CrossMark

Measuring quality of life in patients with stress urinary incontinence: is the ICIQ-UI-SF adequate?

Zuzanna Kurzawa[1] · Jason M. Sutherland[1] · Trafford Crump[2] · Guiping Liu[1]

Abstract

Purpose The International Consultation on Incontinence Questionnaire Short Form (ICIQ-UI-SF) is a widely used four-item patient-reported outcome (PRO) measure. Evaluations of this instrument are limited, restraining user's confidence in the instrument. This study conducts a comprehensive evaluation of the ICIQ-UI-SF on a sample of urological surgery patients in Canada.

Methods One hundred and seventy-seven surgical patients with stress urinary incontinence completed the ICIQ-UI-SF pre-operatively. Methods drawing from confirmatory factor analysis (CFA), measures of reliability, item response theory (IRT), and differential item functioning were applied. Ceiling effects were examined.

Results Ceiling effects were identified. In the CFA, the factor loadings of items one and two differed significantly ($p < 0.001$) from item three indicating possible multidimensionality. The first two items reflect symptom severity not quality of life. Reliability was moderate as measured by Cronbach's alpha (0.63) and McDonald's coefficient (0.65). The IRT found the instrument does not discriminate between individuals with low incontinence-related quality of life.

Conclusions Due to low/moderate reliability, the ICIQ-UI-SF can be used as a complement to other data or used to report aggregated surgical outcomes among surgical patients. If the primary objective is to measure quality of life, other PROs should be considered.

Introduction

Both acute and chronic forms of urinary incontinence (UI) have significant impact on the quality of life of those affected. The International Consultation on Incontinence Questionnaire Short Form (ICIQ-UI-SF) is a widely used patient-reported outcome (PRO) for screening, measuring the symptoms, treatment outcomes, and impact of symptoms on quality of life [1, 2].

The International Continence Society's Consultation on Incontinence awarded the ICIQ-UI-SF a 'Grade A' status [3] for assessing symptoms and quality of life, applying evaluation criteria of validity, reliability, and sensitivity to change. These psychometric evaluations found reliability to be high, as measured by Cronbach's alpha (0.71–0.95 [1, 4, 5]). In addition, the ICIQ-UI-SF was sensitive to change in scores, including interventions for conservative management, anti-muscarinic therapy, and tension-free vaginal tape [1, 4, 6,

✉ Jason M. Sutherland
jason.sutherland@ubc.ca

Zuzanna Kurzawa
zuza.kurzawa@gmail.com

Trafford Crump
trafford@crumpemail.com

Guiping Liu
guiping.liu@ubc.ca

[1] Centre for Health Services and Policy Research, School of Population and Public Health, University of British Columbia, 201-2206 East Mall, Vancouver V6T 1Z3, Canada

[2] Department of Surgery, University of Calgary, Calgary, AB, Canada

7]. The ICIQ-UI-SF is widely used in pre-post studies of UI interventions [8].

Despite these past evaluations, a number of knowledge gaps exist regarding both what the ICIQ-UI-SF is measuring and how well it measures it. In the foundational publication, the ICIQ-UI-SF was described as having a "single strong underlying factor" [1]; however, it is also often described as an instrument for measuring UI severity and UI-related quality of life [4]. While these factors are related, it remains unclear whether the ICIQ-UI-SF measures one or both of these constructs.

Despite advancements in psychometric assessment, a limited number of methods have been applied to answer how well the ICIQ-UI-SF measures the purported construct(s). While Cronbach's alpha and kappa statistics have been used to assess reliability [1, 4, 5], the values ranged significantly. Moreover, the ICIQ-UI-SF has not been assessed for differential item functioning between respondents, or for ceiling effects, which have implications to calculating sensitivity to change and the minimally clinically important difference [9]. Lastly, validations of the instrument have generally been conducted on patients with lower incontinence burden recruited through outpatient urology clinics [4, 10] or undergoing conservative management of their condition [1]. It is unclear whether the findings from past studies are generalizable to the patients with high symptom burden proceeding with surgical treatment for their UI.

The purpose of this research is to evaluate the ICIQ-UI-SF among a cohort of patients with moderate-to-severe stress urinary incontinence and proceeding with surgical treatment for their UI. The results of this analysis will inform recommendations regarding applying this instrument in a population of patients with high symptom burden.

Materials and methods

This study is based on a prospectively recruited sample of men with stress urinary incontinence (SUI) who have elected for insertion of an artificial sphincter or urethral sling for treatment of their condition. Note that while the sample has SUI, a small group of patients had unspecific diagnostic codes which may have corresponded to mixed UI. Patients were identified from a convenience sample of three urologists who agreed to have their population of patients contacted to participate in the study. Only patients undergoing initial treatment for SUI were contacted to participate. The setting of the study is Vancouver Coastal Health authority (VCH), a geographically defined region encompassing Vancouver, Canada. Up to two telephone calls were made by VCH surveyors to contact potential participants. Patients not successfully contacted were mailed a study invitation.

All participants completed a survey package pre-operatively. The survey package included a description of the study, study personnel and contact information, a checklist of common health conditions, the ICIQ-UI-SF, and generic instruments (not reported in this study). Participants were given an option of completing the instruments online or by mail, and received two reminder emails or telephone calls to complete their survey. Participants returned their surveys between September 2013 and August 2016.

The VCH Legal and Privacy Office completed a Privacy Impact Assessment to ensure the protocol was consistent with privacy legislation, and patient information was adequately secured. The University of British Columbia's Behavioural Research Ethics Board approved the study (approval number H12-02062).

The ICIQ-UI-SF is a four-item instrument. The first three items' scores are summed to produce a total score ranging from 0 to 21. The first two items ask questions related to symptom severity, and the third asks about how much incontinence interferes with daily life. While the purported construct has not been consistently defined, a higher score implies lower UI-quality of life, or higher UI severity; in this analysis, the hypothesized construct is assumed to be 'UI-quality of life.' The last item of the ICIQ-UI-SF is an unscored diagnostic item assessing the perceived cause of incontinence; the diagnostic item was not evaluated in this study.

For all participants, demographic variables and addresses were available. Participants' address was linked to a neighborhood-level indicator of socioeconomic status (SES) reflecting highest educational achievement, unemployment, income, and housing [11]. The SES variable is a five-level categorical variable representing SES quintiles. This potential confounder was developed on comprehensive census data independent from this study. Missing SES indices could be attributable to new housing developments not yet assigned an index, living on reserve (municipalities legally associated with First Nations or Indian bands), or living outside of BC.

To characterize the study sample, descriptive statistics were tabulated. In addition, instrument ceiling effects were evaluated. This study looked at the percentage of respondents that had extreme total scores; the threshold of 15% is an indication of ceiling effects, in concordance with other studies [12].

Dimensionality was assessed using confirmatory factor analysis (CFA), as the foundational study of the ICIQ-UI-SF reported just one underlying factor [1]. At least three items are required to perform a CFA; however, this yields a 'just-identified model' where the parameter estimates perfectly fit the data [13], and goodness of fit statistics cannot be computed. The CFA was performed treating all responses as categorical, as well as sensitivity analyses treating all responses

as continuous. Wald Tests were applied to test whether factor loadings across the items are equal with significance set at p value < 0.05.

The first measure of reliability computed was Cronbach's alpha. This statistic ranges from 0 to 1; an alpha value greater or equal to 0.70 [14] is the widely used threshold for acceptability. Cronbach's alpha with deleted items was examined to see which item's exclusion improves the overall alpha. The second measure of reliability was McDonald's coefficient, which uses factor loadings extracted from the CFA. Similar to Cronbach's alpha, a value closer to 1 implies higher reliability/redundancy; a cutoff value for McDonald's coefficient is ≥ 0.70 [15].

A two-parameter graded-response model (GRM) was applied to the item-level data. The GRM accommodates ordered categorical responses, and items that have varying number of response levels [16]. The two parameters in this model are discrimination and location. The discrimination parameter indicates how well items discriminate between responses along the levels of UI-quality of life. When individuals have similar levels of UI-quality of life, highly discriminating items will predict with greater accuracy whether respondents will provide different responses to adjacent response levels. The location parameters indicate whether response categories measure along the continuum of UI-quality of life; this can also be inspected visually through the category response functions (CRFs).

Along with unidimensionality, local independence (LI) was examined. LI means that after controlling for the latent trait, UI-quality of life items are uncorrelated [17]. In this analysis, the Pearson's X^2 statistic, often applied to polytomous data, was used. X^2 values exceeding 0.20 generally indicate LI may be violated [17].

CRFs can be transformed into item and test information functions (IIF/TIF), each of which is a continuously valued index illustrating the item or instrument's ability to differentiate across individuals as a function of their UI-quality of life. Tall and narrow plots of IIFs/TIFs will characterize high discrimination.

To assess GRM model fit, the CRFs were visually inspected for clear distinction across response levels, and peaks for each response level somewhere across the continuum of UI-quality of life. In addition, it was noted if discrimination parameters are high (> 1.70) [18]. Shallow and wide CRFs are evidence of poor discrimination. Assessing model fit statistically is not possible with a three-item instrument.

Lastly, differential item functioning (DIF) was assessed. DIF was performed on participants' SES category. Studies have found that low SES groups not only report higher impairment, but also provide lower valuations of their health once impaired [19]; the latter can introduce bias at a group level when SES differences are unaccounted for. Other

research reported that among a cohort of stress UI patients, a number of factors, including socioeconomic status, independently impacted scores on instruments valuing quality of life, providing rationale for considering DIF on SES [20]. The two highest SES quintiles were compared to the three lowest SES quintiles. For sensitivity analyses, the highest SES quintile was compared to the lowest quintile. DIF was assessed using the likelihood ratio test (LRT) [21].

The CFA was conducted using Mplus Version 6.11, and subsequent analyses were conducted using R Version 3.3.1 and the MIRT package.

Results

There were a total of 177 participants. The response rate of among eligible patients was 64.5%. The average age of participants was 68 years. Most participants were in the highest or second highest SES quintile (44.1%). As shown in Table 1, over one-half of participants were scheduled for insertion of urethral sling.

The rate of missing responses to any question was 1.02% in the whole study cohort. As such, a complete case analysis was used; this approach was not expected to bias the findings. All responses to items were left-skewed. Among participants, 15% scored 21 out of 21 indicating ceiling effects. See Fig. 1 for distribution of ICIQ-UI-SF total scores.

In the CFA, factor loadings for items 1 and 2 were much more similar than item 3. For item 3, only 19% of the variance was attributable to UI-quality of life. The Wald Test value was 16.5, with a p value of 0.0003, providing evidence that parameters of items 1 and 2 are not equal to the

Table 1 Demographic characteristics of the sample of participants ($N = 177$) and summary of ICIQ-UI-SF scores

Age (years)	
Mean	68.86
Standard deviation	8.71
Range	21–87
SES quintiles	N (%)
1 (highest SES)	37 (20.90)
2	41 (23.16)
3	29 (16.38)
4	25 (14.12)
5 (lowest SES)	30 (16.95)
Surgery type	
Urethral sling	94 (53.11)
Artificial sphincter	83 (46.89)
ICIQ-UI-SF response summary	
Mean	15.86
Standard deviation	3.94
Range	3–21
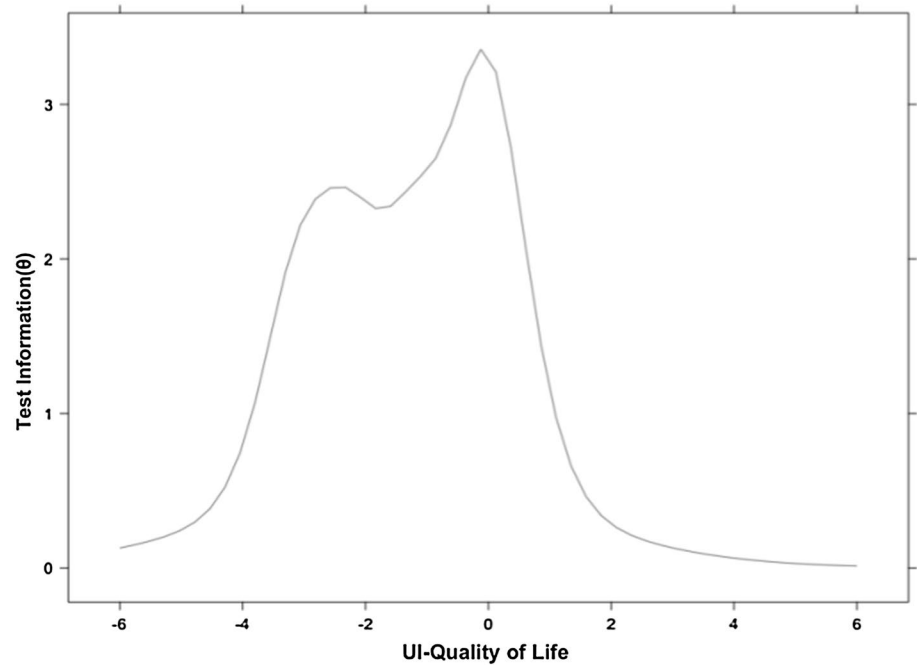
Table 2 Results from confirmatory factor analysis

	Factor loading	Standard error	R^2	p value
Item 1	0.82	0.10	0.67	< 0.0001
Item 2	0.80	0.10	0.64	< 0.0001
Item 3	0.44	0.09	0.19	< 0.0001

Table 3 GRM model parameters

Parameter	Item 1	Item 2	Item 3
A	2.55	2.44	0.88
B1	− 3.11	− 1.19	− 4.41
B2	− 2.75	0.17	− 3.56
B3	− 2.31		− 2.95
B4	− 0.18		− 1.97
B5			− 1.65
B6			− 0.92
B7			− 0.6
B8			0.01
B9			0.81
B10			1.36

A discrimination, *B* difficulty

parameter associated with item 3 (Table 2). Most variance attributable to the underlying factor was captured by items 1 and 2, suggesting the ICIQ-UI-SF primarily measures symptom severity.

Cronbach's alpha was 0.63, below the acceptability value of 0.70. Deletion of the third item resulted in higher reliability ($\alpha^* = 0.67$) than deletion of items one ($\alpha^* = 0.46$) or two ($\alpha^* = 0.45$). McDonald's coefficient was 0.65, also below the acceptability value.

Although unidimensionality is a strong assumption for this analysis based on the evidence from the CFA, the degrees of freedom are too low to investigate a multidimensional model. This analysis proceeded under the assumption of unidimensionality—although the directionality of bias introduced is unclear, past research and simulation studies suggest this violation should not significantly change results [22].

Item pairs 1 and 3 exhibited violations to local independence ($X^2 = 0.24$), although the value was not significant (p value = 0.402).

As shown in Table 3, the discrimination parameters for the first two items were considered high [18]. The curves for the lower response levels did not have unique peaks,

meaning that the association between the level of UI-quality of life and each response level is weak. Item 1 had a bimodal IIF, with highest discrimination when patients had high UI-quality of life (2–4 SD below the mean). Item 2 had a peaked and very narrow IIF suggesting highest discrimination 2 SD below and above the mean. For item 3, the discrimination parameter indicated low/moderate discrimination [18], and the CRFs were flat and overlapping; fit could be improved by collapsing this item's response levels. The IIF was correspondingly wide and flat.

The TIF indicated that the most information on the latent trait is revealed when respondents are below the average UI-quality of life (up to 4 standard deviations below the mean) as shown in Fig. 1. Correspondingly, the

ICIQ-UI-SF does not differentiate participants with low UI-quality of life.

No items were flagged as exhibiting DIF when comparing the two lowest SES quintiles with the three highest SES quintiles, nor when comparing the lowest SES quintile with the highest quintile.

Discussion

There was evidence to support two possible constructs underlying the ICIQ-UI-SF: symptom severity and interference. The implication for this population of patients is that the ICIQ-UI-SF is primarily a measure of symptom severity, not UI-related quality of life. If measuring UI-related quality of life is of interest to clinicians or researchers, other instruments should be considered.

A number of domains associated with quality of life have been reported to be affected by UI including travel, social activities, emotional health, activities of daily living, and sexual function [23]. In particular, those who experience shame, anxiety, and/or avoid social interaction report lower quality of life [2, 24–26]. Instruments that reflect some of these domains may be more appropriate for measuring UI-related quality of life, particularly among patients who have elected for surgical intervention, and so more likely to have more burdensome UI. The Incontinence Quality of Life Instrument (I-Qol), for example, contains subscales evaluating psychosocial impacts and social embarrassment related to UI [27].

The reliability of the instrument was low/moderate, implying differences between scores (e.g., 10 or 11) are not necessarily attributable to clinical differences between patients, but to measurement error. For this reason, ICIQ-UI-SF scores should not be used to inform priority setting or triage of urological surgery patients.

The instrument also does not discriminate between individuals with low UI-quality of life and it has significant ceiling effect. Ceiling effects can limit an instrument's sensitivity to change, meaning it may undermine its ability to measure the effects of interventions.

Item 3 had very low discrimination across the scale of UI burden. This is likely attributable, in part, to the number of response levels. Research in optimizing response levels is recommended.

This analysis had a number of limitations. Only men with moderate/high symptom burden were included in the sample. The results may not be generalizable to women with UI. Since there were few respondents with low levels of UI burden, model fit may have been undetectably poor for this level of symptom severity. Lastly, because the ICIQ-UI-SF is so brief, the model fit of the CFA and GRM cannot be assessed statistically. Future work should evaluate DIF for age, sex, and cultural background, as research suggests this can affect reported UI impairment [25].

Conclusions

This study found that, among men, the ICIQ-UI-SF primarily measures symptom severity. Other instruments should be considered if the objective is to measure UI-related quality of life. A range of methods demonstrated reliability is low/moderate, and particularly compromised for patients suffering from severe incontinence.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Avery, K., Donovan, J., Peters, T. J., Shaw, C., Gotoh, M., & Abrams, P. (2004). ICIQ: A brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourology and Urodynamics, 23*, 322–330. https://doi.org/10.1002/nau.20041.

2. Kwon, B. E., Kim, G. Y., Son, Y. J., Roh, Y. S., & You, M. A. (2010). Quality of life of women with urinary incontinence: A systematic literature review. *International Neurourology Journal, 14*, 133–138. https://doi.org/10.5213/inj.2010.14.3.133.

3. Donovan, J., Badia, X., Corcos, M., Gotoh, M., Kelleher, C., Naughton, M., et al. (2003). Symptom and quality of life assessment. In L. Cardozo, S. Khoury, A. Wein (Eds.), *Proceedings of the 2nd International Consultation on Incontinence Health* (pp. 519–584). Plymouth: Health Publication Ltd.

4. Hashim, H., Avery, K., Mourad, M. S., Chamssuddin, A., Ghoniem, G., & Abrams, P. (2006). The Arabic ICIQ-UI SF: An alternative language version of the English ICIQ-UI SF. *Neurourology and Urodynamics, 25*, 277–282. https://doi.org/10.1002/nau.20212.

5. Pereira, S. B., Thiel, R. D. R. C., Riccetto, C., da Silva, J.M., Pereira, L. C., Herrmann, V., et al. (2010). Validation of the international consultation on incontinence questionnaire overactive bladder (ICIQ-OAB) for Portuguese. *Revista Brasileira De Ginecologia E Obstetricia, 32*, 273–278. https://doi.org/10.1590/S0100-72032010000600004.

6. Espuña Pons, M., Montserrat Puig, C., Rebollo, P., Vanrell Díaz, J. A., & Iglesias Guiu, X. (2005). Evaluation of the results of surgery treatment for female stress urinary incontinence with the ICIQ-UI SF questionnaire. *Medicina Clínica, 124*, 772–774.

7. Seckiner, I., Yesilli, C., Mungan, N. A., Aykanat, A., & Akduman, B. (2007). Correlations between the ICIQ-SF score and

urodynamic findings. *Neurourology and Urodynamics, 26*, 492–494. https://doi.org/10.1002/nau.20389.

8. Patki, P., Woodhouse, J. B., Patil, K., Hamid, R., & Shah, J. (2008). An effective day case treatment combination for refractory neuropathic mixed incontinence. *International Brazilian Journal of Urology, 34*, 63–72.

9. Hamilton, D. F., Giesinger, J. M., MacDonald, D. J., Simpson, A. H. R. W., Howie, C. R., & Giesinger, K. (2016). Responsiveness and ceiling effects of the Forgotten Joint Score-12 following total hip arthroplasty. *Bone & Joint Research, 5*, 87–91. https://doi.org/10.1302/2046-3758.53.2000480.

10. Tamanini, J. T. N., Dambros, M., D'Ancona, C. A. L., César, P. J. N., Palma, R., & Rodrigues, N. (2004). Validation of the "International Consultation on Incontinence Questionnaire—Short Form" (ICIQ-SF) for Portuguese. *Revista de Saúde Pública*. https://doi.org/10.1590/S0034-89102004000300015.

11. Vincent, K., & Sutherland, J. M. .(2013). *A review of methods for deriving an index for socioeconomic status in British Columbia*. Vancouver: University of British Columbia.

12. Lim, C. R., Harris, K., Dawson, J., Beard, D. J., Fitzpatrick, R., & Price, A. J. (2015). Floor and ceiling effects in the OHS: An analysis of the NHS PROMs data set. *British Medical Journal Open, 5*, e007765. https://doi.org/10.1136/bmjopen-2015-007765.

13. Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press. https://doi.org/10.1198/tas.2008.s98.

14. Elkin, E. (2012). Are you in need of validation? Psychometric evaluation of questionnaires using SAS. In *SAS Global Forum* (pp. 1–9).

15. Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*, 375–385. https://doi.org/10.1177/014662169802200407.

16. Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks: Sage Publications.

17. Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265–289. https://doi.org/10.3102/10769986022003265.

18. Baker F. B. (2001). *The basics of item response theory* (2nd edn.). Washington DC: ERIC Clearinghouse on Assessment and Evaluation.

19. Mielck, A., Vogelmann, M., & Leidl, R. (2014). Health-related quality of life and socioeconomic status: Inequalities among adults with a chronic disease. *Health and Quality of Life Outcomes, 12*, 58. https://doi.org/10.1186/1477-7525-12-58.

20. Tincello, D., Sculpher, M., Tunn, R., Quail, D., Van Der Vaart, H., Falconer, C., et al. (2010). Patient characteristics impacting health state index scores, measured by the EQ-5D of females with stress urinary incontinence symptoms. *Value in Health, 13*, 112–118. https://doi.org/10.1111/j.1524-4733.2009.00599.x.

21. Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847–862. https://doi.org/10.3758/BRM.42.3.847.

22. Zhang, B. (2008). Application of unidimensional item response models to tests with items sensitive to secondary dimensions. *Journal of Experimental Education, 77*, 147–166. https://doi.org/10.3200/JEXE.77.2.147-166.

23. Mallah, F., Montazeri, A., Ghanbari, Z., Tavoli, A., Haghollahi, F., & Aziminekoo, E. (2014). Effect of urinary incontinence on quality of life among Iranian women. *Journal of Family Planning and Reproductive Health Care, 8*, 13–19.

24. Hajjar, R. R. (2004). Psychosocial impact of urinary incontinence in the elderly population. *Clinics in Geriatric Medicine, 20*, 553–564. https://doi.org/10.1016/j.cger.2004.04.009.

25. Chiaffarino, F., Parazzini, F., Lavezzari, M., & Giambanco, V. (2003). Impact of urinary incontinence and overactive bladder on quality of life. *European Urology, 43*, 535–538. https://doi.org/10.1016/S0302-2838(03)00097-6.

26. Hajebrahimi, S., Corcos, J., & Lemieux, M. C. (2004). International consultation on incontinence questionnaire short form: Comparison of physician versus patient completion and immediate and delayed self-administration. *Urology, 63*, 1076–1078. https://doi.org/10.1016/j.urology.2004.01.005.

27. Patrick, D., Martin, M., Bushnell, D., Yalcin, I., Wagner, T., & Buesching, D. (1999). Quality of life of women with urinary incontinence: Further development of the incontinence quality of life instrument (I-QOL). *Urology, 53*, 71–76. https://doi.org/10.1016/S0090-4295(98)00454-3.