Validation of prostate cancer index and SF-12 short forms

Thierry Lebeau, MD,[1] Paul Perrotte, MD,[1] Luc Valiquette, MD,[1]
François Bénard, MD,[1] Michael McCormack,[1] MD, Fred Saad, MD,[1]
Pierre I. Karakiewicz, MD[2]

[1]Department of Urology, University of Montreal, Montreal, Quebec, Canada
[2]Cancer Prognostics and Health Outcomes Unit, University of Montreal, Montreal, Quebec, Canada

LEBEAU T, PERROTTE P, VALIQUETTE L, BENARD F, MCCORMACK M, SAAD F, KARAKIEWICZ PI. Validation of prostate cancer index and SF-12 short forms. The Canadian Journal of Urology. 2005;12(6):2873-2879.

Background: *Assessment of prostate cancer (PCa) specific and generic health-related quality-of-life (HRQOL) is frequently omitted due to several obstacles, such as respondent burden and infrastructure-related limitations. We attempted to reduce the number of items of two commonly used HRQOL assessment tools, namely the UCLA PCa Index (PCI) and the RAND SF-12, with the intent of generating the most parsimonious, yet psychometrically valid and reliable HRQOL assessment tool.*
Methods: *The PCI and SF-12 were administered to 2415 radical prostatectomy patients, and re-tested in a convenience sample of 35 men with PCa. Multivariate linear regression models defined the most predictive and item-reduced SF-12 and PCI item combinations. These were subjected to standard psychometric reliability and validity tests.*
Results: *The 8-item PCI sexual function (SF) scale was reduced to three items. The 5-item PCI urinary function (UF) scale was reduced to three items. The 6-item SF-12 mental health scale was reduced to three items, and the 6-item SF-12 physical scale was also reduced to three items. The total number of items was reduced from 27 to 12 (44%). The item-reduced scales accounted for over 85% of full-scale variance. All reliability and validity tests yielded highly satisfactory results.*
Conclusion: *We developed SF-12 and PCI short-forms, which consist of 12 of 27 (44%) original items and can be completed by most men within 2 minutes. The short-forms represent a valid substitute for the full scales, as they provide over 85% of full-scale information and demonstrate excellent reliability statistics. The short forms have the potential for decreasing respondent burden and infrastructure-related requirements, which may in turn promote HRQOL assessment after radical prostatectomy.*

Key Words: prostate cancer, health-related quality-of-life, psychometrics

Introduction

Ideally, health-related quality-of-life (HRQOL) should be routinely assessed and documented in men with localized prostate cancer (PCa). The SF-36 generic health-related quality-of-life questionnaire and the UCLA-Prostate Cancer Index (PCI) have been validated for that purpose.[1-2] Although, several new PCa-specific HRQOL questionnaires have been developed, the PCI and the SF-36 represent the most widely used tools.[1-4]

Despite the availability of validated HRQOL measures, HRQOL assessment is infrequently performed on a routine basis. Moreover, it is common to see that select items are administered instead of complete tools. For example, items extracted from the International Index of Erectile Function are commonly used to assess distinct parts of erectile function, such as rigidity of erections or ability to

Accepted for publication November 2005

Address correspondence to Dr. Pierre I. Karakiewicz, Cancer Prognostics and Health Outcomes Unit, University of Montreal, 1058 St-Denis, Montreal, Quebec H2X 3J4 Canada. Email: pierre.karakiewicz@umontreal.ca

achieve an erection.[5] Such practice is not uncommon, as clinicians attempt to reduce respondent and infrastructure-burdens related to HRQOL assessment, such as data collection, entry and analysis.

Questionnaire short-forms have been developed to address these difficulties. A 12-item version of the SF-36 (SF-12) has been validated.[6] Similarly, a 15-item short-form of the PCI has been developed.[7] However, additional item reduction may be justified to further encourage and facilitate routine HRQOL assessment.

Herein, we relied on previously established methodology, which uses multivariate linear regression modelling, to derive the shortest and most informative item-reduced SF-12 and PCI short-forms.[6] Moreover, we tested whether the item-reduced scales fulfil the psychometric criteria of reliability and validity.[8,9]

Material and methods

The principal sample consisted of men treated with radical prostatectomy (RP) in the Province of Quebec, between January 1, 1988 and January 16, 1996. These men were identified from the Quebec Health Care Plan Database using RP-specific billing code, which resulted in virtually complete ascertainment. A self-administered survey was mailed to 4546 that were alive.[10] The retest sample consisted of 35 consecutive men recruited at a PCa follow-up clinic, and focused on temporal reliability at 2-4 weeks after the initial survey. The Patient Privacy Protection Branch of the Quebec Health Care Plan approved the study.

Generic HRQOL was assessed with the 12-item RAND SF-12 questionnaire, which represents the short form of the SF-36 questionnaire.[6] SF-12 reliability and validity have been previously confirmed in populations of men with and without PCa.[6,7] PCa-specific outcomes were measured with the PCI.[1] The original PCI scales target sexual function (SF), sexual bother (SB), urinary function (UF), urinary bother (UB), and bowel function and bother. The psychometric properties of the PCI are scale specific, and each scale can be used independently. As bowel function and bother scales specifically target radiation-induced proctitis, for purpose of brevity these scales have been excluded in this study. A successful French- and English-Canadian validation of the SF-36 and of PCI urinary and sexual scales has been previously reported.[10]

Linear regression analysis was used to test the proportion of explained variance (R^2) in the full-scale score, when either individual items or item combinations were used as predictors. According to previously established methodology, forward selection was used to identify the most predictive and most parsimonious item combinations, which consisted of fewer items than contained within the full scale.[6] The cut-off for determining the ability of the reduced item combinations to explain full-scale variance was defined as $R^2 \geq 0.85$, which indicates that selected item(s) account for 85% or more of full-scale score. This methodology was previously used to derive the SF-12 from the SF-36.[6] The R^2 method also represents a validity test, as it ensures that the item-reduced scale defines 85% or more of the original latent variable, described by the full-scale. Urinary and sexual bother scales consist of single items, and are therefore not amenable to item reduction.

The most informative multi-item, reduced-scales were subjected to psychometric testing, which focused on reliability and validity.[11] Cronbach's alpha quantified internal consistency of item combinations.[12] Temporal reliability (test-retest reliability) was quantified with the Pearson product-moment coefficient. Validity testing used the R^2 method, as described above.[6] In all tests, two-sided significance level of 0.05 was used and all statistical analyses were performed using the Statistical Package for Social Sciences version 10.0 (SPSS Inc. Chicago Ill).

Results

Of 4546 eligible responders, 2415 (53.1%) participated in the survey, and will be referred to as the principal sample. Their mean age was 63.6 years, 51.3% had an annual household income above $30000, 45.8% had high school, or some college education, and 24.9% had college or higher education. Finally, 87.8% lived with a spouse, and 17.3% worked full or part time. The retest sample consisted of 35 participants, whose average age was 65.6 years.

Table 1 demonstrates the amount of full-scale score variance explained by individual scale items, in the principal sample. The PCI urinary function scale items individually explain between 55.2% and 74.8% of full-scale variance. This range was 41.8%- 73.6% for PCI sexual function scale items, 44.1%-65.2% for SF-12 mental scale items, and 5.4%-71.2% for SF-12 physical scale items.

Table 2 shows the amount of full-scale variance (R^2) that may be explained by the reduced number of items. Of five PCI urinary function items, the two most informative items addressed frequency of incontinence and problems with dripping urine or wetting pants. These items explained 86.4% of full-scale variance. Their Cronbach's alpha was 0.77

TABLE 1. **Amount of full-scale score variance observed in the principal sample, which was explained by individual PCI and SF-12 items. R^2 indicates percentage of full-scale variance explained by each single item. The most parsimonious and most predictive ($R^2 \geq 0.85$) item combinations, that were included in the final questionnaire short-form, are shown in bold.**

Scale	Item	R^2
PCI urinary function scale	**1. Over the last 4 weeks, how often have you leaked urine?**	**0.748**
	2. Dripping urine or wetting your pants?	**0.660**
	3. Which of the following best describes your urinary control during the last 4 weeks?	0.645
	4. How many pads or adult diapers per day did you usually use to control leakage during the last 4 weeks?	0.556
	5. Urine leakage interfering with your sexual activity?	0.552
PCI sexual function scale	**1. Overall, how would you rate your ability to function sexually during the last 4 weeks?**	**0.736**
	2. How would you describe the usual quality of your erections?	**0.730**
	3. Your ability to have an erection?	0.720
	4. How would you describe the FREQUENCY of your erections?	0.692
	5. Your ability to reach orgasm (climax)?	0.627
	6. During the last 4 weeks did you have vaginal or anal intercourse?	0.532
	7. How often have you awakened in the morning or night with an erection?	0.433
	8. Your level of sexual desire?	0.418
SF-12 mental function scale	**1. Accomplished less than you would like**	**0.652**
	2. Were limited in the kind of work or other activities	**0.580**
	3. Cut down on the amount of time you spent on work or other activities	**0.529**
	4. Did you have a lot of energy?	0.482
	5. Have you felt downhearted and depressed?	0.460
	6. Have you felt calm and peaceful?	0.441
SF-12 physical function scale	**1. Were limited in kind of work or other activities**	**0.712**
	2. Accomplished less than you would like	**0.699**
	3. Limited in moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	**0.513**
	4. Climbing several flights of stairs	0.511
	5. In general, would you say that your health is?	0.063
	6. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?	0.054

versus 0.85 that was recorded for the full-scale. Their test-retest product-moment was 0.93 versus 0.95 for full-scale.

Of eight PCI sexual function scale items, the two most informative items addressed the ability to function sexually and the quality of erections. These items explained 88.8% of full-scale variance. Their Cronbach's alpha was 0.77 versus 0.90 that was recorded for full-scale. Their test-retest product moment was 0.95 versus 0.99 for the full-scale.

Of six SF-12 mental function items, three explained 85.9% of the full-scale variance, Table 1. Their Cronbach's alpha was 0.76 versus 0.81 that was recorded for the full-scale. Their retest product moment was 0.96 versus 0.95 for the full-scale.

Of six SF-12 physical function items, three explained 86.9% of the full-scale variance, Table 1. Their Cronbach's alpha was 0.83 versus 0.72 that was recorded for the full-scale. Their retest product moment was 0.92 versus 0.93 for the full-scale.

Comment

Litwin and Ware contributed two landmark instruments, which respectively address PCa-specific and generic HRQOL.[1,6] Despite the availability of these and other tools and in spite of the importance of HRQOL in men with localized prostate cancer, HRQOL is not assessed on a routine basis. Respondent burden and infrastructure-related barriers, such as data collection, data entry and data analysis represent important contributors to lack of

TABLE 2. **Reliability statistics and the amount of full-scale variance (R2) accounted for by the most predictive and most parsimonious combination of items**

		Original Sample (n = 2415)					Retest Sample (n = 35)				
Domain	**Scale**	**NB of items**	**R^2**	**Cronbach's alpha**	**Mean score**	**SD**	**Test mean score**	**Test SD**	**Retest mean score**	**Retest SD**	**Test retest product-moment**
PCI urinary function	Reduced	2	0.864	0.77	63.	32.46	69.06	33	68.11	31.88	0.93
	Original	5	1.000	0.85	71	27.06	76.16	26	76.07	26.18	0.95
PCI sexual function	Reduced	2	0.888	0.77	22	27.58	42.76	35	43.37	34.07	0.95
	Original	8	1.000	0.90	21	23.11	36.22	29	37.54	28.69	0.99
SF-12 mental function	Reduced	3	0.859	0.76	76.9	31.32	90.71	21.8	91.90	22.00	0.96
	Original	6	1.000	0.81	72.5	23.59	80.40	17.7	83.38	17.51	0.95
SF-12 physical function	Reduced	3	0.869	0.83	72.1	36.28	85.71	27.5	87.62	26.30	0.92
	Original	6	1.000	0.72	72.0	24.13	84.05	20.8	84.52	21.09	0.93

routine HRQOL assessment. To circumvent this obstacle, we performed an item-reduction followed by psychometric validation of two established HRQOL assessment tools: the SF-12 and the urinary and sexual function scales of the PCI. Our reliability and validity tests have confirmed that the 12-item SF-12 could be successfully reduced to six items, and that the 15 urinary and sexual items of the PCI could be successfully reduced to six items. Our effort resulted in a 12-item questionnaire (44% of original items), instead of the original 27 items. The item reduction was accomplished without compromising of established reliability properties and with minimal detriment to content. As in the original questionnaire, the item reduced scale scores range from 0 to 100 and higher scores indicated better HRQOL. The 12-item questionnaire is shown in Table 3.

Questionnaire item reductions have been performed by other investigators. Ware and colleagues derived the original 12-item SF-12 from the 36-item SF-36, using the same methodology that we employed in the current analysis, namely step-forward linear regression with a 90% R^2 cut-off for item selection.[6] The two 6-item physical and mental component scales explained over 90% of the information contained within the original SF-36 component scales, and maintained the ability to measure eight original SF-36 domains, in either two-item or single item formats.[6] However, no objective, methodological justification was provided to explain the choice of the 90% R^2 cut-off.

Litwin used the same R^2, linear regression-based methodology to confirm the validity and R^2 characteristics of PCI short-forms. PCI 4-item urinary and 5-item sexual function short-forms demonstrated R^2 values between 95% and 96%. Item reduction was not based on R^2 considerations. Instead, duplicative items and lack of detrimental effects on reliability represented criteria for item removal.[7]

We used forward-step regression analysis for item reduction. This method allowed us to define item-reduced PCI scales that are shorter than the original PCI short-forms proposed by Litwin and colleagues, where four of five urinary function items and five of eight sexual function items have been selected.[7] Litwin's short forms demonstrated R^2 values between 95% and 96%. Our PCI item-reduction resulted in scales consisting of two of five original PCI urinary function items, and two of eight original PCI sexual function items. Our item-reduced urinary function and sexual function scales respectively explained 86.4% and 88.8% of the original PCI scale variances (R^2) and their selection was entirely based on the effect on R^2 in forward-step regression models. For both item-reduced scales the most informative items were different from those selected by Litwin.[7] This difference may reflect methodological differences. Cultural differences might also have contributed, as differences in the perception of HRQOL detriments may place more emphasis on some items then others. Therefore, questionnaire short-forms may be culture-specific. Litwin's and our PCI urinary and sexual

TABLE 3. **Questions comprising the 12-item generic and prostate cancer-specific tool.**

Mental function items:

1. During the past four weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? Accomplished less than you would like…
   All of the time, Most of the time, Some of the time, A little of the time, None of the time
2. During the PAST past four weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? Did work or other activities less carefully than usual …
   All of the time, Most of the time, Some of the time, A little of the time, None of the time
3. During the past four weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? Cut down on the amount of time you spent on work or other activities…
   All of the time, Most of the time, Some of the time, A little of the time, None of the time

Physical function items:

4. During the past four weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health? Were limited in kind of work or other activities …
   All of the time, Most of the time, Some of the time, A little of the time, None of the time
5. During the past four weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health? Accomplished less than you would like…
   All of the time, Most of the time, Some of the time, A little of the time, None of the time
6. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf…
   Yes, limited a lot, Yes, limited a little, No, not limited at all

Urinary function items:

7. Over the last 4 weeks, how often have you leaked urine?
   Every day, About once a week, Less than once a week, Not at all
8. How big a problem, if any, has each of the following been for you? Dripping urine or wetting your pants?
   No problem, Very small problem, Small problem, Moderate problem, Big problem

Urinary bother item:

9. Overall, how big a problem has your urinary function been for you during the last 4 weeks?
   No problem, Very small problem, Small problem, Moderate problem, Big problem

Sexual function items:

10. Overall, how would you rate your ability to function sexually during the last 4 weeks?
    Very poor, Poor, Fair, Good, Very good
11. How would you describe the usual QUALITY of your erections?
    None at all, Not firm enough for any sexual activity, Firm enough for masturbation and foreplay only, Firm enough for intercourse

Sexual bother item:

12. Overall, how big a problem has your sexual function been for you during the last 4 weeks?
    No problem, Very small problem, Small problem, Moderate problem, Big problem

function item-reduced scales need to be complemented with respectively one urinary bother and one sexual bother item.

To our knowledge, we are the first to perform an item reduction on the SF-12 mental and physical scales, from 12 to 6 items. Therefore, we cannot compare our results to existing data. However, we have demonstrated that despite significant item

reduction (50%), the reliability statistics of the SF-12 short-forms substantially exceeded the required cut-off of 0.7: internal consistency ranged from 0.76 to 0.83 and product-moments ranged from 0.92 to 0.96, Table 2. Excellent internal consistency has also been recorded for the PCI scales (0.77) and indicates that within the item-reduced scales, prostate cancer-specific and generic items correlate very well with one another. Similarly, excellent product-moments have been demonstrated for the PCI item-reduced scales (0.93 and 0.95). These findings indicate that assessment of generic and prostate cancer-specific HRQOL with item-reduced scales is stable over time (test-retest). Finally, the item-reduced cancer-specific and generic scales demonstrated excellent validity: between 85.9% and 88.8% of the original scale information could be obtained with the reduced scales. Taken together, these data indicate that the reduced scales are reliable (internal consistency and temporal reliability) and provide a valid estimate of the latent variable, which is defined by the original scale content. These excellent reliability and validity characteristics confirm the applicability of the item-reduced SF-12 and PCI scales, which contain as few as 12 of 27 original items (44%). The twelve items can be easily fit on a single page.[6] Moreover, most subjects can complete 12 items with 2 minutes.[6] These two considerations represented criteria used by Ware and colleagues in the development of the SF-12.[6] We find them equally important. In addition to reducing the responder burden, a shorter questionnaire also reduces infrastructure requirements, such as data collection, entry and analysis.

Our short-form is not intended to compete with established tools such as the PCI, the EPIC, or the Giesler PCa questionnaire.[1,3,4] Neither, is it meant to compete with the SF-36 or SF-12.[6] Arguably, the short-forms may not be capable of highly accurately discriminating between fine nuances in HRQOL detriments, to the same extent as the full length scale versions. Instead, the short-forms are designed to rapidly, quantify HRQOL with adequate validity and reliability and could be of greatest use in large scale studies, where hundreds or thousands of men are surveyed.

Our study has several limitations. A low response rate represents a caveat of our study, as only 53.1% of men invited to participate complied with the invitation. Non-response bias may undermine the validity of our findings. Non-respondents may favour other questions then those identified as most important in the current study. The validation was performed on a sample of men that received treatment for localized PCa. Therefore, we could not test the properties of this tool in men awaiting treatment or in men who opted for watchful waiting. Moreover, we chose to focus on urinary and sexual function scales, which precludes the applicability of the proposed tool to men treated with radiotherapy. Criticism may also be directed at the potential loss of discriminant properties, if the short-forms are used.[6] According to Ware, the benefits related to the ease of administration of a short questionnaire greatly outweigh the detriments related to the decrease in the number of health dimensions and detail that it may measure.[6] Although, we agree with Ware and Litwin, regarding a certain redundancy of items within both questionnaires, some investigators could disagree with that opinion.[1,6] In smaller scale HRQOL studies, a detailed assessment of numerous HRQOL domains may be very important. Alternatively, detailed assessment of a particular outcome might be required, such as for example the use of pads or diapers. This type of assessment might not be possible with an abbreviated questionnaire from which this particular item might have been removed.

Conclusion

We developed SF-12 and PCI short-forms, which consist of 12 of 27 (44%) original items, which can be completed by the majority of patients within 2 minutes.[6] The short-forms represent a valid substitute for the full scales, as they provide over 85% of full-scale information and demonstrate excellent reliability statistics, which all exceed the 0.7 threshold. We believe that these short forms will decrease respondent burden and infrastructure requirements, and that they will facilitate and promote assessment of HRQOL after radical prostatectomy. □

References

1. Litwin MS, Hays RD, Fink A, Ganz PA, Leake B, Brook RH. The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure. *Med Care* 1998;36(7):1002-1012.
2. Litwin MS, Hays RD, Fink A, Ganz PA, Leake B, Leach GE, Brook RH. Quality-of-life outcomes in men treated for localized prostate cancer. *JAMA* 1995 Jan 11;273(2):129-135.
3. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. *Urology* 2000 Dec 20;56(6):899-905.

4. Giesler RB, Miles BJ, Cowen ME, Kattan MW. Assessing quality of life in men with clinically localized prostate cancer: development of a new instrument for use in multiple settings. *Qual Life Res* 2000;9(6):645-665.
5. 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 Jun;49(6):822-830.
6. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34(3):220-233.
7. Litwin MS, McGuigan KA. Accuracy of recall in health-related quality-of-life assessment among men treated for prostate cancer. *J Clin Oncol* 1999 Sep;17(9):2882-2888.
8. Litwin MS. Health related quality of life in older men without prostate cancer. *J Urol* 1999 Apr;161(4):1180-1184.
9. Rosen RC, Cappelleri JC, Smith MD, Lipsky J, Pena 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 Dec;11(6):319-326.
10. Karakiewicz PI, Kattan MW, Tanguay S, Elhilali MM, Bazinet M, Scardino PT, Aprikian AG. Cross-cultural validation of the UCLA prostate cancer index. *Urology* 2003 Feb;61(2):302-307.
11. Shrout PE and Fleiss JL. Intraclass correlation: uses in assessing rater reliability. Psychol Bulletin 1999;86:420-428.
12. Cronbach LJ, Mehl PE. Construct validity in psychological tests. Psychological bulletin 1955;52:281-302.