Reproducibility of the international spinal cord injury quality of life basic data set

Post, Marcel W M; Forchheimer, Martin B; Charlifue, Susan; D'Andréa Greve, Julia Maria; New, Peter W; Tate, Denise G

Link to publication in University of Groningen/UMCG research database

ISCoS
The International
Spinal Cord Society

Check for
updates

ARTICLE

Reproducibility of the international spinal cord injury quality of life basic data set: an international psychometric study

Marcel W. M. Post [1,2] · Martin B. Forchheimer[3] · Susan Charlifue[4] · Julia Maria D'Andréa Greve[5] · Peter W. New [6,7,8] · Denise G. Tate[3]

Abstract

Study design Psychometric study.

Objectives To examine the reproducibility of the international spinal cord injury quality of life basic data set (QoL-BDS) in an international sample of community-dwelling adults with spinal cord injury or disease (SCI/SCD) and in subgroups with respect to age (< 50 vs. ≥ 50 years) and etiology.

Setting Outpatient rehabilitation clinics and community.

Methods Participants were people living with SCI/D in four countries, at least 1 year post onset and at least 18 years of age. The QoL-BDS consists of three items on satisfaction with life as a whole, physical health and psychological health rated on a 0–10 numerical rating scale. A fourth item on satisfaction with social life was included based on feedback from participants.

Results A total of 79 people were included. Median age was 52 years. Most participants (69.6%) had SCI, paraplegia (53%), and 40.5% reported a motor complete injury. Median time between tests was 14 days (range 4–27). Intra-class correlation (ICC) values of the items ranged from 0.66 to 0.80. ICC values of the three-item and four-item total scores were identical and good (0.83; 95% CI 0.75–0.89). Subgroup analyses showed ICC values ranging from 0.76 to 0.83. Bland–Altman plots suggested no bias for the three-item total score, but some bias for the four-item total score. The limits of agreement of both scores were similar, wide at individual level and small at group level.

Conclusions This study provides evidence of reproducibility of the current version of the QoL-BDS. When adding a fourth item, reproducibility was maintained.

Introduction

Spinal cord injury or disease (SCI/D) is associated with decreased quality of life (QoL) of the individuals involved [1]. QoL is therefore an important factor when evaluating outcomes that best capture the effects of rehabilitation treatments for persons with SCI. Studies on QoL following SCI/D are abundant, including objective and subjective evaluations of QoL and reflecting its multidimensional nature [2]. Results from previous research are difficult to compare, however, because of variation in definitions of QoL and the measures used [3]. In response to this situation,

✉ Marcel W. M. Post
  m.post@dehoogstraat.nl

1   Center of Excellence for Rehabilitation Medicine, Brain Center
    Rudolf Magnus, University Medical Center Utrecht, University
    Utrecht and De Hoogstraat Rehabilitation, Utrecht, The
    Netherlands

2   University of Groningen, University Medical Center Groningen,
    Center for Rehabilitation, Department of Rehabilitation Medicine,
    Groningen, The Netherlands

3   Department of Physical Medicine and Rehabilitation, University of
    Michigan, Ann Arbor, MI, USA

4   Craig Hospital, Englewood, CO, USA

5   Orthopedic Department Medical School, University of São Paulo,
    São Paulo, Brazil

6   Spinal Rehabilitation Service, Caulfield Hospital, Alfred Health,
    Melbourne, VIC, Australia

7   Epworth-Monash Rehabilitation Medicine Unit, Monash
    University, Melbourne, VIC, Australia

8   Department of Epidemiology and Preventive Medicine, School of
    Public Health & Preventive Medicine, Monash University,
    Melbourne, VIC, Australia

the International SCI QoL Basic Data Set (QoL-BDS) was developed as a three-item self-report questionnaire as part of the International SCI Data Sets Project [4, 5]. It was designed to include a minimal number of data elements, to be collected in clinical practice and to be included in any SCI/D study, in addition to the preferred QoL measure, if applicable [4]. The QoL-BDS is recommended for use in clinical practice and research by the International Spinal Cord Society (ISCoS) and the American Spinal Injury Association (ASIA), and included in the National Institute of Neurologic Disorders and Stroke Common Data Elements Project [6].

The International SCI Data Sets were not designed to be used as measures. Nevertheless, they need to be valid and reliable, across the continuum of care and internationally [7]. The first data on cross-cultural validity of the QoL-BDS from the United States and Brazil were published in 2014 [8, 9]. After that a study from The Netherlands showed validity of the three items and the option to use a total QoL-BDS score [10]. Secondary analysis of merged data from five countries (these three plus Australia and India) provided further indication of its concurrent and divergent validity [11]. Finally, a small study of inter-rater reliability of several International SCI Data Sets in an inpatient rehabilitation setting showed good inter-rater reliability of the QoL-BDS items when administered by different physicians a few days apart [12].

Based on these encouraging results, a comprehensive prospective international validation project was planned to establish the reliability, validity and responsiveness of the QoL-BDS across four countries (Australia, Brazil, Netherlands and United States). For a set of data elements to be accepted as reasonably reliable and valid, information on its reliability and reproducibility must be reported.

The current study is part of this larger project and focuses on the reproducibility of the QoL-BDS and answers the following question: What is the test–retest reliability and agreement between two points in time of the Qol-BDS, (a) in the whole sample; and (b) in subgroups with respect to age and etiology? It was hypothesized that the reliability of all three items and the total scale would be satisfactory, and that agreement would be good, in the whole sample and in these subgroups.

Methods

Design

Pre–post test. Repeated administration of the QoL-BDS to persons with SCI/D living in the community within a period of 14 days between the first test (T1) and the second (T2).

Participants

We aimed to include a total of 80 participants from five study sites: the University of Michigan (UM) in Ann Arbor, Michigan; Craig Hospital in Englewood, Colorado; De Hoogstraat Rehabilitation in Utrecht, The Netherlands; Caulfield Hospital in Melbourne, Australia; and Hospital das Clínicas in Sao Paulo, Brazil. Inclusion criteria were: at least 18 years of age at the time of the study, living with SCI/D for at least 1 year and not having substantial cognitive or psychiatric problems. We aimed for a balanced sample with respect to age (< 50 or ≥ 50 years) and etiology (SCI or SCD).

Procedures

Convenience sampling was performed at each site. Participants were recruited from people visiting the hospital for their regular follow-up, from the hospital's medical files, registries, and from the community or a combination of these. Potential participants received written and oral information about the study and signed written informed consent. The study was approved at all sites by their respective IRBs and Ethic Committees. The QoL-BDS was administered by the same rater in an oral interview or telephone interview during two time periods (T1, T2).

Instruments

QoL-BDS

The QoL-BDS fits the definition of subjective QoL, reflecting an individual's overall perception of and satisfaction with how things are in his/her life [13]. It includes three items on the individual's satisfaction with their life as a whole, physical health, and mental health. All items use a time frame of the past four weeks and a 0–10 numerical rating scale with higher scores indicating better QoL [4].

Including a fourth item for the domain, social life in the QoL-BDS was considered at the time of its development. It was decided not to include such an item because a separate International SCI Activities and Participation Data Set was developed simultaneously [14]. However, as described in greater details elsewhere, cognitive interviews conducted with participants from all sites as part of an earlier phase of the current project made clear that the QoL-BDS would be incomplete without such an item [15]. Therefore, a fourth item to rate satisfaction with social life, with the same time frame and response scale, was asked in addition to the QoL-BDS.

The QoL-BDS was developed in English and this version was used in the USA and Australia. It had already been translated into Dutch and Brazilian Portuguese for use in previous studies [8–10], following the recommendations of the International SCI Data Sets project [7].

Demographic and injury/disease data questionnaire

In addition to the QoL-BDS, demographic and injury/disease-related questions were also asked at the first administration. Demographic information included age (in years), gender, marital status, education (years of formal schooling), and employment status. Questions on SCI/D characteristics included date of onset, etiology, level, and completeness. Since many participants could not indicate the completeness of their SCI/D, a question on the degree of voluntary movement below the level of the lesion (response categories: none, some, full) was used as a proxy measure for lesion completeness.

Analysis

Age was dichotomized as up to 50 or 50 years or older. Etiology was dichotomized into non-traumatic (SCD) or traumatic (SCI), marital status into married or not married, and employment status into employed or not employed. Differences between QoL-BDS item and total scores with respect to age and etiology were tested using the Mann–Whitney test. Differences between QoL-BDS scores at t1 and t2 were tested using the Wilcoxon Matched Pairs test.

Internal consistency reliability was examined using Cronbach's alpha and inspection of corrected item-total correlations. For group comparisons, an alpha of at least 0.70 is "sufficient", an alpha of 0.80 or higher is "good" and an alpha of 0.90 is "excellent". Corrected item-total correlations should be at least 0.40 [16].

Agreement of QoL-BDS scores at T1 and T2 was examined using weighted Kappa for single items and intraclass correlations (ICC) for the total scores. A weighted kappa of 0.21–0.40 is considered fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect [17]. For the total scores, the two-way random effects model for absolute agreement was used. An ICC above 0.70 is "sufficient" and above 0.80 is 'good' [18].

Bland–Altman plots were used to analyze agreement of total QoL-BDS scores at T1 and T2 related to the mean scores of T1 and T2 in the whole sample. The "limits of agreement" were computed, defined as ± 1.96 SD of the difference score. This figure indicates the minimum difference between scores exceeding chance at an individual level [19]. Similarly, the limits of agreement at group level were calculated as ± 1.96 times the standard error of the difference score (SE = SD/$\sqrt{N}$). To express both figures in terms of effect size, both were divided by the SD of the baseline score. Cohen's approach was used to interpret these effect sizes: 0.2 is "small" 0.5 is "moderate", and 0.8 is "large" [20]. Finally, we visually inspected the Bland–Altman plots for possible bias, meaning an association between the differences between the two scores and the mean of the two scores.

All analyses except the Bland–Altman plots were performed for the individual item scores and the three-item and four-item total scores of the QoL-BDS. Results are presented for both total scores to evaluate the impact of adding a fourth item and to facilitate comparisons with other studies. Subgroup analyses were performed with respect to age (up to 50 years vs. 50 years or older) and etiology (SCI vs. SCD). Because of the low sample sizes per country, no country-specific analyses were performed.

Results

Between 15 and 19 participants per site were included, for a total of 83. Four participants' data were excluded after collection because their time between tests was outside of the established window, making the final sample 79.

The time between T1 and T2 was between 4 and 27 days (median 14 days, interquartile range 11–15 days) in the whole sample. Participant characteristics are displayed in Table 1. Sample composition varied substantially across sites. Three sites included predominantly males with SCI injured as young adults, whereas two other sites included more females, more often with SCD and injured at a higher age. Most participants had long-standing SCI/D (median 10 years). Most had paraplegia and 40.5% indicated no voluntary movement below the level of injury.

At T1, the distributions of the total scores were approximately normal for both the three-item score (Skewness − 0.28, Kurtosis − 0.54) and the four-item score (Skewness − 0.24, Kurtosis 0.59). The distributions of the QoL-BDS item and total scores are displayed in Table 2. The score distributions of the three-item and the four-item total scores were similar.

There were no significant differences in QoL-BDS-scores with respect to etiology or age group. Scores at T2 were generally slightly higher than at T1, but this difference was only significant ($p < 0.01$) for the item that asks for ratings in relation to life as a whole (Table 2).

Cronbach's alpha values were good (range 0.84–0.86) for both total scores at both test occasions. All subgroup analyses also showed good alphas (range 0.82–0.90). All but one corrected item-total correlation were at least 0.40, with the social life item in the SCD at T1 (0.35) being the exception.

Test–retest reliability of all QOL-BDS scores are displayed in Table 3. The weighted Kappa values of the four single items in the whole sample were all substantial. The ICC values of the three-item and four-item total scores were identical and good (both 0.83; 95% CI: 0.75–0.89). Subgroup analyses showed only one item with a weighted Kappa value below 0.60 (physical health in the younger age group), and satisfactory to good ICC values of the two total scores in all subgroups.

Table 1 Characteristics of the study group (*N* = 79)

	Age		Etiology		All (*n* = 79)
	Up to 50 N = 34	50 and older N = 45	SCD N = 24	SCI N = 55	
Age (median, IQR)	38 (32–43)	60 (54–71)	65 (37–74)	50 (39–58)	52 (38–62)
Age at onset (median, IQR)	26 (21–32)	50 (38–60)	54 (28–65)	32 (23–44)	37 (25–52)
Sex (% male)	76.5	60.0	37.5	80.0	67.1
Marital status (% married)	44.1	71.1	62.5	58.2	59.5
Education years (median; IQR)	14 (12–16)	14 (13–17)	14 (12–15)	14 (12–16)	14 (12–16)
Employment (% employed)	26.5	22.2	25	23.6	24.1
Level of lesion					
Paraplegia	47.1	57.8	54.2	52.7	53.2
Tetraplegia	50.0	26.7	16.7	45.5	36.7
Unknown	2.9	15.6	29.2	1.8	10.1
Voluntary movement below the lesion					
Full	5.9	15.6	16.7	9.1	11.4
Some	52.9	44.4	62.5	41.8	48.1
None	41.2	40	20.8	49.1	40.5
Site					
Ann Arbor, MI	14.7	20.0	4.2	23.6	17.7
Denver, CO	35.3	8.9	8.3	25.5	20.3
Melbourne	11.8	28.9	41.7	12.7	21.5
Sao Paulo	29.4	6.7	4.2	21.8	16.5
Utrecht	8.8	35.6	41.7	16.4	24.1

Table 2 Descriptive statistics QOL-BDS by age and etiology

	Age				Etiology					
	Up to 50 n = 34		50 and older n = 45		SCD n = 24		SCI n = 55		All n = 79	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
Life as whole 1	**6.8**[a]	**2.3**	6.6	1.9	6.5	2.1	6.8	2.1	6.7	2.1
Physical health 1	6.8	2.1	6.5	2.4	6.3	2.1	6.8	2.3	6.6	2.3
Psychol health 1	6.7	2.2	7.2	1.9	6.8	1.9	7.1	2.2	7.0	2.1
Social life 1	6.4	2.7	7.2	2.2	7.5	1.7	6.5	2.6	6.8	2.4
Total 3 items 1	6.8	2.0	6.8	1.8	6.5	1.8	6.9	1.9	6.8	1.9
Total 4 items 1	6.7	2.1	6.9	1.7	6.8	1.5	7.0	2.0	6.8	1.9
Life as whole 2	**7.4**[a]	**1.7**	7.2	1.5	7.2	1.5	7.3	1.6	7.3	1.6
Physical health 2	6.7	1.9	6.3	1.9	6.4	1.8	6.5	2.0	6.5	1.9
Psychol health 2	6.7	2.2	7.3	1.8	6.9	1.9	7.1	2.0	7.1	2.0
Social life 2	6.9	2.6	7.1	1.9	7.3	2.1	6.9	2.2	7.0	2.2
Total three items 2	6.9	1.7	6.9	1.7	6.9	1.5	6.8	1.6	6.9	1.6
Total four items 2	6.9	1.8	7.0	1.4	7.0	1.5	7.0	1.7	7.0	1.6

[a]Significant difference between T1 and T2 (Wilcoxon matched pairs test, *p* < 0.01)

Agreement of the three-item and four-item total scores at T1 and T2 are displayed in Table 4 and Fig. 1a, b. Visual inspection of these plots suggested no bias for the three-item total score. However, a slightly decreasing trend with increasing mean scores is found in the plot, suggesting some bias for the four-item total score. The limits of agreement of the two scores were similar, both showing that large score differences (effect size > 1.1) are needed to exceed these limits of agreement and thereby indicate change beyond random chance at the individual level.

Table 3 Test–retest reliability QOL-BDS items and total scores by age and etiology (mixed model; absolute agreement)

| | Age | | | | Etiology | | | | | |
| | Up to 50 | | 50 and older | | SCD | | SCI | | All | |
	ICC	CI	ICC	CI	ICC	CI	ICC	CI	ICC	CI
Life as whole	0.68	0.43–0.83	0.65	0.43–0.80	0.65	0.45–0.78	0.70	0.39–0.86	0.66	0.49–0.78
Physical health	0.52	0.22–0.73	0.80	0.66–0.88	0.71	0.55–0.82	0.63	0.31–0.82	0.69	0.55–0.79
Psychol health	0.83	0.69–0.91	0.75	0.58–0.85	0.84	0.74–0.90	0.68	0.38–0.85	0.80	0.70–0.86
Social life	0.81	0.65–0.90	0.70	0.52–0.82	0.77	0.64–0.87	0.71	0.44–0.86	0.76	0.65–0.84
Total three items	0.81	0.66–0.90	0.85	0.75–0.92	0.86	0.77–0.91	0.76	0.53–0.89	0.83	0.75–0.89
Total four items	0.83	0.68–0.91	0.84	0.73–0.91	0.85	0.75–0.91	0.79	0.57–0.90	0.83	0.75–0.89

Table 4 Bland–Altman analysis of agreement between QOL-BDS scores on T1 and T2

	Three-item total score	Four-item total score
Range of differences T1–T2	−3–2.7	−2.8–3.8
Mean of differences T1–T2 (SD)	0.18 (1.05)	0.20 (1.08)
95% CI of difference T1–T2 (*p* value)	−0.04–0.41 (0.111)	−0.03–0.44 (0.088)
Limits of agreement	−1.86–2.23	−1.91–2.32
Effect size needed to exceed chance (individual)	1.08	1.12
Effect size needed to exceed chance (group)	0.12	0.12
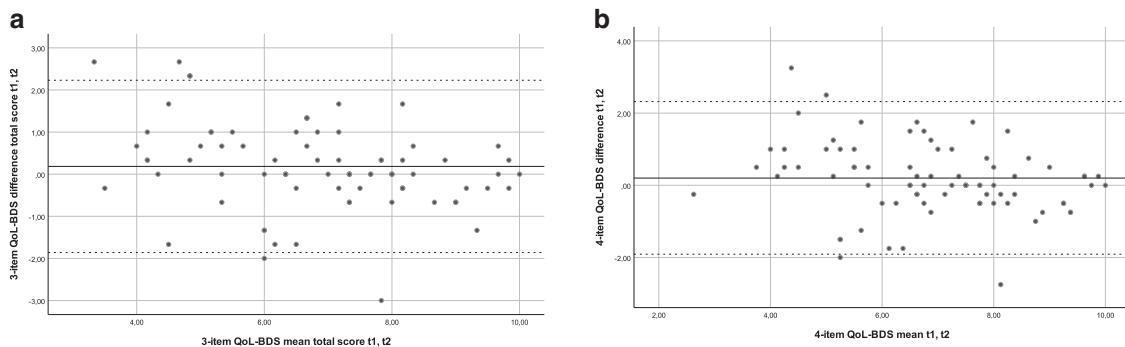
Fig. 1 a, **b** Bland–Altman plots of the three-item and four-item scores

In contrast, small score differences (effect size > 0.1) are sufficient to show change beyond chance at group level.

Discussion

This is the first prospective, multicenter, international psychometric evaluation of the QoL-BDS. The results provide evidence of its reproducibility in an international sample and in subgroups based on age and etiology of SCI/D. Internal consistency of the three-item and four-item total scores were good. Test–retest reliability was substantial for the four single items and good for the three-item and four-item total scores. The Bland–Altman analyses showed that the QoL-BDS is sensitive to small changes in QoL at the group level, but not at the individual level.

The addition of a fourth item on satisfaction with social life did not impact the reproducibility of the QoL-BDS. Also, the score distributions of the three-item and four-item versions were remarkably similar. We did not formulate a-priori hypotheses about possible differences, but it is reassuring that the extra item did not increase the heterogeneity of the measure. Future studies should clarify whether or not the extra item on satisfaction with social life increases the validity of the QoL-BDS.

The ICC values for the individual items in this study (0.66–0.80) were lower than the range of ICC's of 0.86–0.94 found in a Dutch study of rehabilitation inpatients [12]. The short time between tests in the latter study (median 4 days) could explain the higher reliability found in that study. We could not find other studies to compare our results with.

The Bland–Altman analysis showed that the QoL-BDS is likely to be sensitive to change when used in trials or longitudinal cohort studies because the analysis would be conducted at a group level. The Bland–Altman analysis, however, also showed that large changes in scores are required to exceed chance at individual level. This implies that the QoL-BDS as currently designed might not be sufficiently sensitive to change in clinical practice. Repeated administrations would be recommended to increase its sensitivity to change in clinical practice. The SPSS-output showed that the ICC of the average of the two test administrations would be 0.91, which is above the recommended 0.90 required for use in individual patient care [16].

Limitations

A few limitations of this study should be noted. First, the sample size per country was small and below the recommended sample size of 30 [7]. Therefore, we refrained from comparing test–retest reliability between countries. In the forthcoming full validation analyses, assessment of differences by country will be a focus.

Also, for five participants the test–retest interviews were not administered by the same interviewer. This, however, hardly influenced the results: for example, the ICC of the four-item total score increases only from 0.83 to 0.84 after exclusion of these five participants.

Finally, although data were collected in four countries in different parts of the world, there was no representation of low-income or lower–middle income countries and countries from Africa, the Middle-East or Asia. It also remains unclear whether the test–retest reliability results of this study extend to the inpatient situation.

Conclusion

This study provides evidence of reproducibility of the QoL-BDS and suggests sensitivity to change in research investigations. The possible addition of an item on satisfaction with social life did not affect the internal consistency and reproducibility of this measure, and conceptually adds a dimension found to be important by persons with SCI. In the context of the continuing debate on the conceptualization and measurement of QoL, the QoL-BDS is a significant step toward unifying our ability to record and report this important information.

Data archiving

The data sets analyzed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

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

Statement of ethics The research protocol was approved by the University of Michigan Medical School Institutional Review Board, protocol number HUM00126164 and HUM00125769. For the Netherlands, permission to execute the study was granted by the Board of Directors of De Hoogstraat after positive advice of the Institute Review Board on 27 July 2017. For Brazil, the research protocol was approved by the Medical Ethics Committee of the Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo in April 2018. Approbation number CAAE:283112917.3.0000.0068. For Australia, the project was approved by the Alfred Hospital Ethics Committee on 14 June 2017 (project no 203/17). We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research.

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

References

1. Post MW, van Leeuwen CM. Psychosocial issues in spinal cord injury: a review. Spinal Cord. 2012;50:382–9.
2. Dijkers MP. Quality of life of individuals with spinal cord injury: a review of conceptualization, measurement, and research findings. J Rehabil Res Dev. 2005;42:87–110.
3. Post MW. Definitions of quality of life: what has happened and how to move on. Top Spinal Cord Inj Rehabil. 2014;20:167–80.
4. Charlifue S, Post MW, Biering-Sørensen F, Catz A, Dijkers M, Geyh S, et al. International spinal cord injury quality of life basic data set. Spinal Cord. 2012;50:672–5.
5. Biering-Sørensen F, Charlifue S, DeVivo M, Noonan V, Post M, Stripling T, et al. International spinal cord injury data sets. Spinal Cord. 2006;44:530–4.
6. Biering-Sørensen F, Noonan VK. Standardization of data for clinical use and research in spinal cord injury. Brain Sci 2016;6:29.
7. Biering-Sørensen F, Alexander MS, Burns S, Charlifue S, DeVivo M, Dietz V, et al. Recommendations for translation and reliability testing of International Spinal Cord Injury Data Sets. Spinal Cord. 2011;49:357–60.

8. Tate D, Forchheimer M. Review of cross-cultural issues related to quality of life after spinal cord injury. Top Spinal Cord Inj Rehabil. 2014;20:181–90.

9. Oyakawa A, Castineira C, Greve J, Furlan J, Forchheimer M, Tate D, et al. Reliability of the Portuguese version of the International Quality of Life Spinal Cord Injury Data Set. Abstract book 52nd annual meeting International Spinal Cord Society, 2103 (abstract 101).

10. Post MWM, Adriaansen JJE, Charlifue S, Biering-Sørensen F, van Asbeck FWA. Good validity of the international spinal cord injury quality of life basic data set. Spinal Cord. 2016;54:314–8.

11. New PW Tate DG, Forchheimer MB D'Andrea Greve JM, Parashar D. Post MWM preliminary psychometric analyses of the international spinal cord injury quality of life basic data set. Spinal Cord. 2019.

12. Nachtegaal J, van Langeveld SA, Slootman H, Post MWM. Dutch-Flemish Spinal Cord Society. Implementation of a standardized dataset for collecting information on patients with spinal cord injury. Top Spinal Cord Inj Rehabil. 2018;24:133–40. Spring

13. Wood-Dauphinee S, Exner G, Bostanci B, Glass C, Jochheim KA, Kluger P, et al. Quality of life in patients with spinal cord injury—basic issues, assessment, and recommendations. Restor Neurol Neurosci. 2002;20:135–49.

14. Post MW, Charlifue S, Biering-Sørensen F, Catz A, Dijkers MP, Horsewell J, et al. Development of the International Spinal Cord Injury Activities and Participation Basic Data Set. Spinal Cord. 2016;54:530–4.

15. Tate D, Forchheimer M, Charlifue S, Coker J, Bennett S, Huey S, et al. Cross cultural comparisons in patients self-reports of physical and psychological health and quality of life. Ann Phys Rehabil Med. 2018;61S:e43.

16. Nunnally JC Psychometric theory. New York: McGraw; 1967.

17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.

18. Hays RD, Anderson RT, Revicki D. Assessing reliability and validity of measurement in clinical trials. In: Staquet MJ, Hays RD, Fayers PM, eds. Quality of life assessment in clinical trials: methods and practice. Oxford: Oxford University Press; 1998. p. 169–83.

19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–10.

20. Cohen J. Statistical power for the behavioral sciences, 2nd edn. Hillsdale, NJ: Erlbaum; 1988.