



# The German linguistic validation of the Wisconsin Stone Quality of Life questionnaire (WisQoL)

Martin Gottstein<sup>1</sup> · Manolis Pratsinis<sup>2</sup> · Sabine Güsewell<sup>3</sup> · Patrick Betschart<sup>2</sup> · Dominik Abt<sup>2</sup> · Thomas Knoll<sup>1</sup>

Received: 20 April 2020 / Accepted: 7 August 2020 / Published online: 12 August 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

**Purpose** WisQoL (Wisconsin Stone Quality of Life questionnaire) is a disease specific, health related quality of life measure designed for patients who form kidney stones. The goal of this study was to develop and validate a German version of WisQoL.

**Methods** The German version of the WisQoL was developed following a standardized multistep process. Patients were recruited prior to stone treatment, and completed the questionnaire as well as the SF-36v2 (36-Item Short Form Health Survey). This was repeated 1, 3, and 6 months after stone surgery. Scores of the 28 questionnaire items were summarized into sum scores for four domains and a total score. The psychometric properties of the questionnaire were statistically analyzed.

**Results** The German WisQoL demonstrated excellent internal consistency (Cronbach's  $\alpha > 0.90$  for all domains at all visits). All inter-domain associations were positive. The test–retest reliability for patients with unchanged self-reported health state was considered satisfactory (Spearman's rho for total score 0.70 [95% CI 0.55 to 0.80]). The German WisQoL demonstrated good convergent validity with the validated SF-36v2 (correlation between corresponding items 0.44 to 0.64). All domain scores showed significant sensitivity to change induced by stone treatment ( $p \leq 0.05$ ). Total WisQoL scores generally improved during the first 3 months following stone treatment, and remained stable thereafter.

**Conclusion** The German WisQoL proved to be a reliable and robust instrument to evaluate health related quality of life measures of kidney stone patients in the clinical setting. It is expected to be of use for further research in patients with kidney stones.

**Keywords** Urolithiasis · Health related quality of life · HRQOL · Wisconsin Quality of Life Score · WISQOL · German WISQOL

---

Martin Gottstein and Manolis Pratsinis contributed equally to this work.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00345-020-03405-7>) contains supplementary material, which is available to authorized users.

---

✉ Martin Gottstein  
m.gottstein@klinikverbund-suedwest.de

Manolis Pratsinis  
Manolis.Pratsinis@kssg.ch

Sabine Güsewell  
Sabine.Guesewell@kssg.ch

Patrick Betschart  
patrick.betschart@kssg.ch

Dominik Abt  
Dominik.Abt@kssg.ch

Thomas Knoll  
t.knoll@klinikverbund-suedwest.de

## Introduction

Kidney stones are associated with a significant morbidity [1] and represent a global challenge due to a rising prevalence all around the world [2]. Each stone episode leaves its footprint in daily life by impairing the quality of life.

<sup>1</sup> Department of Urology, Kliniken Sindelfingen-Böblingen, Arthur-Gruber-Str. 70, 71065 Sindelfingen, Germany

<sup>2</sup> Department of Urology, Cantonal Hospital St. Gallen, Rorschacherstrasse 95, 9007 St. Gallen, Switzerland

<sup>3</sup> Clinical Trials Unit, Cantonal Hospital St. Gallen, Rorschacherstrasse 95, 9007 St. Gallen, Switzerland

Health related quality of life (HRQOL) represents a key consideration of clinical research [3, 4]. Assessing the HRQOL not only provides a valuable insight of the course of a disease, but also offers a better understanding of the impact of the disease on the individual concerned. Therefore, HRQOL deserves recognition and careful consideration in clinical practice.

Measuring HRQOL using validated test forms focuses on various aspects of physical and mental health conditions and, therefore, serves as a tool to delineate progress as well as success, and efficacy of an ongoing treatment [5]. While instruments measuring general HRQOL provide a broad perspective [6], they might not be ideal to gain a profound understanding of the impact of a specific disease on patients.

Recently a disease specific HRQOL measure for kidney stones was developed, the “Wisconsin Stone Quality of Life questionnaire” (WisQoL) [7]. It has been validated [8] and utilized [9] and was translated and validated in Spanish, Russian and Turkish [10–12]. The purpose of this study was to develop and validate a German version of the WisQoL.

## Materials and methods

### Translation process and pilot testing

Translation and linguistic validation was conducted in accordance with the multistep process established by Hutchinson [13]. The final consensus version was checked for readability [14] and successfully used for pilot testing including semi-structured interviews of 10 patients [15] suffering from urolithiasis.

### Patient selection and data collection

Patients were recruited in Sindelfingen, Germany, and St. Gallen, Switzerland. Inclusion criteria were patient age  $\geq 18$  years, written informed consent, patient scheduled for urinary stone treatment, and sufficient German language skills.

Collection of patient characteristics and administration of the German WisQoL and validated Short Form Health Survey (SF-36v2) was performed prior to stone surgery (baseline), as well as 1, 3 and 6 months thereafter. As the final part of the WisQoL asks for the presence of urinary stones, stone-related symptoms or interventions and other events interfering with quality of life during the last 4 weeks, this repeated administration allowed for analyses of intra-individual changes including test–retest reliability and sensitivity to change.

## Statistical analyses

Statistical analyses were performed with R, version 3.5.1. (R-Core-Team 2018) and the package psych, version 1.8.12.

WisQoL total score was calculated as the sum of scores of all items from 1a to 7f (28 items). Domain scores were calculated in accordance with the original publication [8]. Sum scores were calculated only from complete data.

Patients’ characteristics were evaluated using descriptive statistics. Approximate 95% confidence intervals for Spearman’s rank correlations ( $\rho$ ) were obtained through Fisher’s transformation. Statistical significance was evaluated at the 5% level without correction for multiple testing across items or domains.

Reliability was assessed by internal consistency (Cronbach’s  $\alpha$ ) separately for each visit and considering either the items of one domain or all items together. Test–retest reliability was assessed by correlation of domain and total scores between two consecutive visits (either at 1 month and 3 month, or at 3 month and 6 month follow-up) at which patients reported unchanged health state in the final items of the WisQoL (8.1 to 8.5). Two measures of correlation were used: The intraclass correlation (ICC), i.e. the fraction of total variation in scores that is attributable to differences among patients, and Spearman’s rank correlation ( $\rho$ ).

To assess convergent validity, relationships between pairs of items in the WisQoL and SF-36 v2 expected to measure the same component of life quality were evaluated with Spearman’s rank correlations ( $\rho$ ).

Sensitivity to change was measured by the mean difference in domain and total scores between two consecutive visits chosen such that the patient reported having stones or stone-related symptoms at the first visit but not at the second (WisQoL item 8.1 and 8.2). The significance of score improvement was analyzed using Wilcoxon signed-rank tests. For comparison, scores were also compared between two consecutive visits without change in stone presence or with unchanged health state to check that there was no (or only little) score improvement in this case.

Construct validity was tested by comparing patients with and without stones or stone-related symptoms (WisQoL item 8.1 or 8.2). Means were compared between groups with one-way analysis of variance followed by Tukey’s honest significant difference tests for pairwise comparisons.

## Results

A total of 227 patients were recruited between February 2018 and April 2019, with 19 dropouts due to non-adherence and insufficient language skills. Data for 208 patients

(140 men and 68 women; median age 54.9 years (range 18.1–86.4) who completed the questionnaires at recruitment were included (baseline, Suppl. Tab. 1).

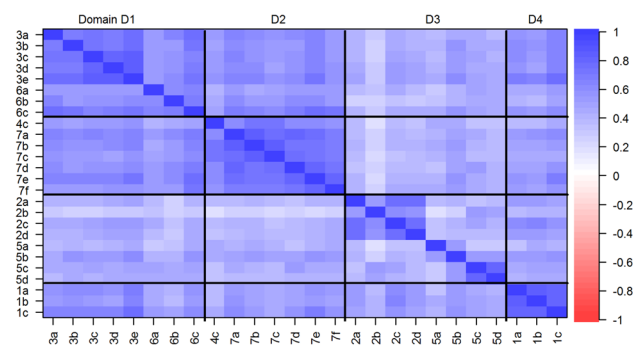
### Descriptive statistics for single items, domain scores, and total scores

All WisQoL items had scores spanning the full range from 1 to 5, and all items had similar variability of scores (SD) at an individual assessment (Suppl. Tab. 2). This indicates that all items assessed life quality on the same scale, i.e. a difference of  $\pm 1$  unit means the same for all items. Domain scores had different means and SD because they are calculated as the sum of different numbers of items.

As expected, total WisQoL scores varied widely among individual patients at baseline and after 1 month, whereas most patients had total scores between 120 and 140 after 3 and 6 months (Suppl. Fig. 1a). Most patients reached their individual maximum either after 1 month or after 3 months, but a few did so only after 6 months or stayed at low values. The contrasting time courses of total scores at the level of individual patients were reflected by an increase in mean total score up to the 3-month visit and no further change up to the 6-month visit (Suppl. Fig. 1b).

### Associations between single items, domain scores and total score

Spearman’s rank correlations between all pairs of individual items were positive ( $\rho > 0$ ), meaning that all items measured quality of life in the same sense, i.e. higher scores always reflect higher quality of life (Fig. 1). Together with the similar variability of item scores (Suppl. Tab. 2) this means that the calculation of simple sum scores based on the raw item scores is a meaningful way of summarizing the information.



**Fig. 1** Spearman rank correlations ( $\rho$ ) between all pairs of individual items. (Each cell of the image corresponds to one pair of items (rows and columns), and the tone indicates the strength of the correlation. Items are ordered according to the four domains, and thick lines indicate the subdivisions.)

Items of the same domain were on average more strongly correlated with each other than with items of the other domains (Fig. 1, Suppl. Tab. 3). Nevertheless, domain scores consistently showed substantial inter-domain correlations ( $\rho$  between 0.62 and 0.78), indicating a high conceptual relationship of the different domains (Table 1). Accordingly, all domain scores were also strongly correlated with the total score ( $\rho$  between 0.79 and 0.93, Table 1).

### Internal consistency and test–retest reliability

All domains had excellent internal consistency ( $\alpha > 0.90$ ). This was similar at baseline and at all follow-ups (all timepoints assessed: Suppl. Tab. 4; baseline and three-month follow-up: Table 3). Domain D3 had only ‘good’ internal consistency at baseline ( $\alpha = 0.87$ ), but this increased progressively during follow-up.

Test–retest reliability was calculated for patients with unchanged health state. Repeated assessments were correlated rather moderately with each other (Table 2).

### Convergent validity

Relationships between corresponding items in the WisQoL and SF36 questionnaire were evaluated with Spearman rank correlations. Questions assessing fatigue (WisQoL 1b—SF36 9g/i), social impact of disease (WisQoL 3a—SF36 6), concentration level (WisQoL 3e—SF36 4d) and levels of pain (WisQoL 5b—SF36 7) were assessed. All correlations were significantly positive (lower limit of 95% CI  $> 0$ ), even if only moderately so (Table 3).

**Table 1** Inter-domain associations: Spearman rank correlations between all pairs of domain scores, and correlations of each domain score with the total score, given with 95% CI

Domain(s)	At baseline	At 3-month follow-up
D1–D2	0.78 (0.71 to 0.84)	0.74 (0.62 to 0.82)
D1–D3	0.67 (0.57 to 0.75)	0.70 (0.58 to 0.80)
D1–D4	0.70 (0.61 to 0.78)	0.64 (0.50 to 0.74)
D2–D3	0.62 (0.52 to 0.71)	0.72 (0.60 to 0.81)
D2–D4	0.62 (0.51 to 0.71)	0.65 (0.51 to 0.75)
D3–D4	0.69 (0.60 to 0.77)	0.70 (0.57 to 0.79)
D1–total	0.93 (0.90 to 0.95)	0.83 (0.75 to 0.89)
D2–total	0.89 (0.84 to 0.92)	0.89 (0.83 to 0.93)
D3–total	0.84 (0.79 to 0.89)	0.92 (0.87 to 0.95)
D4–total	0.81 (0.74 to 0.86)	0.79 (0.70 to 0.86)

Domain 1—social impact (3a, 3b, 3c, 3d, 3e, 6a, 6b, 6c), Domain 2—emotional impact (4c, 7a, 7b, 7c, 7d, 7e, 7f), Domain 3—disease impact (2a, 2b, 2c, 2d, 5a, 5b, 5c, 5d) and Domain 4—impact on vitality (1a, 1b, 1c)

**Table 2** Internal consistency and test–retest reliability

Domain	Cronbach's alpha		Test–retest reliability	
	At baseline	3-months follow-up	ICC	Spearman's rho
D1	0.94 (0.92 to 0.95)	0.95 (0.94 to 0.96)	0.63 (0.49 to 0.74)	0.57 (0.40 to 0.70)
D2	0.95 (0.93 to 0.96)	0.95 (0.94 to 0.96)	0.66 (0.53 to 0.77)	0.72 (0.59 to 0.82)
D3	0.87 (0.84 to 0.89)	0.92 (0.90 to 0.93)	0.65 (0.50 to 0.76)	0.62 (0.46 to 0.75)
D4	0.92 (0.90 to 0.94)	0.91 (0.89 to 0.93)	0.60 (0.42 to 0.73)	0.57 (0.40 to 0.71)
Total	0.96 (0.95 to 0.97)	0.97 (0.97 to 0.98)	0.69 (0.55 to 0.79)	0.70 (0.55 to 0.80)

Internal consistency was tested for each domain and for all items of the questionnaire. Test–retest reliability was the assessment of correlations between domain and total scores at two consecutive assessments with unchanged state of health. This was defined as identical answers to items 8.1 to 8.5, a condition fulfilled either at 1 and 3 months or at 3 and 6 months by 88–89 patients for the four domains. All measures are given with 95% CI

(D1: social impact, D2: emotional impact, D3: disease impact, D4: impact on vitality)

**Table 3** Spearman rank correlations between corresponding items in the WisQoL and SF36 questionnaires (Approximate 95% CI are provided)

Items	At baseline	At 3-month follow-up
W 1b—SF36 9g	0.64 (0.54 to 0.73)	0.44 (0.27 to 0.58)
W 1b—SF36 9i	0.64 (0.54 to 0.72)	0.53 (0.37 to 0.65)
W 3a—SF36 6 (reversed)	0.50 (0.38 to 0.60)	0.49 (0.33 to 0.62)
W 3e—SF36 4d	0.60 (0.50 to 0.69)	0.47 (0.31 to 0.61)
W 5b—SF36 7* (reversed)	0.51 (0.39 to 0.61)	0.50 (0.34 to 0.63)

### Sensitivity to change

Domain and total scores consistently increased between two consecutive visits chosen such that either stone presence or current perception of stone-related symptoms improved between these visits. All changes were significant at the 5% level in Wilcoxon signed-rank tests (Suppl. Tab. 5), meaning that positive score changes were always more frequent than negative changes. In fact, scores improved even in the absence of a change in self-reported stone presence or symptoms (Suppl. Tab. 5). However, mean changes were much larger after stone removal and after symptom disappearance than in situations of unchanged self-reported condition.

### Construct validity

Mean total WisQoL scores did not differ between patients who did and those who did not report currently having stones in their urinary system (Suppl. Fig. 2a), but were significantly different between patients who did and did not report currently having stone-related symptoms (Suppl. Figure 2b).

## Discussion

Urinary stones are a common disease with reported prevalence rates up to 20% [16], and have been steadily increasing over the past years [17–19]. Urinary stones and their treatments can negatively affect patients and their quality of life [20]. In 2016, Penniston et al. developed and validated the WisQoL hereby creating a stone specific instrument with robust psychometric properties [8].

The psychometric properties of the German questionnaire showed excellent validation results, with a high internal consistency ( $\alpha > 0.9$ ), good sensitivity to change, and satisfactory convergent and construct validity. Minor exceptions were limited to single parameters and can be explained by the study's methodology. The test–retest reliability for patients with unchanged self-reported health state ranged considerably for the single domains (Spearman's rho 0.40–0.82), but was overall considered satisfactory (Spearman's rho for total score 0.70 [95% CI 0.55 to 0.80]). In principle, the uniformity of scales might promote a tendency to mechanically check similar scores for all items, especially towards the end of the questionnaire, leading to reduced differentiation. However, most pairs of items were only moderately correlated ( $\rho < 0.7$ ), suggesting that they were rated separately by the patients, and therefore not redundant (Fig. 1, Suppl. Tab. 3).

Repeated assessments were only moderately correlated with each other (Suppl. Tab. 5). Even in the absence of self-reported changes in condition, there could be distinct score improvements or score reductions. However, mean changes were much larger after stone removal and after symptom disappearance than in situations of unchanged condition, showing the ability of all scores to reflect changes in subjective stone-related condition over time.

Valid scores should not only reflect temporal changes in individual patients, but also differences between patients, so that they could e.g. be used to compare two treatment groups

after the treatment. In the original English WisQoL [8], this was tested by comparing mean scores of patients with and without stones, or with and without stone-related symptoms. Interestingly, our results showed that mean total scores differed between patients who did and those who did not report currently having stone-related symptoms (Suppl. Fig. 2b), but not between patients who did and did not report currently having stones in their urinary system (Suppl. Fig. 2a). This is particularly relevant, as this underlines the quality of WisQoL in identifying the impact of health related quality of life measures.

We conducted our study in accordance with the standard linguistic validation [13]. Due to the large number of patients and the repeated follow-up, we were able to assess changes over a period of time, even in patients who required repeated interventions.

Our study has some limitations. Enrollment was conducted by availability at both recruiting sites. Thus, a bias of patients recruited may have been introduced. Less than half of the initially recruited patients completed all items for all four visits. This loss in follow-up is in part due to the repeated questionnaires patients received. It is of note, that this close follow-up allowed us to assess the test–retest validity of WisQoL, a test that had not been performed in the original WisQoL validation study.

## Conclusions

The German version of the WisQoL is a robust measure of health related quality of life in patients with urinary stones. The validation is internally consistent and can satisfactorily discriminate between patients with a different stone and symptom status. Thus, the German WisQoL will be a helpful patient reported outcome measure in clinical studies assessing the impact of urinary stones and stone treatment on quality of life. The questionnaire is readily available from the authors.

**Acknowledgements** Pascal Viktorin, Frederik König, Prof. Dr. Thorsten Bach, Prof. Dr. Dr. h. c. mult. Jens Rassweiler, Dr. Mehmet Özsoy, Prof. Dr. Axel Häcker, Prof. Dr. Dr. Arkadiusz Miernik.

**Authors' contributions** MG, DA, TK: study design, data collection and interpretation, manuscript writing; MP, PB: data collection and interpretation, manuscript writing; SG: statistical analysis, data interpretation, manuscript writing.

**Funding** Study costs were covered by unspecified internal research funds of the two recruiting study centers.

**Availability of data, material, and code** Requests for sharing of data, and statistical code will be considered if the purposed use aligns with public good purposes, contingent on approval from the responsible local ethics committees.

**Conflicts of interest/Competing interests** The authors have no conflicts of interest.

**Ethics approval** Both study sites obtained approval by the ethics committee (F2017-033, EKOS2017-01944).

**Consent to participate and for publication** All participants provided written informed consent for study inclusion and publication of study results.

## References

- Whitehurst L, Jones P, Somani BK (2019) Mortality from kidney stone disease (KSD) as reported in the literature over the last two decades: a systematic review. *World J Urol* 37(5):759–776. <https://doi.org/10.1007/s00345-018-2424-2> (Epub 2018 Aug 27)
- Sorokin I, Mamoulakis C, Miyazawa K, Rodgers A, Talati J, Lotan Y (2017) Epidemiology of stone disease across the world. *World J Urol* 35(9):1301–1320
- Yousefi P, Rasekhi S, Heshmati H (2016) Quality of life in medical sciences. *Int J Med Res Health Sci* 5(S):43–46
- WHOQOL—Annotated Bibliography WHO/MNH/MHP/98.4.Rev.2Page 3 (1998) Annotated Bibliography of the WHO Quality of Life Assessment Instrument—WHOQOL. Department of Mental Health, World Health Organization
- Lu G, Brazier JE, Ades AE (2013) Mapping from disease-specific to generic health-related quality-of-life scales: a common factor model. *Value Health* 16:177
- Pietersma S, van den Akker-van Marle ME, de Vries M (2013) Generic quality of life utility measures in health-care research: conceptual issues highlighted for the most commonly used utility measures. *Int J Wellbeing* 3:173
- Penniston KL, Nakada SY (2013) Development of an instrument to assess the health related quality of life of kidney stone formers. *J Urol* 189:921
- Penniston KL, Antonelli JA, Viprakasit DP, Averch TD, Sivalingam S, Sur RL, Pais VM Jr, Chew BH, Bird VG, Nakada SY (2017) Validation and reliability of the Wisconsin stone quality of life questionnaire. *J Urol* 197(5):1280–1288. <https://doi.org/10.1016/j.juro.2016.11.097> (Epub 2016 Nov 23)
- Penniston KL, Antonelli JA, Viprakasit DP, Averch TD, Bird VG, Chew BH, Sivalingam S, Sur RL, Nakada SY, Pais VM Jr (2018) The effect of thiazide and potassium citrate use on the health related quality of life of patients with urolithiasis. *J Urol* 200(6):1290–1294
- Atalay HA, Ülker V, Canat L, Özer M, Can O, Penniston KL (2018) Validation of the Turkish version of the Wisconsin stone-quality of life questionnaire. *Turk J Urol* 45(2):118–123. <https://doi.org/10.5152/tud.2018.35305>
- Basulto-Martínez M, Mendez-Probst C, Olvera-Posada D, Velueta-Martínez I (2018) Quality of life in patients with kidney stones. Validation of the Spanish version of WISQOL questionnaire. <https://doi.org/10.13140/rg.2.2.28882.20169>
- Shestaev AY, Paronnikov MV, Protoshchak VV, Sinelnikov LM, Karpuschenko EG, Kiselev AO (2018) Russian version of the Wisconsin questionnaire for assessing the quality of life in patients with urolithiasis (Wisconsin stone-QOL (WISQOL)): the results of a pilot study. *J Exp Clin Urol* 2:50–57 (Russian)
- Hutchinson A, Bentzen N, König-Zahn C (1996) Cross-cultural health outcome assessment: a user's guide. European Research Group on Health Outcomes, Ruinen

14. Amstad T (1978) Wie verständlich sind unsere Zeitungen? [How readable are our newspapers?]. Dissertation, University of Zurich, Switzerland
15. Betschart P, Staubli S, Zumstein V, Babst C, Sauter R, Schmid HP, Abt D (2019) Improving patient education materials: a practical algorithm from development to validation. *Curr Urol* 13(2):64–69. <https://doi.org/10.1159/000499291>
16. Trinchieri A et al (2003) Epidemiology. In: Segura J, Conort P, Khoury S (eds) *Stone disease*, 21st edn. ICUD, Paris
17. Stamatelou KK et al (2003) Time trends in reported prevalence of kidney stones in the United States: 1976–1994. *Kidney Int* 63:1817
18. Hesse A et al (2003) Study on the prevalence and incidence of urolithiasis in Germany comparing the years 1979 vs. 2000. *Eur Urol* 44:709
19. Sanchez-Martin FM et al (2007) Incidence and prevalence of published studies about urolithiasis in Spain. A review. *Actas Urol Esp* 31:511
20. Raja A, Hekmati Z, Joshi HB (2016) How do urinary calculi influence health-related quality of life and patient treatment preference: a systematic review. *J Endourol* 30:727

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.