Measuring Spinal Cord Injury

A Practical Guide of Outcome Measures

Giovanni Galeoto
Anna Berardi
Marco Tofani
Maria Auxiliadora Marquez
Editors

Springer

Measuring Spinal Cord Injury

Giovanni Galeoto
Anna Berardi • Marco Tofani
Maria Auxiliadora Marquez
Editors

Measuring Spinal Cord Injury

A Practical Guide of Outcome Measures

Springer

Editors
Giovanni Galeoto
Department of Human Neurosciences
Sapienza University of Rome
Rome
Italy

Anna Berardi
Department of Human Neurosciences
Sapienza University of Rome
Rome
Italy

Marco Tofani
Department of Neurorehabilitation and Robotics
Bambino Gesù Paediatric Hospital
Rome
Italy

Maria Auxiliadora Marquez
Universidad Fernando Pessoa-Canarias
Las Palmas
Spain

ISBN 978-3-030-68381-8 ISBN 978-3-030-68382-5 (eBook)
https://doi.org/10.1007/978-3-030-68382-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction on Measuring Spinal Cord Injury

Giorgio Scivoletto, Giovanni Galeoto, Marco Tofani, Anna Berardi, and Maria Auxiliadora Marquez

1 Introduction

Every year, around the world, between 250,000 and 500,000 people suffer a spinal cord injury (SCI). There is no reliable estimate of global prevalence, but the estimated annual global incidence is 40–80 cases per million. Traffic accidents were typically the most common cause of SCI, followed by falls in the elderly population [1]. Substantial variation in mortality and longevity within the SCI population, compared to the general population, and between World Health Organization regions and country income level was found. People with an SCI are two to five times more likely to die prematurely than people without an SCI, with worse survival rates in low- and middle-income countries [2].

Living with an SCI requires strategies to face a wide range of health-related problems. Apart from the paralysis, problems in various body functions, such as the bladder, bowel and sexual function, autonomic function, and pain, will be of concern. Functional problems can lead to limitations in activities and participation restrictions typically related to mobility, self-care activities, difficulties in regaining work, maintaining social relationships, participating in leisure activities, and being active members of the community [3, 4]. Participation restrictions are highly dependent on environmental factors, such as mobility equipment and transportation [5]. After an SCI, long-term functional outcomes result in a combination of acute neurological recovery and medical intervention, rehabilitation, and social participation [6]. Establishing the value of medical interventions from multiple parties' perspective is essential to the availability and adoption of therapies [7]. This study was conducted by a research group comprised of medical doctors and health professionals from the "Sapienza" University of Rome and from "Rehabilitation & Outcome Measure Assessment" (R.O.M.A.) association. R.O.M.A. association in the last few years has dealt with several studies and validated Italy's outcome measures for the spinal cord injury population [8–21].

The research group aims to develop a systematic review to identify all the evaluation tools developed and validated to measure different aspects of SCI. This study aims to provide clini-

G. Scivoletto
Foundation Santa Lucia, Rome, Italy

G. Galeoto · A. Berardi
Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

M. Tofani (✉)
Department of Neurorehabilitation and Robotics, Bambino Gesù Paediatric Hospital, Rome, Italy
e-mail: marco.tofani@uniroma1.it

M. Auxiliadora Marquez
Universidad Fernando Pessoa-Canarias, Las Palmas, Spain

G. Galeoto et al. (eds.), *Measuring Spinal Cord Injury*,
https://doi.org/10.1007/978-3-030-68382-5_1

cians and researchers information regarding the existing outcome measures to assess people with SCI based on reviewing, analyzing, comparing, and critically appraising the available outcome measures and their distribution in the international literature.

2 Spinal Cord and International Classification of Functioning Disability and Health

The International Classification of Functioning, Disability, and Health (ICF) [22] provides a comprehensive and universally accepted framework to classify and describe functioning, disability, and health. According to the ICF framework, the problems associated with a disease may involve *body functions and body structures* and the *activities and participation* in life situations. Health status and disability are modified by contextual factors such as environmental and personal factors [22]. According to ICF, many people living with SCI face different body function problems and structure domains according to lesion levels and according to early post-acute or long-term settings. These differences also appear in the activity and participation domains. New efforts to understand differences within people living with SCI, according to the ICF framework, were made.

Table 1 describes the ICF core set for acute SCI [23]. The formal consensus process integrating evidence from preparatory studies and expert knowledge at the ICF Core Set conference for SCI led to the definition of a Comprehensive ICF Core Set for SCI in the early post-acute context for multidisciplinary assessment and clinical studies.

A formal consensus process integrating evidence and expert opinion based on the ICF framework and classification led to the definition of ICF core sets for SCI in the long-term context. The brief ICF core set includes 33 second-level categories that were selected out of the second-level categories of the comprehensive ICF core set, using a two-step ranking procedure and a final cutoff decision. Table 2 describes the ICF core set for people living with SCI in a long-term context [24].

Table 1 ICF core set for acute SCI adapted and modified from Kirchberger et al. [23]

Body functions	b730	Muscle power functions
	b620	Urination functions
	b525	Defecation functions
	b280	Sensation of pain
	b440	Respiration functions
	b735	Muscle tone functions
	b152	Emotional functions
	b810	Protective functions of the skin
Body structures	s120	Spinal cord and related structures
	s430	Structure of respiratory system
	s610	Structure of urinary system
Activities and participation	d420	Transferring oneself
	d410	Changing basic body position
	d445	Hand and arm use
	d530	Toileting
	d550	Eating
	d450	Walking
	d510	Washing oneself
	d540	Dressing
	d560	Drinking
Environmental factors	e310	Immediate family
	e355	Health professionals

Tables 1 and 2 just described well represent the needs of people living with an SCI during the early phase or the life-span perspective. Researchers, clinicians, and students should approach the ICF framework for clinical practice, research, and health reporting.

3 Outcome Measures

An outcome measure is a tool used to investigate a different aspect of patients' status. An outcome measure can provide baseline data and/or provide information on patient change during recovery and/or rehabilitation. According to the ICF framework, different outcome measures could investigate body function and structure, activity and

Table 2 ICF core set for people living with SCI in a long-term context adapted and modified from Cieza et al. [24]

Body functions	b730	Muscle power functions
	b620	Urination functions
	b280	Sensation of pain
	b525	Defecation functions
	b640	Sexual functions
	b810	Protective functions of the skin
	b735	Muscle tone functions
	b710	Mobility of joint functions
	b152	Emotional functions
Body structures	s120	Spinal cord and related structures
	s610	Structure of urinary system
	s810	Structure of areas of skin
	s430	Structure of respiratory system
Activities and participation	d530	Toileting
	d420	Transferring oneself
	d230	Carrying out daily routine
	d465	Moving around using equipment
	d410	Changing basic body position
	d445	Hand and arm use
	d470	Using transportation
	d455	Moving around
	d520	Caring for body parts
	d550	Eating
	d240	Handling stress and other psychological demands
Environmental factors	e310	Immediate family
	e120	Products and technology for personal indoor and outdoor mobility and transportation
	e115	Products and technology for personal use in daily living
	e150	Design, construction, and building products and technology of buildings for public use
	e155	Design, construction, and building products and technology of buildings for private use
	e110	Products or substances for personal consumption
	e355	Health professionals
	e340	Personal care providers and personal assistants
	e580	Health services, systems, and policies

participation, environmental and personal factors. Some of the assessment tools commonly used in clinical practice can also link to the ICF domains. Regardless of the area investigated, the outcome measures have a different structure and basic concept. For a comprehensive overview, the main categories of outcome measures are presented:

3.1 Patient-Reported Outcome Measure (PROM)

Self-report measures are typically captured in the form of a questionnaire. The questionnaires are scored by applying a predetermined point system to the patient's responses. Although self-report measures seem subjective, PROM objectifies a patient's perception. PROM can be pencil-based or in electronic format. When examining equivalence between paper and electronic versions, formats are usually judged by authors to be equivalent [25].

3.2 Clinician-Reported Outcomes (ClinRo)

The use of a ClinRO assessment requires specialized professional training to evaluate the patient's health status. Conducted and reported by a trained health care professional, the ClinRO assessment reflects the evaluation of a patient's reported condition. A ClinRO assessment that is an appropriate outcome assessment in one context of use may or may not be adequate in a different context. ClinRO assessments are commonly used in end points that form the basis for reviewing and approving medical interventions [26].

3.3 Observer-Reported Outcomes (ObsRo)

An ObsRO is a measurement based on an observation by someone other than the patient or a health professional. In general, ObsRO is reported by a parent, caregiver, or someone who observes the patient in daily life and is particularly useful for patients who cannot report for themselves (e.g., infants or individuals who are cognitively impaired) [27].

3.4 Performance-Based Outcome Measure (PbOM)

Performance-based measures involve presenting examinees with functional tasks in a standardized format. The clinician does not apply judgment to quantifying the performance but is administering and monitoring the performance of the PbOM. A performance-based measure entails having the patient prepare a meal using a lab/hospital-based mock kitchen. For instance, rather than merely asking a patient or a collateral informant about the patient's cooking skills and safety (limited by their insight, candor, and objectivity), or going to the patient's home and watching them prepare a meal (while potentially valuable and informative, can be impractical), the patient's performance of a defined task is quantified in a specified way, and it does not rely on judgment to determine the rating [28].

4 International Project for Measuring SCI

"Good science and good clinical practice depend upon sound information, which in turn relies on sound measurement. Measurement enables health care professionals and researchers to describe, predict and evaluate in order to provide benchmarks and summarize change related to the condition and care of individuals with spinal cord injury." This is the incipit to the Spinal Cord Injury Rehabilitation (SCIRE) collaboration (https://scireproject.com/). SCIRE outcome measures provide information on the psychometric properties and the clinical use of 104 measures, giving the reader the necessary confidence to move their clinical practice and research forward on a more rigorous basis. The project started to analyze and search different outcome measures in studies involving the SCI population investigating psychometric properties, namely reliability, validity, and responsiveness. SCIRE's efforts provided in these years the development of the spinal cord injury outcome measures toolkit: a standardized set of outcome measures for use in the SCI practice consisting of 33 outcome measures that have been psychometrically validated for the SCI population. The outcome measures are grouped into different categories: assistive technology, community integration, lower limb and walking, mental health, neurological impairment and autonomic dysfunction, pain, quality of life and health status, self-care and daily living, skin health, spasticity, upper limb and wheeled mobility.

Another concrete effort to spread evidence-based practice using validated outcome measures is the Rehabilitation Measures Database (www.sralab.org/rehabilitation-measures). The database is organized for a specific health condition and can be filtered according to specific categories (e.g., assessment type, area of assessment, costs). There is a specific section for the SCI population; however, for each category, the database reports the purpose of the scale, time, and administration, psychometric properties, population, references.

In conclusion, the international landscape is engaging in promoting using valid outcome measures that can lead to better quality of care, reduced health care costs, and promote a continuous evaluation of the effectiveness of health interventions. Health care workers are called to improve their work and credentials continually. The adaptation of valid outcome measures is an ethical and moral issue and a duty enshrined in their individual professional profiles. The present manual aims to propose a critical and systematic analysis of the validated outcome measures for the SCI population at international levels. We hope that less experienced readers will be inspired to approach their work with seriousness and determination. Instead, we trust that more experienced readers will enthusiastically grasp our effort to systematize the outcome measures available to the population with SCI. Our hope is to create an international synergy to improve the care process and the quality of people's lives.

References

1. Singh A, Tetreault L, Kalsi-Ryan S, Nouri A, Fehlings MG. Global prevalence and incidence of traumatic spinal cord injury. Clin Epidemiol. 2014. https://doi.org/10.2147/CLEP.S68889.
2. Chamberlain JD, Meier S, Mader L, Von Groote PM, Brinkhof MWG. Mortality and longevity after a spinal cord injury: systematic review and meta-analysis. Neuroepidemiology 2015. https://doi.org/10.1159/000382079
3. Gerhart KA, Bergstrom E, Charlifue SW, Menter RR, Whiteneck GG. Long-term spinal cord injury: functional changes over time. Arch Phys Med Rehabil. 1993. https://doi.org/10.1016/0003-9993(93)90057-H.
4. Lidal IB, Huynh TK, Biering-Sørensen F. Return to work following spinal cord injury: a review. Disabil Rehabil. 2007. https://doi.org/10.1080/09638280701320839.
5. Whiteneck G, Meade MA, Dijkers M, Tate DG, Bushnik T, Forchheimer MB. Environmental factors and their role in participation and life satisfaction after spinal cord injury. Arch Phys Med Rehabil. 2004. https://doi.org/10.1016/j.apmr.2004.04.024.
6. McKinley W, Meade MA, Kirshblum S, Barnard B. Outcomes of early surgical management versus late or no surgical intervention after acute spinal cord injury. Arch Phys Med Rehabil. 2004. https://doi.org/10.1016/j.apmr.2004.04.032.
7. Walton MK, Powers JH, Hobart J, et al. Clinical outcome assessments: conceptual foundation-report of the ISPOR clinical outcomes assessment-emerging good practices for outcomes research task force. Value Heal. 2015. https://doi.org/10.1016/j.jval.2015.08.006.
8. Castelnuovo G, Giusti EM, Manzoni GM, et al. What is the role of the placebo effect for pain relief in neurorehabilitation? Clinical implications from the Italian consensus conference on pain in neurorehabilitation. Front Neurol. 2018. https://doi.org/10.3389/fneur.2018.00310.
9. Marquez MA, De Santis R, Ammendola V, et al. Cross-cultural adaptation and validation of the "spinal cord injury-falls concern scale" in the Italian population. Spinal Cord. 2018;56(7):712–8. https://doi.org/10.1038/s41393-018-0070-6.
10. Berardi A, De Santis R, Tofani M, et al. The Wheelchair Use Confidence Scale: Italian translation, adaptation, and validation of the short form. Disabil Rehabil Assist Technol. 2018;13(4) https://doi.org/10.1080/17483107.2017.1357053.
11. Anna B, Giovanni G, Marco T, et al. The validity of rasterstereography as a technological tool for the objectification of postural assessment in the clinical and educational fields: pilot study. In: Advances in intelligent systems and computing; 2020. https://doi.org/10.1007/978-3-030-23884-1_8.
12. Panuccio F, Berardi A, Marquez MA, et al. Development of the pregnancy and motherhood evaluation questionnaire (PMEQ) for evaluating and measuring the impact of physical disability on pregnancy and the management of motherhood: a pilot study. Disabil Rehabil. 2020:1–7. https://doi.org/10.1080/09638288.2020.1802520.
13. Amedoro A, Berardi A, Conte A, et al. The effect of aquatic physical therapy on patients with multiple sclerosis: a systematic review and meta-analysis. In: Mult Scler Relat Disord; 2020. https://doi.org/10.1016/j.msard.2020.102022.
14. Dattoli S, Colucci M, Soave MG, et al. Evaluation of pelvis postural systems in spinal cord injury patients: outcome research. J Spinal Cord Med. 2018;43:185–92.
15. Berardi A, Galeoto G, Guarino D, et al. Construct validity, test-retest reliability, and the ability to detect change of the Canadian occupational performance measure in a spinal cord injury population. Spinal Cord Ser Cases. 2019. https://doi.org/10.1038/s41394-019-0196-6.
16. Ponti A, Berardi A, Galeoto G, Marchegiani L, Spandonaro C, Marquez MA. Quality of life, concern of falling and satisfaction of the sit-ski aid in sit-skiers with spinal cord injury: observational study. Spinal Cord Ser Cases. 2020. https://doi.org/10.1038/s41394-020-0257-x.
17. Panuccio F, Galeoto G, Marquez MA, et al. General sleep disturbance scale (GSDS-IT) in people with spinal cord injury: a psychometric study. Spinal Cord. 2020. https://doi.org/10.1038/s41393-020-0500-0.
18. Monti M, Marquez MA, Berardi A, Tofani M, Valente D, Galeoto G. The multiple sclerosis intimacy and sexuality questionnaire (MSISQ-15): validation of the Italian version for individuals with spinal cord injury. Spinal Cord. 2020. https://doi.org/10.1038/s41393-020-0469-8.
19. Galeoto G, Colucci M, Guarino D, et al. Exploring validity, reliability, and factor analysis of the Quebec user evaluation of satisfaction with assistive technology in an Italian population: a cross-sectional study. Occup Ther Heal Care. 2018. https://doi.org/10.1080/07380577.2018.1522682.
20. Colucci M, Tofani M, Trioschi D, Guarino D, Berardi A, Galeoto G. Reliability and validity of the Italian version of Quebec user evaluation of satisfaction with assistive technology 2.0 (QUEST-IT 2.0) with users of mobility assistive device. Disabil Rehabil Assist Technol. 2019. https://doi.org/10.1080/17483107.2019.1668975.
21. Berardi A, Galeoto G, Lucibello L, Panuccio F, Valente D, Tofani M. Athletes with disability' satisfaction with sport wheelchairs: an Italian cross sectional study. Disabil Rehabil Assist Technol. 2020. https://doi.org/10.1080/17483107.2020.1800114.
22. World Health Organization. The ICF: an overview. Geneva: WHO; 2001.

23. Kirchberger I, Cieza A, Biering-Sørensen F, et al. ICF Core sets for individuals with spinal cord injury in the early post-acute context. Spinal Cord. 2010. https://doi.org/10.1038/sc.2009.128.
24. Cieza A, Kirchberger I, Biering-Sørensen F, et al. ICF Core sets for individuals with spinal cord injury in the long-term context. Spinal Cord. 2010. https://doi.org/10.1038/sc.2009.183.
25. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. Qual Life Res. 2015. https://doi.org/10.1007/s11136-015-0937-3.
26. Powers JH, Patrick DL, Walton MK, et al. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR clinical outcome assessment emerging good practices task force. Value Heal. 2017. https://doi.org/10.1016/j.jval.2016.11.005.
27. Benjamin K, Vernon MK, Patrick DL, Perfetto E, Nestler-Parr S, Burke L. Patient-reported outcome and observer-reported outcome assessment in rare disease clinical trials: an ISPOR COA emerging good practices task force report. Value Heal. 2017. https://doi.org/10.1016/j.jval.2017.05.015.
28. Moore DJ, Palmer BW, Patterson TL, Jeste DV. A review of performance-based measures of functional living skills. J Psychiatr Res. 2007. https://doi.org/10.1016/j.jpsychires.2005.10.008.

Psychometric Properties of Assessment Tools

Marco Monticone, Giovanni Galeoto, Anna Berardi, and Marco Tofani

1 Introduction

Clinicians and researchers are encouraged to proficiently analyze the psychometric properties of assessment tools essential for evaluating patients with medical problems [1]. Advances in diagnosis and care are possible when appropriate assessment tools are made available to decision makers. They know the measure used is adequate for its purpose, how it compares with similar measures, and how to interpret findings [2]. For every patient or population group, several instruments can be used to evaluate clinical conditions or health status. However, many instruments have been poorly or incompletely validated over time, thus limiting their use for specific diseases or among populations and countries [3]. Assessment is common in the medical sciences and varies from questions asked during history-taking to physical evaluations, imaging techniques, laboratory tests, or self-reported questionnaires.

M. Monticone
Physical Medicine and Rehabilitation, Department of Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy

G. Galeoto (✉) · A. Berardi
Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy
e-mail: giovanni.galeoto@uniroma1.it

M. Tofani
Department of Neurorehabilitation and Robotics, Bambino Gesù Paediatric Hospital, Rome, Italy

Irrespective of the tool clinicians and researchers may wish to select, a correct way of assessment aims to replace an empirical approach with a scientific methodology, therefore increasing the effectiveness of everyday practice [4].

This study was conducted by a research group comprised of medical doctors and health professionals from the "Sapienza" University of Rome and "Rehabilitation & Outcome Measure Assessment" (R.O.M.A.) association. R.O.M.A. association in the last few years has dealt with several studies and validated outcome measures in Italy for the spinal cord injury (SCI) population [5–18].

This chapter serves as a synthetic guide to present the core psychometric properties of measurement instruments in the medical field. The method of assessments, strengths, and criticisms for each psychometric property is directly based on current literature in the field. Terminology and definitions based on recent consensus-based standards for the selection of health measurement instruments (also known as COSMIN) are reported throughout the chapter [19].

2 Reliability

This method is defined as "the degree to which the measurement is free from measurement error" [19], and when repeated measurements are conducted, it is worthy of investigation. Reliability

G. Galeoto et al. (eds.), *Measuring Spinal Cord Injury*,
https://doi.org/10.1007/978-3-030-68382-5_2

varies depending on issues that include the instrument under investigation, the evaluators, and the patients under study. These possibilities led to different types of reliability: (1) test–retest reliability, when measurements are repeated over time; (2) inter-rater reliability, when they are conducted by different evaluators but on the same occasion; (3) intra-rater reliability, when they are conducted by the same evaluator but on different occasions, and (4) internal consistency, when different sets of items from the same tool are employed [1, 19].

When test–retest, inter, and intra-rater reliability are addressed, readers should know they state the way evaluations can be distinguished from each other despite the presence of measurement error. Statistics to calculate test–retest, inter, and intra-rater reliability vary according to variables adopted. When continuous variables are studied, the adoption of intraclass correlation coefficients, which consist of a ratio of variances, is recommended [20]. The coefficient values vary from 0 (i.e., the error variance is considered negligible compared to patient variance) to 1 (i.e., the error variance is very large regarding patient variance as may happen when homogenous samples are encountered). A coefficient value of 0.70 is considered acceptable, and values of 0.80 and 0.90 as good and very good [21]. When categorical variables are studied, literature advises the use of Cohen's kappa, which adjusts for the agreement expected by chance, calculated by assuming independence of measurements as obtained by multiplications of the marginals. Estimates vary from −1 to 1: figures equal to 1 state there is a perfect agreement; figures of 0 mean there is no agreement, which can be expected by chance; figures near to −1 are usually caused by reversed scaling by one of the two raters. Cohen's kappa may be influenced by sample differences, marginals distribution, number of classes, and between-raters' systematic differences [22].

Measurement error constitutes a related but different concept from reliability and corresponds to the difference between an amount that can be measured and its true value [23]. It can be calculated in three ways. First, the standard error of measurement corresponds to the standard deviation around a single measurement. It is a measure of how far apart the findings of repeated measures are. Clinicians and researchers easily interpret it as it is reported in the reference unit of the tool under study [24]. It is calculated by the following formula:

$$\text{SEM} = \text{SD}\sqrt{1 - \text{ICC}_{2,1}}$$

where ICC corresponds to test–retest reliability of the reference population as assessed by intraclass coefficient correlation statistics.

Second, the limits of agreement represent a graphical method to compare two measurements where differences between the two techniques are plotted against the averages of the two techniques. Relating the limits of agreement to the tool range may give an impression of the magnitude of the measurement error. By definition, 95% of the differences between repeated measurements fall between the limits of agreement [1, 24]. Third, by the coefficient of variation, which relates the standard deviation of repeated measurements to the mean value, with higher percentage figures representing higher heterogeneity [1, 24].

Internal consistency is defined as "the degree of interrelatedness among the items" and represents the level to which items belonging to an assessment tool assess the same construct [19]. There are three parameters for calculating internal consistency: Cronbach's alpha, inter-item correlations, and item-total correlations. The first determines how closely related a set of items are as a group, with values greater than 0.7 and 0.8 showing acceptable and good internal consistency, respectively. The second indicates whether an item is part of the assessment tool, with values higher than 0.7 suggesting items evaluating the same construct. The third gives an indication of whether the items discriminate patients on the construct under investigation, with figures below 0.3 suggesting a low contribution to distinctions [25].

Sample size. A minimum number of 50 patients are required in order to calculate reliability by avoiding the risk of bias due to insufficient populations [26].

3 Validity

Validity is defined as "the degree to which an instrument truly measures the construct it purports to measure" [1, 19]. An adequate definition of the construct (i.e., an explanatory variable to be measured which is not directly observable) is imperative. The construct itself has to be part of the conceptual model within a theoretical and clinical framework. There are three different types of validity: content validity, criterion validity, and construct validity.

3.1 Content Validity

This has been defined as "the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured" [19]. It has been recommended as the starting point of each validation process [3]. With special reference to multi-item measures, content validity aims to investigate their relevance and comprehensiveness about the construct under study. The first issue evaluates if all items refer to relevant aspects of the construct to be measured, whether they are all relevant for the population being studied, and if they are really relevant for the object of the tool usage. The second issue investigates if the construct chosen is entirely covered by the items. Content validity can be qualitatively evaluated by an independent panel of independent experts to avoid the risk of bias. A full description of the outcome measure, including procedures of administration, has to be warranted [27].

An additional aspect of the content validity is represented by face validity, which has been defined as "the degree to which a measurement instrument looks, indeed, as though it is an adequate reflection of the instrument to be measured."[19] It stands for an overall view of the tool and it is related to a subjective assessment. This property is still undervalued because no standards as to how it should be evaluated are clearly recommended.

3.2 Criterion Validity

It has been defined as "the degree to which the scores of a measurement instrument are an adequate reflection of a gold standard (i.e., diagnostic test that is regarded as definitive in determining whether an individual has a disease process)" [19]. Criterion validity can be subdivided into two main sides: (1) concurrent validity and (2) predictive validity. When assessing the first side, clinicians and researchers take into account, and at the same time, the score deriving from the gold standard they have chosen and that from the measurement under investigation. When the second side is addressed, clinicians and researchers aim to know whether the outcome measure under study forecasts the gold standard. The hypothesis is that the outcome measure should perform as efficiently as the gold standard for both sides. Moreover, an appropriate target population should be individuated. The level of agreement of the gold standard and the instrument under study should be stated as early as possible, and the scores should be obtained independently. Statistics to apply depends again on the variables included: criterion validity is expressed through sensitivity and specificity estimates whether the gold standard and the instrument under study show dichotomous outcomes; receiver operating characteristics curves are advised if the tool relies upon an ordinal or a continuous scale; correlation coefficients are adopted if a continuous gold standard variable is represented [1]. Sample size: a minimum number of 50 patients are required [26].

3.3 Construct Validity

It has been defined as "the degree to which the scores of a measurement instrument are consistent with hypotheses, e.g., with regard to internal relationships, relationships with scores of other instruments or differences between relevant groups" [19]. In other terms, evaluating this property, clinicians and researchers state a measurement tool that validly assesses the construct under investigation. There are three subtypes of construct validity: (1) structural validity, (2) hypothesis testing, and (3) cross-cultural validity.

The first subtype is defined as "the degree to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured" [19]. Overall structural validity can be assessed by factorial analyses that mainly include exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is chosen when there are no clear ideas on the factorial characteristics (i.e., number and types of dimensions) composing a multi-item instrument. An initial analysis is initially performed using Cattel's scree test to determine the line plot of extracted factors with eigenvalues greater than 1 selected. Furthermore, orthogonal (also known as Varimax) rotation of the item is habitually applied, leading to defining a component matrix made of all the items under study. Those with loadings on dimensions greater than 0.50 are included in the factor. The expected explained variance by this factor analysis should be more than 50% to be considered acceptable [28]. CFA is implemented when, based on previous researches and findings, predefined hypotheses on dimensions are available. Each item needs to be specified to load onto its subscale, as originally described. Model fit is assessed using the ratio between the χ^2-test and degrees of freedom (i.e., χ^2/d.f.), the comparative fit index, the normed fit index, and the root-mean square error of approximation and its 90% confidence intervals [29]. The following thresholds are considered to represent a good fit: χ^2/d.f. < 3, comparative fit index ≥ 0.90, normed fit index ≥ 0.90 and root-mean square error of approximation ≤ 0.08 [30]. Sample size: from 4 to 10 patients per item with a minimum of 100 patients are required [26].

The second subtype of construct validity occurs when hypotheses are formulated a priori on the relationships of scores on the instrument under investigation with scores deriving from other measures evaluating related or dissimilar constructs. One idea suggests formulating a series of hypotheses between the tool under study and related measures, describing the expected direction (i.e., positive or negative) and magnitude (i.e., small, moderate, large) of the above relationships. It is important to evaluate if the findings are consistent with the preformulated hypotheses by counting how many are established and how many refuted (usually they are expressed in percentage), and discuss the findings [31]. Sample size: a minimum of 50 patients are recommended are preferred [26].

The third subtype (i.e., cross-cultural validity) is defined as "the degree to which the performance of the items on a translated or culturally adapted tool is an adequate reflection of the performance of items in the original version of the instrument" [19]. This issue is important after the translation of a questionnaire. There may be differences in cultural aspects; some items may be irrelevant in other cultures, causing the risk of lowering the subsequent psychometric assessment [32, 33]. The first step is represented by the *forward translation*, where items are translated to retain the tool's original concepts. Two professional translations conduct two independent translations and ensure the language is compatible with a reading age of 14 years. Discrepancies are resolved through reconciliation which ends when a common adaptation is agreed. The second step is defined as *backward translation*. Two bilingual translators independently back-translate the initial translation; the researchers review these translations and ensure the adapted version essentially reflects the same item content as the original version. The third step is the *evaluation of the pre-final version by an expert committee*, where the translations are submitted to a bilingual committee of clinicians and methodologists who explore items semantic, idiomatic, and conceptual equivalence and answers to identify any discrepancies or mistakes. This third step ends when a pre-final version is agreed upon. A fourth step is represented by the *on-field test of the pre-final version*, where the pre-final version is tested to assess the comprehensibility and cognitive equivalence of the translation; this is pursued by cognitive interviews performed by trained psychologists (and eventually by health carers) who administer the instrument to selected subjects. At conclusion, the expert committee reviews the results, identifying any modification to improve the adapted form further. Finally, the step consisting of the *evaluation of the process by the developers*, where the shared version of the questionnaire is sent to the developers to receive further suggestions and final approval. All of the above steps rely on qualitative assessment. However, the new tool actually performs as the original version in different populations can be tested through factor analysis methods (especially CFA) or logistic regression analysis techniques [1].

4 Responsiveness

This psychometric property has been defined as "the ability of an instrument to detect change over time in the construct to be measured" [19], as it is important to know if the clinical status of patients has changed over time. When a tool shows to be responsive to change if patients change on the construct of interest, their scores on the measurement tool assess this construct change accordingly. Responsiveness is crucial with longitudinal studies and when evaluation scopes are pursued.

There are two ways of assessing responsiveness: almost similar to what above describes validity but based on a criterion that uses a construct approach. When a gold standard for change is available, a criterion approach is recommended to assess the degree in the scores of the tool under study is an adequate reflection of changes in scores of the gold standard. An appropriate target population should be individuated. The level of agreement between changes of the gold standard and the instrument under study should be defined. The scores should be obtained independently and over the same time period. Coefficients correlations are used when the gold standard is a continuous (i.e., change in score) variable and receiver operating characteristics (ROC) curves when it is a dichotomous (i.e., change vs. no change) variable. The area under the ROC curve is considered to measure an instrument's ability to distinguish between patients who are considered improved (or deteriorated) from those who are not considered improved (or deteriorated) in relation to the gold standard. An area under the curve of at least 0.70 is considered an appropriate responsive measure [34]. Hypothesis testing is useful to test the responsiveness of a measure when no gold standard is available. The hypotheses test correlations between changes in scores of the tool under study and changes in scores on other measures with satisfactory responsiveness. Relative correlations (i.e., comparisons between comparisons) can also be considered. The hypotheses should include the expected direction and magnitude between the change scores [34–36]. Previous studies should help formulate hypotheses.

4.1 Inappropriate Measures of Responsiveness

A series of methods to evaluate responsiveness, such as effect size, Guyatt's approach, the standardized response means, and the paired *t*-test, have been suggested and widely used over time. Effect size calculates on a whole sample by dividing the pre- and posttest scores by the pretest standard deviation (SD). As for Guyatt's approach, the change computed on the whole sample gets divided by the pretest SD calculated only for those subjects whose status remains unchanged. The standardized response mean (also known as the responsiveness treatment coefficient or efficacy index) is the ratio between individual change and the SD of that change. As for all of the measures, estimates of 0.20, 0.50, and 0.80 represent small, moderate, and large changes. However, evidence concludes they are inadequate measures of responsiveness as they express the magnitude of the change scores and not the validity of those changes. The paired *t*-test measures the change scores' statistical significance and not, again, the validity of the change scores [35, 36].

5 Interpretability

Interpretability has been defined as "the degree to which one can assign qualitative meaning, that is clinical or commonly understood connotations, to an instrument's qualitative scores or change in scores" [19]. It is crucial for every measurement instrument and a powerful information for clinicians and researchers as it refers to what the scores on an instrument mean. The minimum detectable change and the minimal important change get reported when addressing the meaning of change scores.

5.1 Minimum Detectable Change

It is the change beyond measurement error. In other words, it corresponds to a change that falls outside the limits of agreement as calculated by the Bland and Altman method [37] and can be estimated by the following formula:

$$\text{MDC} = \text{SEM} * z \ \text{ value} * \sqrt{2}$$

where SEM is the standard error of measurement, and z value corresponds to 1.96 or 1.64 when 95% or 90% confidence levels are chosen, respectively.

If a subject achieves a change score greater than the threshold estimated as per the MDC, it is possible to state (with % confidence) this change is real and not due to measurement error. A smaller change in score should be attributed to measurement error [24].

5.2 Minimal Important Change (MIC)

It is defined as "the smallest change in score in the construct to be measured which patients perceive as important" [19]. When outcomes based on patients' perspectives are considered, the MIC should consider their perspective, while when different instruments are used the clinician's point of view is of importance [38, 39]. MIC is estimated by anchor-based or distribution-based methods.

Anchor-based methods. They utilize an external criterion, or anchor, to determine clinically important improvements or to worsen. A globally perceived effect for patients' or clinicians' use and evaluated by the question: "Overall, how much did the treatment you received help your current problem?" or "Overall, how much did the treatment you delivered help your patient's current problem?"; then the perceived effect is determined using a Likert-type scale characterized by improvement levels (e.g., "it helped a lot" and "it helped"), no change level (e.g., "it did not help") and worsening levels (e.g., it made things worse; it made things much worse) [40]. A first method based on the use of an anchor is the mean change method, where the minimal important change corresponds to the mean change in score on the measurement instrument in the subcategory of patents who are minimally importantly changed. A second method is based on the receiver operating characteristics curves: subjects are dichotomized into two groups based on the global perceived effect scores (e.g., improved vs. non-improved; improved vs. worsened); sensitivity (i.e., the probability that the measure correctly classifies subjects who demonstrate change when an external criterion of clinical change is used) and specificity (i.e., the probability that the measure correctly classifies subjects who do not demonstrate change when the external criterion is used) of each value of change in the measure are calculated and used to plot the curve. Sensitivity values and false-positive rates (1-specificity) are then plotted at the y and x-axis on the curve. Using the Youden index, an optimal cutoff point (i.e., the minimal important change figure) is computed and taken as the MIC, which indicates the change score associated with the least misclassification [34–36].

Anchor-based methods represent powerful methods to calculate the minimal important change as they explicitly define and incorporate the concept of minimal importance; however, they limit by not considering the variability of tool scores within the sample.

5.3 Distribution-Based Approaches

They are based on the distributional features of the population under study and explicit the observed change in the outcome measure to some form of variation to get a standardized metric. Several methods have been proposed over time, such as the effect size and the standard error of measurement; they should be used with caution because they do not directly indicate the importance of the observed change [34–36].

Two issues are crucial before addressing interpretability: (1) scores distribution and (2) floor and ceiling effects. The first delineate if the sample is distributed over the whole range of the scale, whether the study sample has high or low scores, or patients are clustered at points on the scale. The second is again important as patients at both ends of a measurement tool cannot show any further improvements or worsen [1].

6 Additional Properties [41, 42]

6.1 Floor and Ceiling Effects

These terms describe how subjects have scored at or near the possible lower or upper limit, preventing from measuring variance above a certain level. They are recognized when more than 15% of patients achieve the lowest (i.e., floor) or the highest (i.e., ceiling) possible score.

6.2 Precision

It represents the instrument's exactness, which is based on the number and accuracy of distinctions made. This issue is raised concerning response categories and numerical values, and for the relationship between the range of difficulty of the items and the distribution of what is being measured. Advanced statistical methods (e.g., the Rasch analysis) or simpler techniques such as ordering the items based on their mean scores may provide a system for examining an instrument's interval characteristics.

6.3 Acceptability

It describes how easy the measure is for respondents to complete. Patients are investigated about the response rate (i.e., missing values) and the time to complete the tool.

6.4 Feasibility

It constitutes how easily the instrument can be administered and processed, including the extent of effort, burden, and disruption to staff and clinical care arising from the use. It also requires gathering information on professional expertise to apply or interpret the instrument and an instruction manual (including its clarity).

7 Conclusions

Researchers are encouraged to ensure outcome measures are psychometrically sound, and they are administered thoughtfully and analyzed correctly. All measures should meet the classical requirements of reliability, validity, and responsiveness. Conducting an appropriate evaluation is a difficult task: evaluating outcome measures should be carefully oriented, using well-chosen judgment criteria defined in terms of precise objective and referring to international standards. The importance of appropriate assessment is clearly demonstrated by the growing number of studies devoted to identifying the best criteria for choosing and applying. Development and refinements of measurement tools are recommended over time as part of an ongoing process. More robust validation is possible when a measurement instrument that has shown an acceptable degree of validity in one situation is also validated in another context, disease, or population. More evidence is recommended to test and enlarge psychometric properties. Despite encouraging results and ongoing processes as described above, improving the quality of assessment still constitutes an important challenge for most medical fields. Better assessment will undoubtedly lead to better planning of care, better communication among health professionals, better evaluations of treatment efficacy, better clinical research, and better knowledge about patients' needs and expectations.

References

1. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Cambridge: Cambridge University Press; 2011. https://doi.org/10.1017/CBO9780511996214.
2. Clinton-McHarg T, Yoong SL, Tzelepis F, et al. Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the consolidated framework for implementation research: a systematic review. Implement Sci. 2016. https://doi.org/10.1186/s13012-016-0512-5.
3. O'Connor A, McGarr O, Cantillon P, McCurtin A, Clifford A. Clinical performance assessment tools in physiotherapy practice education: a systematic review. Physiother (United Kingdom). 2018. https://doi.org/10.1016/j.physio.2017.01.005.
4. Walters SJ, Stern C, Robertson-Malt S. The measurement of collaboration within healthcare settings: a systematic review of measurement properties of instruments. JBI Database Syst Rev Implement Reports. 2016. https://doi.org/10.11124/JBISRIR-2016-2159.
5. Castelnuovo G, Giusti EM, Manzoni GM, et al. What is the role of the placebo effect for pain relief in neurorehabilitation? Clinical implications from the Italian consensus conference on pain in neurorehabilitation. Front Neurol. 2018. https://doi.org/10.3389/fneur.2018.00310.
6. Marquez MA, De Santis R, Ammendola V, et al. Cross-cultural adaptation and validation of the "spinal cord injury-falls concern scale" in the Italian population. Spinal Cord. 2018;56(7):712–8. https://doi.org/10.1038/s41393-018-0070-6.
7. Berardi A, De Santis R, Tofani M, et al. The Wheelchair Use Confidence Scale: Italian translation, adaptation, and validation of the short form. Disabil Rehabil Assist Technol. 2018;13(4) https://doi.org/10.1080/17483107.2017.1357053.
8. Anna B, Giovanni G, Marco T, et al. The validity of rasterstereography as a technological tool for the objectification of postural assessment in the clinical and educational fields: pilot study. In: Advances in Intelligent Systems and Computing; 2020. https://doi.org/10.1007/978-3-030-23884-1_8.
9. Panuccio F, Berardi A, Marquez MA, et al. Development of the pregnancy and motherhood evaluation questionnaire (PMEQ) for evaluating and measuring the impact of physical disability on pregnancy and the management of motherhood: a pilot study. Disabil Rehabil. August 2020:1–7. https://doi.org/10.1080/09638288.2020.1802520.
10. Amedoro A, Berardi A, Conte A, et al. The effect of aquatic physical therapy on patients with multiple sclerosis: a systematic review and meta-analysis. Mult Scler Relat Disord. 2020. https://doi.org/10.1016/j.msard.2020.102022
11. Dattoli S, Colucci M, Soave MG, et al. Evaluation of pelvis postural systems in spinal cord injury patients: outcome research. J Spinal Cord Med. 2018;43:185–92.
12. Berardi A, Galeoto G, Guarino D, et al. Construct validity, test-retest reliability, and the ability to detect change of the Canadian occupational performance measure in a spinal cord injury population. Spinal Cord Ser Cases. 2019. https://doi.org/10.1038/s41394-019-0196-6.
13. Ponti A, Berardi A, Galeoto G, Marchegiani L, Spandonaro C, Marquez MA. Quality of life, concern of falling and satisfaction of the sit-ski aid in sit-skiers with spinal cord injury: observational study. Spinal Cord Ser Cases. 2020. https://doi.org/10.1038/s41394-020-0257-x.
14. Panuccio F, Galeoto G, Marquez MA, et al. General sleep disturbance scale (GSDS-IT) in people with spinal cord injury: a psychometric study. Spinal Cord. 2020. https://doi.org/10.1038/s41393-020-0500-0.
15. Monti M, Marquez MA, Berardi A, Tofani M, Valente D, Galeoto G. The multiple sclerosis intimacy and sexuality questionnaire (MSISQ-15): validation of the Italian version for individuals with spinal cord injury. Spinal Cord. 2020. https://doi.org/10.1038/s41393-020-0469-8.
16. Galeoto G, Colucci M, Guarino D, et al. Exploring validity, reliability, and factor analysis of the Quebec user evaluation of satisfaction with assistive Technology in an Italian Population: a cross-sectional study. Occup Ther Heal Care. 2018. https://doi.org/10.1080/07380577.2018.1522682.
17. Colucci M, Tofani M, Trioschi D, Guarino D, Berardi A, Galeoto G. Reliability and validity of the Italian version of Quebec user evaluation of satisfaction with assistive technology 2.0 (QUEST-IT 2.0) with users of mobility assistive device. Disabil Rehabil Assist Technol. 2019. https://doi.org/10.1080/17483107.2019.1668975.
18. Berardi A, Galeoto G, Lucibello L, Panuccio F, Valente D, Tofani M. Athletes with disability' satisfaction with sport wheelchairs: an Italian cross sectional study. Disabil Rehabil Assist Technol. 2020. https://doi.org/10.1080/17483107.2020.1800114.
19. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010. https://doi.org/10.1016/j.jclinepi.2010.02.006.
20. McGraw KO, Wong SP. Forming inferences about some Intraclass correlation coefficients. Psychol Methods. 1996. https://doi.org/10.1037/1082-989X.1.1.30.

21. Nunnally JC. Psychometric theory. 1979. https://doi.org/10.1109/PROC.1975.9792
22. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960. https://doi.org/10.1177/001316446002000104.
23. Streiner DL, Norman GR. A Practical guide to their development and use: health measurement scales. 2008. https://doi.org/10.1093/acprof:oso/9780199231881.001.0001.
24. Portney A, Washington RD. Review of: foundations of clinical research applications to practice (3rd edition). J Allied Health. 2010;8:3.
25. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ. 1997. https://doi.org/10.1136/bmj.314.7080.572.
26. Kline P. Handbook of psychological testing. London: Routledge; 2013. https://doi.org/10.4324/9781315812274.
27. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. J Clin Epidemiol. 2018. https://doi.org/10.1016/j.jclinepi.2017.11.005.
28. Child D. The essentials of factor analysis (2nd ed.). Cassell Educational; 1990.
29. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res. 1992. https://doi.org/10.1177/0049124192021002005.
30. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999. https://doi.org/10.1080/10705519909540118.
31. Smith GT. On construct validity: issues of method and measurement. Psychol Assess. 2005. https://doi.org/10.1037/1040-3590.17.4.396.
32. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000. https://doi.org/10.1097/00007632-200012150-00014.
33. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Heal. 2005. https://doi.org/10.1111/j.1524-4733.2005.04054.x.
34. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993. https://doi.org/10.1093/clinchem/39.4.561.
35. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008. https://doi.org/10.1016/j.jclinepi.2007.03.012.
36. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness. A critical review and recommendations. J Clin Epidemiol. 2000. https://doi.org/10.1016/S0895-4356(99)00206-1.
37. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999. https://doi.org/10.1191/096228099673819272.
38. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary Pretense. J Man Manip Ther. 2008. https://doi.org/10.1179/jmt.2008.16.4.82e.
39. Engel L, Beaton DE, Touma Z. Minimal clinically important difference. A review of outcome measure score interpretation. Rheum Dis Clin North Am. 2018. https://doi.org/10.1016/j.rdc.2018.01.011.
40. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009. https://doi.org/10.1179/jmt.2009.17.3.163.
41. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007. https://doi.org/10.1016/j.jclinepi.2006.03.012.
42. Barat M, Franchignoni F. Assessment in physical medicine and rehabilitation. Pavia: Maugeri Foundation Books; 2004.

Methodological Approach to Identifying Outcome Measures in Spinal Cord Injury

Giovanni Galeoto, Marco Tofani, Giulia Grieco, Marina D'Angelo, and Anna Berardi

1 Introduction

Measuring health status and health interventions is fundamental to guarantee the quality of services and good health for all. The biomedical model of health had focused well-being assessment and related interventions on strictly biomedical parameters in the past. The biopsychosocial model's emergence has instead overturned this paradigm, defining health as a set of biological, psychological, and environmental contingencies. The *International Classification of Functioning, Disability, and Health (ICF)*—approved by the World Health Assembly in 2001 [1]—describes a universal framework of functioning, and health examining different components: body functions and structures, activities and participation, and contextual factors. The increasing recognition of the patient perspective and, more specifically, functioning and health has led to an impressive effort in research to develop concepts and instruments to measure them [2]. Comparing selected instruments may provide clinicians and researchers with new insights when selecting health-status measures for clinical studies [3]. High-quality clinical care requires people to provide information regarding how they feel, their symptoms, the possibility of restoring community living, and achieving a satisfying life quality. Keeping in mind the great varieties of areas to consider, how can clinicians choose the more appropriate outcome measures? In the specific context of the SCI population, different efforts are made: the ICF framework was investigated [4, 5], using the Delphi method with occupational therapists [6], physical therapists [7], and customers [8].

Furthermore, many authors tried to investigate the appropriateness of outcome measures for newly acquired SPI population [9] or for measuring function or mobility [10]. Studies suggested a good general methodology, but with a limited sample. Furthermore, for the development of clinical practice guidelines including recommendations for assessments in initial SCI rehabilitation, the psychometric properties of outcome measures and their clinical relevance need to be considered [9, 11]. Unfortunately, in everyday clinical activities, this is a widespread practice. In many countries, the lack of interest and funding opportunities in validating outcome measures may also lead to a worse quality of care and poor adherence/compliance of healthcare plans. The COSMIN (*COnsensus-based Standards for the selection of*

G. Galeoto · A. Berardi (✉)
Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy
e-mail: anna.berardi@uniroma1.it

M. Tofani
Department of Neurorehabilitation and Robotics, Bambino Gesù Paediatric Hospital, Rome, Italy

G. Grieco · M. D'Angelo
R.O.M.A. Rehabilitation Outcome Measures Assessment, Non-Profit Organization, Rome, Italy

G. Galeoto et al. (eds.), *Measuring Spinal Cord Injury*,
https://doi.org/10.1007/978-3-030-68382-5_3