Readability assessment of commonly used urological questionnaires

Patrick Betschart¹, Dominik Abt¹, Hans-Peter Schmid¹, Pascal Viktorin¹, Janine Langenauer¹, Valentin Zumstein¹,²
¹Department of Urology, Cantonal Hospital of St. Gallen, St. Gallen, Switzerland, ²Department of Urology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Purpose: This study was performed to assess readability of the most commonly used questionnaires in urology including a separate analysis of their single-items to identify questions that might be especially demanding for patients.

Materials and Methods: The guidelines of the European Association of Urology were screened for recommended questionnaires. Readability was analyzed for complete questionnaires as well as their single-items separately using well established readability assessment tools, including Flesch-Kincaid grade level (FKGL), Simple Measure of Gobbledygook grade level (SMOG), Coleman-Liau Index (CLI), Gunning-Fog Index, and the Flesch Reading Ease formula.

Results: A total of 13 questionnaires were included to the analysis. Calculation of grade levels (FKGL, SMOG, CLI, FGI) showed readability scores of 2.7th to 16.7th grade. Easiest readability as calculated by median grade levels was found for the short form of the International Consultation on Incontinence Questionnaires-Female Lower Urinary Tract Symptoms short form (FLUTS-SF) while the short form of the International Index of Erectile Function (IIEF-5) showed the hardest readability. Based on the FKGL between 0% (FLUTS-SF) and 80% (IIEF-5) of the single-items were written above the recommended grade levels.

Conclusions: The questionnaires that are used most frequently in urology mainly show a satisfactory overall readability. Inadequate readability levels were not only found for individual questionnaires but also for single-items of the majority of assessed questionnaires. This requires consideration for the interpretation of results and when developing novel health-related surveys.

Keywords: Comprehension, reading; Surveys and questionnaires; Urology

© The Korean Urological Association

INTRODUCTION

Questionnaires represent an integral part in everyday medical practice and leading urological guidelines recommend their use in various subspecialties of urology. They facilitate a time saving and structured assessment of current complaints and allow a systematic assessment of changes during a longer-term course. Many clinical trials are based on validated questionnaires as they represent the most reliable way to assess patient reported outcomes. Thus, a PubMed search reveals over 5,200 hits for the International Prostate Symptom Score (IPSS), over 3,300 for the International Index of Erectile Function (IIEF) and its short form IIEF-5 and more than 700 hits for the Chronic Prostatitis Symptom Index (CPSI).

Readability of health-related content represents a fundamental component of comprehensibility but is often neglected, which has especially been shown for patient...
education material [1,2]. Therefore, the USA National Institutes of Health (NIH) and the American Medical Association (AMA) recommend a 7th to 8th grade or 5th to 6th grade reading level for written health materials respectively [3,4].

Previously, Bergman et al. [5] analyzed the readability of 76 different health-related quality of life instruments and reported on a median 65th grade reading level. Like similar studies in different fields of medicine [6,7], they reported on generally satisfactory results. However, all of these studies evaluated the average readability of the complete questionnaires. This approach provides no information on potential variations in the readability of single-items. Thus, very difficult single-items can be covered up by easier ones and compromise a meaningful completion of a questionnaire and the informative value of its results [8].

The aim of our study was to assess readability of the most commonly used questionnaires in urology including a separate analysis of their single-items to identify questions that might be especially demanding for patients and, therefore, deserve particular attention when interpreting the results of such surveys.

**MATERIALS AND METHODS**

We reviewed the guidelines published by the European Association of Urology (EAU) for recommended urological questionnaires. In addition, the 36-Item Short Form Health Survey version 2 (SF-36v2) questionnaire [9] was included into the analysis as a benchmark. This survey represents the most widely used generic measure of health-related quality of life, has already undergone a revision to improve comprehensibility [10] and is also recommended by the EAU guidelines.

The English versions of the questionnaires were assessed by copying each single-item into a Microsoft Word document (Microsoft Corp, Redmond, WA, USA). Several items within the surveys (e.g., sentence fragments with subsequent question choices) are not recognized by the readability-test programs as a complete sentence or a question. Therefore, incomplete phrases were combined with potential question choices to form and test only complete sentences as recommended elsewhere [11]. Response options which were single words and therefore are likely to score as very easy to read, were not assessed. Copyright notices, disclaimers, acknowledgments, author information, citations, and references were excluded from the analysis. Each extracted text was analyzed for readability using the software package Readability Studio Professional Edition version 2015 for Mac (Oleander Software, Ltd., Vandalia, OH, USA). Readability assessment tools used for the analyses are described in Table 1 [12-15].

Descriptive statistics using median and range were performed for all tests corresponding to a grade level (i.e., Flesch-Kincaid grade level [FKGL], Simple Measure of Gobbledygook grade level [SMOG], Coleman-Liau Index [CLI], Gunning-Fog Index [GFI]). The Flesch Reading Ease (FRE) formula is charted and results in a score between 0 and 100.

**RESULTS**

Totally 13 questionnaires were included to the analysis (Table 2) [9,16-25]. The median length of the questionnaires was 238 words with a range of 57 words (International Cystitis Symptom Index, ICSI) to 403 words (International Consultation on Incontinence Questionnaires-Male Lower Urinary Tract Symptoms long form, ICIQ-MLUTS-LF) for specific urological questionnaires. The SF-36v2 has a length of 546 words. Analysis of word complexity (3 or more syllables) of the questionnaires showed a range of 4% (ICIQ-Female Lower Urinary Tract Symptoms short form, ICIQ-FLUTS-SF) to 28% (IEIEF-5). Analysis of word length showed 17.9% (IPSS) to 38% (Short form-Qualiveen, SF-Qualiveen) of words with six or more characters.

Calculation of grade levels (i.e., FKGL, SMOG, CLI, GFI) showed readability scores of 27th to 167th grade (Table 2). Easiest readability as calculated by median (range) grade level of FKGL, SMOG, CLI, and GFI was found for the ICIQ-FLUTS-SF (4.4 [2.7–6.4]), while the IIEF-5 showed the hardest readability (13.95 [11.2–16.7]). These results were confirmed by the FRE, where scores of 90 points (corresponding to ‘very easy’ to ‘easy’ readability) were found for the short and long form of the ICIQ-FLUTS. With a score of 42 points (corresponding to ‘difficult’ readability) the IIEF-5 proved to be the most difficult questionnaire (Fig. 1) [12]. Using the SF-36v2 as a benchmark, seven of the urological questionnaires performed better, while five showed a harder readability.

Analyses of single-items showed a wide range of readability scores for all of the different questionnaires. Fig. 2 shows the results of the three most commonly cited surveys according to a PubMed search (i.e., IPSS, IIEF, CPSI) using FKGL, SMOG, CLI, and GFI. Single-item assessments for the other questionnaires are presented in Fig. 3.

Assessing the single-items of all questionnaires by only using the FKGL, which has been used most frequently and is considered as the gold standard of calculation of reading levels [26,27], readability scores still showed a wide range (Fig. 4). The widest range was found for the SF-36v2, incontinence
| Readability assessment tool | Characteristic | Analyzed parameter | Formula | Strength | Limitation |
|----------------------------|---------------|-------------------|---------|----------|------------|
| FKGL [12]                  | Used in the field of education. Results corresponding to USA grade level. | Average number of syllables per word and average number of words per sentence | $0.39 \times \text{average number of words per sentence} + 11.8 \times (\text{average number of syllables per word}) - 15.59$ | Wide availability, extensively validated, quick and easy to administer, correlates highly with other readability formulas | Solely based on polysyllable words and long sentence, therefore may underestimate reading difficulty of medical jargon (short but unfamiliar words) |
| SMOG grade level [13]      | Widely used for checking health messages. Results corresponding to USA grade level. | Number of words with 3 or more syllables | $1.043 \times \sqrt{\left( \frac{\text{number of polysyllables}}{30/\text{number of sentences}} \right) \times 3.1291}$ | Considered as gold standard measure of readability for consumer-oriented healthcare material | Complex calculation; May underestimate reading difficulty of medical jargon (short but unfamiliar words); Only validated in English |
| Coleman-Liau Index [14]    | Used for calibration of the readability of textbooks for the public school system (USA). Results corresponding to USA grade level. | Average number of letters and sentences per 100 words | $0.0588 \times \text{average number of letters per 100 words} - 0.296 \times \text{average number of sentences per 100 words} - 15.8$ | Relies on characters instead of syllables per word, therefore more accurately counted by computer programs than are syllables | Designed samples of 100 words, not for entire documents; May underestimate reading difficulty of medical jargon (short but unfamiliar words) |
| Gunning-Fog Index [15]     | Estimates the years of formal education a person needs to understand the text on the first reading. Commonly used to confirm that the intended audience can read the text easily. Results corresponding to USA grade level. | Average number of words per sentence, average number of Gunning-Fog hard words (3 or more syllables) | $0.4 \times (\frac{\text{average number of words per sentence}}{\text{average number of Gunning-Fog hard words}}) + 100 \times (\frac{\text{average number of 3+syllable words}}{\text{average number of syllables}})$ | Less time-consuming compared to FRE and FKGL if calculated manually | May underestimate reading difficulty of medical jargon (short but unfamiliar words) |
| FRE [12]                   | Best suited for school textbooks and technical manuals. Possible value 0–100. Higher scores indicate material that is easier to read. | Average number of syllables per word and average number of words per sentence | $206.835 - (1.015 \times \text{average number of words per sentence}) - (84.6 \times \text{average number of syllables per word})$ | Easily available through Microsoft Word Office package | Explanation points, colons and semicolons serve as sentence delimiters |

FKGL, Flesch-Kincaid grade level; SMOG, Simple Measure of Gobbledygook grade level; FRE, Flesch Reading Ease.

Data from the article of Kasabwala et al. Otolaryngol Head Neck Surg 2012;147:466-71 [2].

Table 1. Description of applied readability assessments (based on Kasabwala et al. [2]).
quality of life (I-QOL), and Qualiveen-questionnaires. Based on the FKGL, a readability level above the recommended 8th grade level was found for the following numbers and proportions of single-items of the questionnaires: IIEF-5: 4 (80%), CPSI: 8 (61%), IIEF: 7 (47%), SF-36v2: 9 (25%), Qualiveen: 6 (20%), IPSS and SF-Qualiveen: 1 (13%), I-QOL:
Fig. 2. Readability grade levels for single-items of the questionnaires. (A) International Prostate Symptom Score (IPSS), (B) International Index of Erectile Function (IIEF), and (C) Chronic Prostatitis Symptom Index (CPSI). FKGL, Flesch-Kincaid grade level; SMOG, Simple Measure of Gobbledygook grade level; CLI, Coleman-Liau Index; GFI, Gunning-Fog Index.

Fig. 3. Readability grade levels for single-items of the questionnaires. (A, B) International Consultation on Incontinence Questionnaires-Female Lower Urinary Tract Symptoms short and long form (ICIQ-FLUTS-SF and -LF), (C, D) ICIQ-Male Lower Urinary Tract Symptoms long and short form (ICIQ-MLUTS-LF and -SF), (E) incontinence quality of life (I-QOL), (F) Qualiveen, (G) SF-Qualiveen, (H) 36-Item Short Form Health Survey version 2 (SF-36v2), (I) International Cystitis Symptom Index (ICSI), (J) International Index of Erectile Function-short form (IIEF-5). FKGL, Flesch-Kincaid grade level; SMOG, Simple Measure of Gobbledygook grade level; CLI, Coleman-Liau Index; GFI, Gunning-Fog Index.
DISCUSSION

This study assessed the readability of questionnaires that are used in everyday urological practice and are recommended by leading guidelines.

A wide range of readability was found for the 13 questionnaires assessed, including results that clearly exceed the 7th to 8th grade or 5th to 6th grade reading level recommended by NIH or AMA respectively [3,4]. While questionnaires assessing male and FLUTS as well as incontinence symptoms (i.e., ICIQ-FLUTS-SF and -LF, ICIQ-MLUTS-SF and -LF, I-QOL, incontinence quality of life; ICSI, International Cystitis Symptom Index; SF-Qualiveen, Short-Form Qualiveen) are generally written very easy or easy, most of the other questionnaires (i.e., SF-36v2, IPSS, ICSI) did at least not clearly exceed recommended reading levels (Table 2). In contrast, the CPSI and the IIEF, including its short form, have a readability level that has to be considered as clearly too difficult according to all tests that were applied. These findings are largely consistent with assessments of questionnaires in other medical fields [5-7,11].

As items exceeding patients’ reading skills often induce them to give an invalid response or simply skip the item, readability analysis of questionnaires should not only report on mean results of the summarized text, but also assess each single-item separately [8,11]. Remarkably, even questionnaires with good average readability showed single-items clearly exceeding acceptable difficulty levels in this study (Figs. 3, 4).

Thus, based on the FKGL, the questionnaires IPSS, IIEF, and CPSI that are used most frequently in the literature have 13%, 47%, and 61% of question-items that are too difficult, respectively (Fig. 4). These numbers were even higher if other tests were applied (Fig. 2).

As reflected by the underlying formula of established assessment tools (Table 1), complexity of the sentence structure and the words used represent the main causes of poor readability. Koo and Yap [1] and Mossanen et al. [28] previously described that simplification of content is often accompanied by creation of longer sentences and words, and therefore, improved readability is not necessarily achieved. Dalziel et al. [29] could show, that the readability of patient education material could be improved by an average 31 grade level by simple substitutions of multisyllabic words. Further improvement has been shown to be achievable by adaption of sentence structure. Though such simplifications are not easily applicable to standardized and validated questionnaires, they might be considered if such questionnaires undergo revisions like it has been performed for the SF-36 (version 1) in the past [10] and more attention should be paid to readability if novel questionnaires are developed. Moreover, in the case of inconsistent or implausible results of questionnaires in clinical practice or trials, special attention should be paid to single-items that clearly exceed recommended readability levels (Figs. 2, 3).

The study has limitations that have to be addressed.

As there is no consensus which readability formulas are most suitable for assessing questionnaires, we decided to use a combination of validated and well-established tests. Moreover, readability tests do not provide information about the complexity of the content and factors other than text quality (e.g., appealing layout, font types, images) also affect comprehensibility [30]. Only English versions of the questionnaires were assessed in the study and, therefore, the results are not transferable to the multitude of translations that are available.

CONCLUSIONS

The questionnaires that are used most frequently in urology mainly show a satisfactory overall readability. Inadequate readability levels were, however, not only found for individual questionnaires but also for single-items of all
assessed questionnaires. This requires consideration for the interpretation of results and when developing novel health-related surveys.

CONFLICTS OF INTEREST

The authors have nothing to disclose.

REFERENCES

1. Koo K, Yap RL. How readable is BPH treatment information on the internet? Assessing barriers to literacy in prostate health. Am J Mens Health 2017;11:300-7.
2. Kasabwala K, Agarwal N, Hansberry DR, Baredes S, Eloy JA. Readability assessment of patient education materials from the American Academy of Otolaryngology—Head and Neck Surgery Foundation. Otolaryngol Head Neck Surg 2012;147:466-71.
3. Medlineplus. How to write easy-to-read health materials [Internet]. Bethesda (MD): U.S. National Library of Medicine, U.S. Department of Health and Human Services National Institutes of Health [updated 2017 Jun 28; cited 2018 Jan 15]. Available from: https://medlineplus.gov/etr.html
4. Weiss BD. Health literacy and patient safety: help patients understand. Manual for clinicians. 2nd ed. Chicago (IL): American Medical Association, American Medical Foundation; 2007.
5. Bergman J, Gore JL, Singer JS, Anger JT, Litwin MS. Readability of health related quality of life instruments in urology. J Urol 2010;183:1977-81.
6. Atcherson SR, Richburg CM, Zraick RI, George CM. Readability of questionnaires assessing listening difficulties associated with (central) auditory processing disorders. Lang Speech Hear Serv Sch 2013;44:48-60.
7. Alas AN, Bergman J, Dunivan GC, Rashid R, Morrisroe SN, Rogers RG, et al. Readability of common health-related quality-of-life instruments in female pelvic medicine. Female Pelvic Med Reconstr Surg 2013;19:293-7.
8. Homan S, Hewitt M, Linder J. The development and validation of a formula for measuring single-sentence test item readability. J Educ Meas 1994;31:349-58.
9. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
10. Ware JE Jr. SF-36 health survey update. Spine (Phila Pa 1976) 2000;25:3130-9.
11. Calderón JL, Morales LS, Liu H, Hays RD. Variation in the readability of items within surveys. Am J Med Qual 2006;21:49-56.
12. Flesch R. A new readability yardstick. J Appl Psychol 1948;32:221-33.
13. Hedman AS. Using the SMOG formula to revise a health-related document. Am J Health Educ 2008;39:61-4.
14. Coleman M, Liu TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60:283-4.
15. Gunning R. The technique of clear writing. New York: McGraw-Hill; 1952.
16. Brookes ST, Donovan JL, Wright M, Jackson S, Abrams P. A scored form of the Bristol Female Lower Urinary Tract Symptoms questionnaire: data from a randomized controlled trial of surgery for women with stress incontinence. Am J Obstet Gynecol 2004;191:73-82.
17. Abrams P, Avery K, Gardener N, Donovan J. The International Consultation on Incontinence Modular Questionnaire: www.iciq.net. J Urol 2006;175:1063-6; discussion 1066.
18. Patrick DL, Martin ML, Bushnell DM, Yalcin I, Wagner TH, Buesching DP. Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). Urology 1999;53:71-6.
19. Bonniaud V, Parratte B, Amarenco G, Jackowski D, Didier JP, Guyatt G. Measuring quality of life in multiple sclerosis patients with urinary disorders using the Qualiveen questionnaire. Arch Phys Med Rehabil 2004;85:1317-23.
20. Bonniaud V, Bryant D, Parratte B, Guyatt G. Development and validation of the short form of a urinary quality of life questionnaire: SF-Qualiveen. J Urol 2008;180:2592-8.
21. Lubeck DP, Whitmore K, Sant GR, Alvarez-Horine S, Lai C. Psychometric validation of the O'Leary-Sant interstitial cystitis symptom index in a clinical trial of pentosan polysulfate sodium. Urology 2001;57(suppl 1):62-6.
22. Barry MJ, Fowler FJ Jr, O'Leary MP, Bruskewitz RC, Holgrewe HL, Mebust WK, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol 1992;148:1549-57; discussion 1564.
23. Clemens QJ, Calhoun EA, Litwin MS, McNaughton-Collins M, Dunn RL, Crowley EM, et al. Rescoring the NIH chronic prostatitis symptom index: nothing new. Prostate Cancer Prostatic Dis 2009;12:285-7.
24. Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. Urology 1997;49:822-30.
25. Rosen RC, Cappelleri JC, Smith MD, Lipsky J, Peña BM. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impot Res 1999;11:319-26.
26. Steinbrook R. Searching for the right search—reaching the
medical literature. N Engl J Med 2006;354:4-7.

27. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care 2008;53:1310-5.

28. Mossanen M, Calvert JK, Wright JL, True LD, Lin DW, Gore JL. Readability of urologic pathology reports: the need for patient-centered approaches. Urol Oncol 2014;32:1091-4.

29. Dalziel K, Leveridge MJ, Steele SS, Izard JP. An analysis of the readability of patient information materials for common urological conditions. Can Urol Assoc J 2016;10:167-70.

30. Kandula S, Zeng-Treitler Q. Creating a gold standard for the readability measurement of health texts. AMIA Annu Symp Proc 2008:353-7.