen [29]</td>
<td>8</td>
<td>4.7</td>
<td>6.7</td>
<td>5.0</td>
<td>5.0 (4.7–6.7)</td>
</tr>
<tr>
<td>SF-36v2 [9]</td>
<td>50</td>
<td>5.0</td>
<td>6.2</td>
<td>5.0</td>
<td>5.0 (5.0–6.2)</td>
</tr>
<tr>
<td>I-QOL [30]</td>
<td>22</td>
<td>5.3</td>
<td>7.0</td>
<td>5.0</td>
<td>5.3 (5.0–7.0)</td>
</tr>
<tr>
<td>ICIQ-MLUTS-LF [28]</td>
<td>43</td>
<td>4.6</td>
<td>5.9</td>
<td>6.0</td>
<td>5.9 (4.6–6.0)</td>
</tr>
<tr>
<td>IPSS [31]</td>
<td>8</td>
<td>6.1</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0 (6.0–6.1)</td>
</tr>
<tr>
<td>Qualiveen [32]</td>
<td>30</td>
<td>6.4</td>
<td>9.3</td>
<td>7.0</td>
<td>7.0 (6.4–9.3)</td>
</tr>
<tr>
<td>ICSI [33]</td>
<td>4</td>
<td>5.2</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5 (5.2–9.0)</td>
</tr>
<tr>
<td>NIH-CPSI [34]</td>
<td>13</td>
<td>7.3</td>
<td>8.5</td>
<td>10.0</td>
<td>8.5 (7.3–10.0)</td>
</tr>
<tr>
<td>IIEF [35]</td>
<td>15</td>
<td>7.7</td>
<td>9.4</td>
<td>9.0</td>
<td>9.0 (7.7–9.4)</td>
</tr>
<tr>
<td>IIEF-5 [36]</td>
<td>5</td>
<td>10.0</td>
<td>10.3</td>
<td>9.0</td>
<td>10.0 (9.0–10.3)</td>
</tr>
<tr>
<td>Median</td>
<td>5.2</td>
<td>6.7</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>4.3–10.0</td>
<td>5.6–10.3</td>
<td>5.0–10.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICIQ-FLUTS-SF = International Consultation on Incontinence Questionnaires–Female Lower Urinary Tract Symptoms short form; ICIQ-FLUTS-LF = ICIQ-FLUTS long form; ICIQ-MLUTS-SF = international consultation on incontinence questionnaires – male lower urinary tract symptoms short form; I-QOL = incontinence quality of life; ICIQ-MLUTS-LF = ICIQ-MLUTS long form; IPSS = international prostate symptom score; ICSI = international cystitis symptom index; NIH-CPSI = national institutes of health chronic prostatitis symptom index; IIEF = international index of erectile function; IIEF-5 = IIEF short form.
(ICIQ) [17] or from official self-help organizations, which provide a German online 5-item version International Index of Erectile Function (IIEF-5) [18], Interstitial Cystitis Symptom Index (ICSI) [19].

The questionnaires had a median length of 232 words, with a range of 43 words (ICSI) to 383 words ICIQ Male Lower Urinary Tract Symptoms Long Form Module (ICIQ-MLUTS LF). The SF-36v2 questionnaire, used as a reference questionnaire, has a length of 822 words. Analysis of word complexity (3 or more syllables) of the questionnaires ranged from 15.5 (ICIQ-MLUTS SF) to 29.6% (IIEF-5). Analysis of word length showed 35.8 (ICIQ-MLUTS SF) to 48.8% (ICSI) of words with 6 or more characters.

Calculation of grade levels using SMOG, WS and Lesbarkeitsindex showed readability scores between 4.3th and 10.3th grade (table 2). The easiest readability, as calculated by the median grade level of SMOG, WS and Lesbarkeitsindex was found for the ICIQ-FLUTS-SF and -LF, ICIQ-MLUTS-SF, SF-Qualiveen and SF-36v2 [5.0 (4.3–6.7)]. The hardest readability was found for the IIEF-5 [10.0 (9.0–10.3)].

These results were confirmed by the Amstad Test, where scores of 76 points (corresponding to “fairly easy” readability) were found for the ICIQ-FLUTS-SF and ICIQ-MLUTS-SF. The IIEF-5 proved to be the most difficult of all assessed questionnaires (fig. 1), with a score of 43 points (corresponding to “difficult” readability). Using the SF-36v2 as a benchmark, six of the urological questionnaires showed an easier or harder readability respectively, while the score of the ICSI was equal, i.e. 66 points.

Analysis of single-items of the different questionnaires showed a wide range of readability scores. In figure 2, the three most commonly used surveys (IPSS, IIEF and CPSI-NIH) according to a PubMed search (all languages included in the search) were assessed using SMOG, WS and Lesbarkeitsindex. Single-item assessments for the other questionnaires are presented in figure 3.

Assessing the single-items of all questionnaires exclusively using the WS, which was specifically created for the German language [12], readability scores still showed a wide range (fig. 4). The widest range was found for the SF-36v2, I-QOL and Qualiveen questionnaires. Based on the WS, a readability level above the maximum recommended 8th grade level [4] was found for the following proportions of single-items of the questionnaires: IIEF-5: 80%; Qualiveen: 53%; IIEF: 47%; NIH-CPSI: 46%; SF-Qualiveen: 38%; SF-36v2: 34%; I-QOL: 32%; ICSI and IPSS: 25%; ICIQ-MLUTS-LF: 16%; ICIQ-FLUTS-SF: 13%; ICIQ-MLUTS-SF: 12% and ICIQ-FLUTS-LF: 11%.

Discussion

In this study we analyzed the readability of commonly used urological questionnaires in German. A wide range of readability levels was found for the assessed questionnaires, with some of them clearly exceeding the 7th to 8th grade reading level recommended by the NIH [4] and the 5th to 6th grade reading level recommended by the
Fig. 3. Readability grade levels for single-items of the questionnaires ICIQ-FLUTS (SF and LF) (A, B), ICIQ-MLUTS (LF and SF) (C, D), I-QOL (E), Qualiveen (F), SF-Qualiveen (G), SF-36v2 (H), ICSI (I), IIEF-5 (J) (SMOG = G-SMOG; WS = wiener sachtextformel; LIX = lesbarkeitsindex).
American Medical Association [3]. Ten of the 13 questionnaires appeared to be comprehensible and did not exceed the recommended reading levels (table 2). The IIEF, IIEF-5 and NIH-CPSI, on the other hand, show clearly increased reading levels in the applied tests and therefore should in general be regarded as too difficult.

In order to conclude from the questionnaire’s result on the presence or severity of a disease, it is important that questionnaires are answered correctly and completely. Thus, items exceeding patients’ reading skills may induce to give an invalid response or to skip the item. It is therefore important not only to evaluate the average readability of entire questionnaires, but also of the individual items of each questionnaire [7,11]. Importantly, even questionnaires with a good median readability (e.g. ICIQ-MLUTS/-FLUTS questionnaires) had individual items, which were clearly above accepted readability levels (fig. 3).

Compared to a recent analysis of the English versions of these questionnaires [8], there are certain analogies. The IIEF-5, IIEF, NIH-CPSI were shown to be the questionnaires with the most difficult readability in English as well as in German. However, twice as many German as English questionnaires meet the recommended over all reading levels (10 vs. 5 surveys respectively). Remarkably, only 6 of the German questionnaires have been validated and 2 of them (i.e. IIEF and NIH-CPSI [20, 21]) are exceeding the recommended reading level.

Different options to improve readability have been described previously, including substitution of multisyllabic words or simplification of sentence structure [22]. As some of the German questionnaires have not been validated yet, such improvements could be implemented in the validation process. Changes in validated questionnaires could be considered while undergoing revisions, as it had been performed for the SF-36 (version 1) in the past [10]. Moreover, the readability testing of individual questions should be considered when creating new surveys.

This study has some limitations, which have to be addressed. There is no consensus, which readability formulas should be used for assessing questionnaires. As performed in previous studies, we therefore used a combination of different assessment tools [2, 23]. Additionally, readability tests do not provide information about the content and other factors like the font types, images or appealing layout which also affect the comprehensibility [24]. Furthermore, not all of the questionnaires assessed had been validated, and therefore might be used in slightly modified versions. Thus it seems to be important to improve and validate such questionnaires in the future.
Conclusion

Readability of commonly used German urological questionnaires largely complies with recommended reading grade levels. However, inappropriate reading levels were found for certain questionnaires and for several single-items of all assessed questionnaires. These results should be considered for the interpretation of outcomes, future revisions and validation of German urological questionnaires.

References