

SPECIAL REPORT

TOWARD STANDARDIZED REPORTING FOR A COHORT STUDY ON FUNCTIONING: THE SWISS SPINAL CORD INJURY COHORT STUDY

Birgit Prodinger, PhD, MSc^{1,2,3}, Carolina S. Ballert, MSc^{1,3}, Mirjam Brach, MBA, MPH^{1,3}, Martin W. G. Brinkhof, PhD^{1,3}, Alarcos Cieza, PhD, MSc, MPH^{1,4,5}, Kerstin Hug, MD⁶, Xavier Jordan, MD⁷, Marcel W. M. Post, PhD^{1,8,9}, Anke Scheel-Sailer, MD^{3,10}, Martin Schubert, MD¹¹, Alan Tennant, PhD^{1,3} and Gerold Stucki, MD, MS^{1,2,3}, for the SwiSCI Study Group

From the ¹Swiss Paraplegic Research (SPF), Nottwil, ²ICF Research Branch, a cooperation partner within the World Health Organization Collaborating Centre for the Family of International Classifications in Germany (at DIMDI), ³Department of Health Sciences and Health Policy, University of Lucerne, Nottwil, Switzerland, ⁴Faculty of Social and Human Sciences, University of Southampton, Southampton, UK, ⁵Public Health and Health Services Research, Department of Medical Informatics, Biometry and Epidemiology (IBE), Ludwig-Maximilians-Universität (LMU) Munich, Munich, Germany, ⁶REHAB Basel, Basel, ⁷Clinique romande de réadaptation, Sion, Switzerland, ⁸Department of Rehabilitation Medicine, Center for Rehabilitation, University of Groningen, University Medical Center Groningen, Groningen, ⁹Brain Center Rudolf Magnus and Center of Excellence in Rehabilitation Medicine, University Medical Center Utrecht and De Hoogstraat, Utrecht, The Netherlands, ¹⁰Swiss Paraplegic Center, Nottwil and ¹¹Paraplegikerzentrum Universitätsklinik Balgrist, Zürich, Switzerland

Objective: Functioning is an important outcome to measure in cohort studies. Clear and operational outcomes are needed to judge the quality of a cohort study. This paper outlines guiding principles for reporting functioning in cohort studies and addresses some outstanding issues.

Design: Principles of how to standardize reporting of data from a cohort study on functioning, by deriving scores that are most useful for further statistical analysis and reporting, are outlined. The Swiss Spinal Cord Injury Cohort Study Community Survey serves as a case in point to provide a practical application of these principles.

Methods and Results: Development of reporting scores must be conceptually coherent and metrically sound. The International Classification of Functioning, Disability and Health (ICF) can serve as the frame of reference for this, with its categories serving as reference units for reporting. To derive a score for further statistical analysis and reporting, items measuring a single latent trait must be invariant across groups. The Rasch measurement model is well suited to test these assumptions.

Conclusion: Our approach is a valuable guide for researchers and clinicians, as it fosters comparability of data, strengthens the comprehensiveness of scope, and provides invariant, interval-scaled data for further statistical analyses of functioning.

Key words: comparability; informed decision-making; epidemiology; health information systems; Rasch analysis.

J Rehabil Med 2016; 48: 189–196

Correspondence address: Birgit Prodinger, ICF Unit, Swiss Paraplegic Research, CH-6207 Nottwil, Switzerland. E-mail: birgit.prodinger@paraplegie.ch

Accepted Aug 20, 2015; Epub ahead of print Feb 1, 2016

INTRODUCTION

In order to construct a comprehensive picture of individual and population health, data on morbidity and mortality must be complemented with information on the disability experienced by a person and how this plays out in daily life. This comprehensive approach to comparing the health of individuals and (sub-)populations assumes that people with the same health condition may experience different levels of functioning in their daily lives, as functioning is an outcome that is influenced not only by the severity of the health condition and the impairments in body functions and structures, but also by environmental and personal factors. The World Health Organization's (WHO) International Classification of Functioning, Disability and Health (ICF) reflects this comprehensive understanding of health, and operationalizes it by means of the notion of functioning (1). Having information on functioning at the individual and population levels is critical for clinicians, health and related researchers and policy-makers, in order to set intervention targets tailored to the health and related needs of individuals and (sub-)populations, and to monitor change over time.

Cohort studies are the most appropriate study design to monitor functioning over time in specified populations. Cohort studies also facilitate the identification of risk groups or intervention targets based on population-based, longitudinal data in order to guide health policy and health services programmes (2, 3). Clarity of outcomes and their operationalization in cohort studies is of paramount importance in order to judge the quality of the study, the conclusions that can be drawn from it, its comparability with other studies and suitability for integration into meta-analyses. Although guidelines have been developed on the reporting of cohort studies (4), for reporting on function-

ing there are some particular issues that should be addressed concerning the specification of variables, data sources and statistical methods. These concerns are addressed in this paper.

Functioning is a complex multidimensional phenomenon, as reflected in the vast array of mutually exclusive and cumulative exhaustive categories contained in the ICF. Many aspects of functioning are not directly observable, but must be operationalized as latent traits with linear transformation along a continuum (5, 6). For example, a person's emotional functions or his or her ability to conduct self-care activities should be measured along a continuum, from having no problems to having complete problems. This is markedly different from mortality or morbidity, both of which can be reported as discrete and single states of affairs or events. Over the past decades, a wide range of instruments and assessment tools have been developed to operationalize many discrete aspects of human functioning. These instruments may assess the same category of functioning from various perspectives, e.g. one instrument measures whether a person is able to move around while another measures whether the person needs assistive devices or other support in order to move around. Both the person's ability and the need for assistive devices or support are important features of the experience of disability that must be measured, but they must also be kept distinct.

For this information to be most useful locally, nationally and internationally, across the course of an individual's life, across health conditions, and along the continuum of care, what is required is a commonly agreed conceptual and methodological framework that can make transparent the domains of functioning that are being measured and how they are being measured. The methodology for specifying which variables should be measured in order to set up a cohort study, relying on the ICF, has been discussed in detail in other publications (7, 8). This methodology foresees the operationalization of each area of functioning in a cohort study in terms of items or (sub-)scales from existing instruments. The question that remains, however, is how to analyse and report this information in a meaningful way so that it remains both coherent with its conceptual foundations and metrically sound so as to ensure that relevant and reliable information is available for use; for example, by a clinician or health system decision-maker. This paper proposes a solution to this issue by providing guiding principles on how to standardize reporting of functioning data acquired from a cohort study, by deriving scores that are most useful for further statistical analysis and reporting. The data collected from the Basic Module of the Swiss Spinal Cord Injury Cohort Study (SwiSCI) (9) are used as an example here in order to provide concreteness to these principles.

METHODS

A cohort study is designed to follow a specified group of people so as to ascertain incidence of some condition, or some specified outcome for those with an established condition. When the topic of interest is functioning, the initial stage involves specifying which variables or aspects of functioning should be assessed, and how they will be assessed. Here the ICF serves as the conceptual framework for determining *what*

to assess. ICF Core Sets constitute a selection of the most relevant ICF categories with respect to a specified health condition or given context (10). As such, they can guide the specification of the most relevant areas of functioning that should be addressed in a cohort study about functioning. Once a set of ICF categories has been specified, existing instruments can be linked to the ICF to identify *how to assess* the specified ICF categories (11). After the relevant information has been collected, the question remains as to how the data can be analysed in order to derive scores that are most useful for further statistical analysis and reporting. The development of such scores must be both conceptually coherent and metrically sound.

Conceptual coherency

The ICF categories within each component of Body Functions and Structures, Activities and Participation, and Environmental Factors are structured hierarchically into chapters, and each chapter contains 2nd, 3rd and, for Body Function, 4th level categories. In integrating information about people's functioning, the guiding principle is to adhere to the components of the ICF and the related classification hierarchy. Keeping the components distinct when integrating information makes it possible in further analysis to examine the associations and interactions between components. Within each component, integration of information can be conducted at any level of the classification depending on the purpose for creating a metric and the availability of information linked to the ICF category. An ICF-based scale can be created for each ICF category, as illustrated by Cieza et al. (12). Doing so requires there to be sufficient items identified for the assessment of the category, and each category must be operationalized with as few items as possible. Otherwise, it will not be possible to create a functioning profile across ICF categories with a single questionnaire in a reasonable time. However, creating a metrically sound score at the level of each ICF category is not always a viable solution. When it is not, one can move the hierarchy upwards in the ICF; for example, to the chapter level or across chapters within each component, and then integrate the information at that higher level of the classification.

Metric properties

To derive a score for further statistical analysis that is suitable for comparisons with other studies, it is important to ensure that each item measures a single latent trait (unidimensionality) and does so in the same way across different groups (invariance). Unidimensionality entails that only the individual's ability in a given latent trait determines his or her score. Having invariant measures ensures that group differences that can be found in the data are true differences between these groups and not measurement artefacts. The Rasch measurement model is most suitable for this purpose as it builds upon the assumptions of invariance (13). Applying the Rasch measurement model involves testing whether the structure of responses fulfils the model's assumptions. Both the parameter of a person's ability and item's difficulty are placed along a single dimension, but calibrated separately (14, 15). Hence, person abilities are independent of item distribution, and likewise, item difficulties are independent of person distribution.

A variety of fit statistics can then be applied to determine whether the data fit the model. Differential item functioning (DIF) is a test to examine invariance of the measure across relevant groups (16). The Person Separation Index (PSI) serves as an estimate of the internal consistency reliability of the scale. Good targeting of the scale is achieved when the distribution of the items along the metric matches well with the distribution of the abilities of the persons on the same scale. Where fit of the data to the Rasch model is evident and the assumption of local independence and unidimensionality are met, interval-scale measures are attained that can be used for any further statistical analysis to describe and understand functioning of a (sub-) population. A logit scale derives from the Rasch analysis that can be transformed into a more user-friendly scale from 0 to 100. This scale can then be associated with the 0 to 100 percentage values that WHO has provided as a reference for the ICF qualifiers (1). The ICF

facilitates a description of the extent of a problem that an individual or (sub-)population experiences with regards to identified ICF domains along a continuum from no problem to extreme problem. As such, it supports the development of a functioning profile, which illustrates the levels of functioning across ICF domains (17).

When integrating information collected with items derived from existing instruments within or across ICF categories, the following points should be taken into consideration. First, the direction of responses from items taken from existing instruments should be reviewed and, if necessary, aligned. For instance, in some instruments the response option “0” may indicate “no problem” in one instrument and “complete problem” in another instrument. To ensure a meaningful co-calibration of information, items are re-directed based on the ICF qualifier, where “0” means “no problem”. Secondly, even if items from various instruments are linked to the same ICF category, the instrument may still adopt a different perspective from that in terms of which latent trait has been operationalized. This is illustrated by the example on performance of getting dressed or support needed to get dressed, mentioned above. While “dressing” is addressed in both instances, one should consider whether it is meaningful to co-calibrate information from different perspectives.

Once these points have taken into consideration, clinicians and researchers can continue using their existing instruments in routine practice and research, while findings about items linked to ICF categories from the relevant ICF Core Sets can then be reported based on the ICF. This approach does not ignore assessment practices that have been established over the years, but rather builds upon these to strengthen the value of existing information and information collected with them.

APPLICATION OF PRINCIPLES AND METHODS TO THE SWISS SPINAL CORD INJURY COHORT STUDY

Description of the Swiss Spinal Cord Injury Cohort Study

The Swiss Spinal Cord Injury Cohort Study (SwiSCI) is a national cohort study conducted in collaboration between the Swiss Paraplegic Association and all 4 SCI rehabilitation facilities in Switzerland, and hosted by Swiss Paraplegic Research. This national cohort study follows a modular structure and contains different pathways to monitor the functioning of people with SCI across the lifespan and the continuum of care (18). As mentioned, the ICF, more specifically ICF Core Sets relevant for SCI, guided the specification of which variables to include in SwiSCI (7, 8). Subsequently, items from existing instruments were identified to operationalize the specified ICF categories. Tables I and SI exemplify this for the identified ICF categories in the *Chapters d4 Mobility to d9 Community, social and civic life*, and *Chapters b1 Mental functions to b8 Functions of the skin and related structures* respectively. The comprehensive list of ICF categories deemed relevant for a cohort study on functioning of people with SCI are outlined in an adjunct paper (7).

The recruitment process for SwiSCI is outlined elsewhere (9). Participants were given the option of completing the questionnaire on paper, online or by telephone interview. The SwiSCI community survey was formally approved by the respective regional Research Ethics Committees. All participants gave written consent for the anonymous use of their questionnaire data.

Sample characteristics

For this paper we used only data from the Basic Module of the SwiSCI community survey (18) in which 1,549 individuals participated. Of the participants 71.5% were male, median age

52 years (interquartile range (IQR) 42–63 years), and median time since injury 13.5 years (IQR 6.2–25.10 years). More than three-quarters (77.6 %) had experienced a traumatic injury, 68.6% had paraplegia and 41.7% had a complete lesion.

Data analysis

In the first step, the available items were examined and a strategy was developed for meaningful integration of the information across ICF categories, based on the principles outlined in this paper. For illustrative purposes, the analysis and results focus on the ICF component Activities and Participation, more specifically the scale for *d6–d9 Involvement in life situations*; the information for the other scales. Most of the ICF categories picked from the Activities and Participation component had only 1 or 2 items assigned to them. Hence, co-calibrating items within *Chapter 4 Mobility* and *Chapter 5 Self-care* was reasonable. Items linked to ICF categories across *Chapters d6 Domestic life to d9 Community, social and civic life* were co-calibrated for several reasons. In addition to the practical considerations of the number of items available for each ICF category and chapter, co-calibrating items across chapters had been found appropriate in a previous psychometric study that had investigated the dimensional structure of the ICF and had concluded that if Activities and Participation is to be divided into different dimensions, integrating d6–d9 into a single dimension entitled *Involvement in life situations* would be a good solution (19).

Based on this strategy, Rasch analysis was performed on the proposed item sets to examine their metric properties. On the assumption that the distances between thresholds vary across response categories, the Partial Credit Model (PCM) was applied (20). If disordered thresholds occurred, collapsing response categories was used to improve fit to the model. Item fit was explored by examining outfit and infit mean square residuals to the Guttman structure. Goodness of fit was confirmed when the mean squared outfit and infit were within a determined range. Sample size was taken into consideration when determining this range to increase the power of analysis (21). Regarding local dependency, correlations of standardized residuals >0.2 were considered as indicators for response dependency.

Locally dependent items were accommodated by integrating them into testlets (22, 23). DIF was tested using analysis of variance (ANOVA) of the Rasch residuals. The level of significance was Bonferroni adjusted to account for multiple testing. In the present analysis, DIF was tested for administration mode (online or paper-pencil questionnaire), language (German or French), gender (male or female), education level in years (\leq and >12 years of education), and SCI aetiology (traumatic or non-traumatic SCI). Furthermore, DIF was examined for age at injury (\leq or >45 years) and time since injury (\leq or >15 years); in creating these groups the year threshold proposed for reporting by the International Spinal Cord Society (ISCoS; 24) closest to the median for the respective group variable was chosen. DIF can be resolved by splitting the respective item by DIF groups, which then provides different estimates of the difficulty based on the belonging of a person to a certain group (25). Data with missing values were used for the analysis. All analyses were

Table I. Overview of the International Classification of Functioning, Disability and Health (ICF) categories identified as relevant for reporting for the Chapters d4 Mobility to d9 Community, social and civic life, their operationalization, and remedies applied to improve fit in the Rasch measurement model

ICF Category	Label	Item identified for operationalization of ICF category	Instrument	Response format	Reversal of response options	Testlet	Collapsing ^a	Item and testlet split ^b	Item deleted in analysis			
d410 Changing basic body position	BodyPosition_1	Movement without assistance/electric aids in bed	SCIM-SR	Polytomous	×		×	Gender				
	BodyPosition_2	Transfers: wheelchair- car	SCIM-SR	Polytomous	×				×			
	BodyPosition_3	Transfers: floor-wheelchair	SCIM-SR	Polytomous	×	Transfer	×	Age at SCI; Time since injury				
d420 Transferring oneself	Transfer_1	Transfers: bed-wheelchair	SCIM-SR	Polytomous	×							
	Transfer_2	Transfers: wheelchair-toilet/tub	SCIM-SR	Polytomous	×							
d455 Moving around	Moving_1	Ascending or descending stairs	SCIM-SR	Polytomous	×	Moving	×	Age at SCI; Time since injury				
d460 Moving around in different locations	Moving_2	Restriction in outdoor mobility	USER-P	Polytomous	×		×					
d465 Moving around using equipment	Moving_3	Moving around indoor	SCIM-SR	Polytomous	×	Moving	×	Age at SCI; Time since injury				
	Moving_4	Moving around moderate distances	SCIM-SR	Polytomous	×							
d510 Washing oneself	Washing_1	Washing upper body and head	SCIM-SR	Polytomous	×							
	Washing_2	Washing lower body	SCIM-SR	Polytomous	×							
d520 Caring for body parts	CaringBodyParts	Grooming	SCIM-SR	Polytomous	×		×					
d530 Toileting	Toileting	Toileting	SCIM-SR	Polytomous	×		×	Time since injury				
d540 Dressing	Dressing	Dressing lower body	SCIM-SR	Polytomous	×		×	Age at SCI				
d550 Eating	Eating	Eating and drinking	SCIM-SR	Polytomous	×		×					
d560 Drinking												
d640 Doing housework	Housework	Restriction in household duties	USER-P	Polytomous	×			Age at SCI				
d710 Basic interpersonal interactions	Interpersonal_1	Restriction in relationship	USER-P	Polytomous	×				×			
	Interpersonal_2	Restriction in contacting others	USER-P	Polytomous	×		×					
d850 Remunerative employment	Employment	Restriction in work or education	USER-P	Polytomous	×		×					
d920 Recreation and leisure	Leisure_1	Restriction in sports	USER-P	Polytomous	×	Outside	×					
	Leisure_2	Restriction in going out	USER-P	Polytomous	×							
	Leisure_3	Restriction in day trips	USER-P	Polytomous	×							
	Leisure_4	Restriction in leisure activities at home	USER-P	Polytomous	×							Gender

^aDetails about the collapsing strategies can be requested from the authors. ^bItem and testlet split, gender (M: male; F: female), age at SCI (A45: >45 years; B45: ≤45 years), time since injury (A15: >15 years; B15: ≤15 years).

SCIM-SR: Spinal Cord Independence Measure—Self Report; USER-P: Utrecht Scale for Evaluation of Rehabilitation-Participation; SCI: spinal cord injury.

performed with the software R (26). For the Rasch analyses the package TAM (27) was used which centres the person abilities around 0 and uses the Marginal Maximum Likelihood Estimation (MML) for item parameter estimation.

Deriving scores from ICF-based scales for further statistical analyses

Five ICF-based interval-scales, 2 for body functions (*b1 Mental functions* and *b2-b8 Functions of body systems*) and 3 for

activities and participation (*d4 Mobility, d5 Self-care, d6–d9 Involvement in life situations*), were created. The ICF categories for Chapters *d4 Mobility* to *d9 Community, social and civic life*, and the items identified for their operationalization are outlined in the first 4 columns of Table I. For instance, the USER-P is a self-report instrument recently developed for assessing the participation of individuals in rehabilitation practice (28) and is described in more detail in an adjunct paper (29). All items linked to *d6–d9 Involvement in life situations* were taken from

the USER-P. Looking more specifically into the findings of the Rasch analysis of *Chapters d6 Domestic life to d9 Community, social and civic life*, the lower part of Tables I and SI (line *d640 Doing housework to d920 Recreation and leisure*) indicates that collapsing of response options was necessary for 2 items, namely “Restriction in contacting others” and “Restriction in work or education” linked to *d710 Basic interpersonal interactions* and *d850 Remunerative employment*. Local dependency was found for one item pair linked to the ICF category *d920 Recreation and leisure* on “Restriction in going out” and “Restriction in day trips”. The item “Restriction in household duties” linked to *d640 Doing housework* showed DIF for age, and the item “Restriction in leisure activities at home” linked to *d920 Recreation and leisure* DIF for gender. Both items were split into sub-group specific items (Tables I and SI, columns 5–10). One item was removed from the present analysis “Restriction in relationships” linked to *d710 Basic interpersonal relationships* as it revealed a high level of structural missing values. This scale is referred to as *d6-d9 Involvement in life situations*.

As shown in Tables II and SII, all items fitted the Rasch model based on the outfit and infit mean square residuals. A PSI of 0.74 is acceptable and can be interpreted in the sense that the scale can differentiate in a reliable way 2 distinct ability groups on the metric. The person-item map for this scale, shown in Fig 1, reflects a floor effect of the scale. Fig 1 illustrate the person and item distribution across the metric for *d6-d9 Involvement in life situations*. The person-item maps for the other scales are presented in Fig S1. The scores for the functioning scales derived from this study thus become available for use to create a functioning profile on the population-level (Fig. 2) or for any further statistical analyses.

DISCUSSION

This paper outlines an approach for analysing data from a cohort study on functioning, in order to derive scores useful for further statistical analyses and reporting. These principles

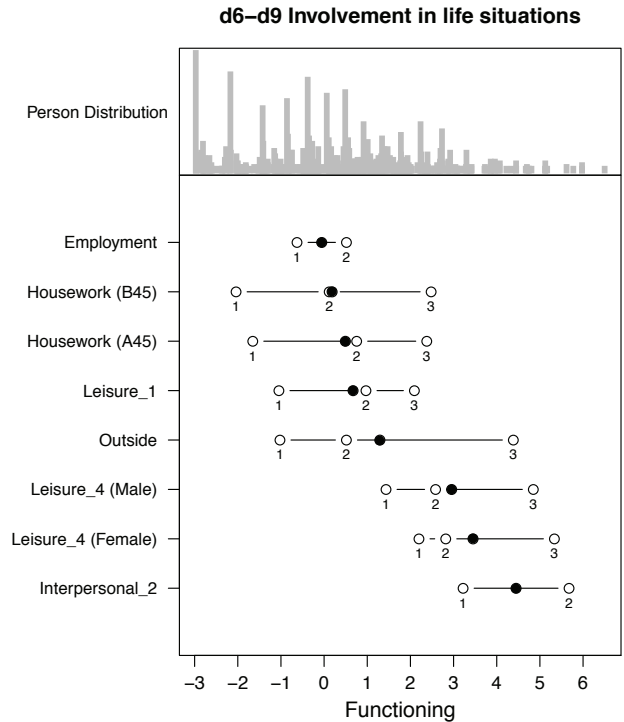


Fig. 1. Example of person-item map for *d6-d9 Involvement in life situations*: the upper part of the figure illustrates the person distribution across the metric. The middle part outlines the item locations (black dot), as well as the threshold distribution (white circles) for each item along the metric. The numbers attached to the circles are the thresholds of an item; that is, the point between 2 adjacent response options where the probability that a person obtains either of the successive scores is equal. The metric ranges from -3 to 6 logits, as indicated at the bottom of the figure. The item “housework” was split by age (≤ 45 years “B45”, and > 45 years “A45”); and item “Leisure_4” by gender.

and methods complement previous research that outlined the methods for setting up a cohort study on functioning and specifying how to operationalize relevant ICF categories (7,

Table II. Summary fit statistics for International Classification of Functioning, Disability and Health (ICF)-based scale for *d6-d9 Involvement in life situations* to the Rasch measurement model

	Label	Subgroup	Location	Outfit MSQ	Infit MSQ
d6-d9 Involvement in Life Situations	Employment		-0.06	0.93	0.98
	Housework	≤ 45	0.19	1.08	1.09
	Housework	> 45	0.49	1.05	1.05
	Leisure_1		0.67	1.03	1.04
	Outside		1.29	0.87	0.90
	Leisure_4	Male	2.96	0.86	1.03
	Leisure_4	Female	3.45	0.97	1.01
	Interpersonal_2		4.45	1.10	1.01
	Smith criteria			0.85–1.15	0.95–1.05
		Mean		Confidence interval	Floor
	Person Abilities	0.18	(0.09; 0.27)	25.00%	0.00%
	Item Difficulties	1.68	(0.50; 2.86)		
	Person Separation Index	0.74			

Note: outfit is more outlier-sensitive, whereas infit is more sensitive to patterns of the responses. MSQ: mean-square residual summary statistic.

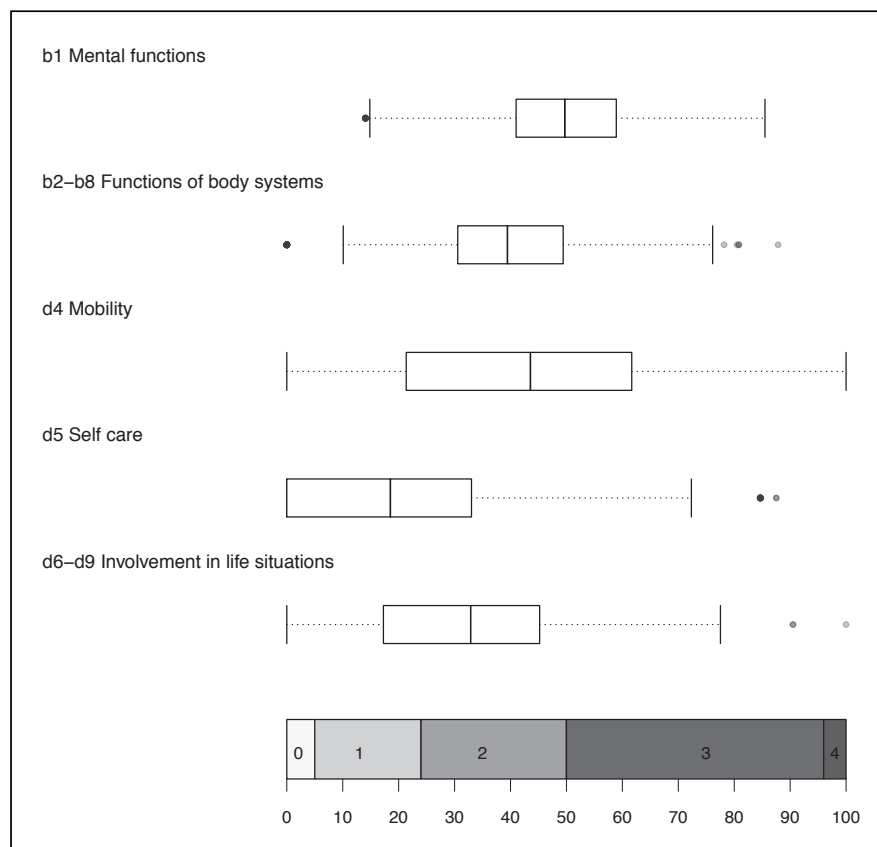


Fig. 2. Demonstration of functioning profile for Swiss Spinal Cord Injury Cohort Study (SwiSCI) population. The scores derived from the International Classification of Functioning, Disability and Health (ICF)-based scales, which have been created by integrating information across ICF categories in a conceptually coherent and psychometrically sound way, can subsequently be deployed for graphical presentation of the persons' abilities, e.g. in a functioning profile, as in this figure. *Boxplots*: Boxes represent upper to lower quartiles with median; whiskers show lowest and highest 25% of scores respectively; dots present outliers. Range from 0 to 100 (*bottom of figure*) can be reported in terms of ICF qualifiers (0=no problem; 1=mild problem; 2=moderate problem; 3=severe problem; 4=complete problem). Note: The scales have been made comparable based on a common distribution; however, the score from one scale cannot be transformed to another as they measure different latent traits (30).

8). Previous research provides evidence that information can be linked to a single ICF category and integrated into a unidimensional scale (12). The current study demonstrates that it is feasible to apply these methods across the whole spectrum of functioning. A functioning profile is a promising tool for monitoring functioning of (sub-)populations over time and identifying areas where further research or interventions are needed. Measuring ICF components is important in order to facilitate subsequent investigations into the association between components, while controlling for relevant mediators or moderators. The principles outlined here respond to the need for transparent and clear reporting of cohort studies to ensure that other researchers can make sense of the variables and outcomes included in a study so that they can decide whether these findings are comparable with other studies or can be included in systematic reviews (4). As cohort studies on functioning are fairly new, having principles for reporting in place at the beginning is of great benefit for future research.

Stakeholders at any operational level in health and related systems need relevant information in order to make decisions about allocation of resources. Given the complexity of functioning, there is a need to integrate information within and across ICF categories so that any decision-maker can retrieve a concise, intuitive description of people's functioning at a glance. Furthermore, being able to link functioning information to the ICF regardless of its mode of collection offers a more comprehensive foundation for decision making. To be able

to respond most effectively to the needs of people with SCI, invariant measures are valuable as they facilitate further stratified analyses and comparisons across groups. Such measures will assist in identifying risk groups or intervention targets most relevant for people with a given health condition to guide health policy and health services initiatives to target efforts to (sub-)populations most prone to experiencing problems in certain aspects of functioning.

Moreover, as functioning occurs in interaction with the immediate and larger environment (31), cross-national comparisons of functioning profiles on individual and population level is important in order to better understand the levels of functioning of the SwiSCI study to people with SCI in other countries. Establishing comparability with other national surveys or specific registries will strengthen the basis for decision making about rehabilitation programmes in response to the needs of individuals and populations living with a health condition, nationally and internationally. Having comparable information available to monitor and most effectively respond to functioning of people with SCI has been called for in the WHO's report *International Perspectives on Spinal Cord Injury* (IPSCI; 32). It is, in short, of considerable value to ensure that the principles and methods described here are in place, so as to facilitate the linking of existing information to the ICF.

There are some further considerations worth mentioning here. When developing ICF-based scales, several conceptual and methodological decisions must be made. The Rasch meas-

urement model offers different strategies for accommodating misfitting items, including removing an item from an existing scale (the least preferred solution applied only when all other strategies failed). Deleting an item does not mean it is irrelevant. In the context of integrating information about items from instruments into an ICF category or across ICF categories, it is worth keeping in mind that ICF categories and the items of an instrument are categorically different. ICF categories are meaningful reference units for reporting aspects of functioning regardless of which instruments is used. Items from an instrument represent indicators of a latent trait to describe variations within it (33). Hence, when reporting functioning in a standardized manner, ICF categories relevant for reporting, for example, because they are part of an ICF Core Set, cannot simply be deleted, even if the item used for its operationalization does not fit the Rasch measurement model. In this case, the item of this ICF category would still need to be reported as a frequency, although not summated into a score. Moreover, the perspective represented in an item used to operationalize an ICF category should be maintained in the creation of an ICF-based interval-scale. For instance, all items linked to ICF categories for the ICF chapters *d6 to d9 Involvement in life situations* adopt a descriptive perspective and focus on the experienced restrictions of a person in performing certain activities. Therefore, the ICF-based interval-scale should also reflect this perspective. Another challenge is the targeting of the scale. For instance, the scale for *d6-d9 Involvement in life situations* reveals a floor effect, which indicates that the scale cannot discriminate well between individuals' level of functioning when they score low ("no problem"). To deal with this it would be necessary to expand the scale with additional items that could discriminate better between people at this level of functioning.

In conclusion, functioning, in addition to morbidity and mortality, is an important outcome in cohort studies. Transparency about how outcomes have been selected, operationalized and analysed in a cohort study is of paramount importance when evaluating the quality of a study and its relevance for systematic literature reviews and meta-analyses. In this paper we have proposed principles and methods for analysing and reporting the data of a cohort study on functioning that was based upon the ICF. This approach is a valuable guide for researchers and clinicians, as it fosters the comparability of data, strengthens the comprehensiveness of scope, and provides invariant, interval-scaled data for further statistical analyses of functioning.

ACKNOWLEDGEMENTS

This study was financed within the framework of the Swiss Spinal Cord Injury Cohort Study (SwiSCI, www.swisci.ch), supported by the Swiss Paraplegic Research.

The members of the SwiSCI Steering Committee are: Xavier Jordan, Bertrand Léger (Clinique Romande de Réadaptation, Sion); Michael Baumberger, Hans Peter Gmünder (Swiss Paraplegic Center, Nottwil); Armin Curt, Martin Schubert (University Clinic Balgrist, Zürich); Margret Hund-Georgiadis, Kerstin Hug (REHAB Basel, Basel); Hans Georg Koch,

(Swiss Paraplegic Association, Nottwil); Hardy Landolt (Representative of persons with SCI, Glarus); Hannjörg Koch (SUVA, Luzern); Mirjam Brach, Gerold Stucki (Swiss Paraplegic Research, Nottwil); Martin Brinkhof, Christine Thyrian (SwiSCI Study Center at Swiss Paraplegic Research, Nottwil).

REFERENCES

1. WHO. International Classification of Functioning, Disability, and Health. Geneva: World Health Organization; 2001.
2. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet* 2002; 359: 341–345.
3. Bonita R, Beaglehole R, Kjellström T, editors. Basic epidemiology. 2nd edition. Geneva: World Health Organization; 2006.
4. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bull World Health Organ* 2007; 85: 867–872.
5. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000; 38: I128–I142.
6. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? *J Rehabil Med* 2012; 44: 97–98.
7. Proding B, Ballert CS, Cieza A. Setting up a cohort study of functioning: from classification to measurement. *J Rehabil Med* 2016; 48: 131–140.
8. Cieza A, Boldt C, Ballert CS, Eriks-Hoogland I, Bickenbach JE, Stucki G. Setting up a cohort study on functioning: deciding what to measure. *Am J Phys Med Rehabil* 2011; 90: S17–S28.
9. Brinkhof, M.W.G., Fekete, C., Chamberlain, J.D., Post, M.W.M., Gemperli, A.: Swiss national community survey on functioning after spinal cord injury: Protocol, characteristics of participants and determinants of non-response. *J Rehabil Med* 2016; 48: 120–130.
10. Selb M, Escorpizio R, Kostanjsek N, Stucki G, Üstün B, Cieza A. A guide on how to develop an International Classification of Functioning, Disability and Health Core Set. *Eur J Phys Rehabil Med* 2014; 51: 105–117.
11. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Üstün B, Stucki G. ICF linking rules: an update based on lessons learned. *J Rehabil Med* 2005; 37: 212–218.
12. Cieza A, Hilfiker R, Boonen A, Chatterji S, Kostanjsek N, Ustun BT, et al. Items from patient-oriented instruments can be integrated into interval scales to operationalize categories of the International Classification of Functioning, Disability and Health. *J Clin Epidemiol* 2009; 62: 912–921.
13. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish National Institute for Educational Research; 1960.
14. Christensen KB, Kreiner S, Meshab M, editors. Rasch models in health. London: ISTE Ltd and John Wiley & Sons Inc.; 2013.
15. Bond TG, Fox CM, editor. Applying the Rasch model: fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.; 2007.
16. Teresi J, Fleishman J. Differential item functioning and health assessment. *Qual Life Res* 2007; 16: 33–42.
17. Bickenbach J, Cieza A, Rauch A, Stucki G, editors. ICF Core Sets. Manual for clinical practice. Göttingen: Hogrefe Publishing; 2012.
18. Post MW, Brinkhof MW, von Elm E, Boldt C, Brach M, Fekete C, et al. Design of the Swiss Spinal Cord Injury Cohort Study. *Am J Phys Med Rehabil* 2011; 90: S5–S16.
19. Ballert CS, Stucki G, Biering-Sørensen F, Cieza A. Determining the most robust dimensional structure of ICF categories across subgroups of persons with Spinal Cord Injury to build the basis for future clinical measures. *Arch Phys Med Rehabil* 2014; 95: 2111–2119.e12.
20. Masters GN. A Rasch model for partial credit scoring. *Psycho-*

- metrika 1982; 47: 149–174.
21. Smith RM. Using item mean squares to evaluate fit to the Rasch model. *J Outcome Meas* 1998; 2: 66–78.
 22. Lundgren-Nilsson Å, Tennant A. Past and present issues in Rasch analysis: the Functional Independence Measure (FIM™) revisited. *J Rehabil Med* 2011; 43: 884–891.
 23. Wainer H, Kiely GL. Item clusters and computer adaptive testing: a case for testlets. *J Educ Meas* 1987; 24: 185–210.
 24. DeVivo M, Biering-Sørensen F, New P, Chen Y. Standardization of data analysis and reporting of results from the International Spinal Cord Injury Core Data Set. *Spinal Cord* 2010; 49: 596–599.
 25. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004; 42: 137–148.
 26. Team RDC. R: a Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2013. [Accessed 2014 Oct]. Available from: <http://www.R-project.org>.
 27. Kiefer T, Robitzsch A, Wu M. TAM: Test Analysis Modules. R package version 1.0-3.18-1. 2014 [accessed 2014 Oct]. Available from: <http://CRAN.R-project.org/package=TAM>.
 28. Post MWM, van der Zee CH, Hennink J, Schafrat CG, Visser-Meily JM, van Berlekom SB. Validity of the Utrecht Scale for Evaluation of Rehabilitation-Participation. *Dis Rehabil* 2012; 34 : 478–485.
 29. Mader, L., Post, M.W.M., Ballert, C.S., Michel, G., Stucki, G., Brinkhof, M.W.G.: Metric properties of the Utrecht Scale for Evaluation of Rehabilitation-Participation (USER-Participation) in persons with spinal cord injury living in Switzerland *Rehabil Med* 2016; 48: 165–174.
 30. Kolen MJ, Brennan RL. Test equating, scaling, and linking (2nd edition). New York: Springer; 2004.
 31. Reinhardt JD, Miller J, Stucki G, Sykes C, Gray DB. Measuring impact of environmental factors on human functioning and disability: a review of various scientific approaches. *Disabil Rehabil* 2011; 33: 2151–2165.
 32. WHO, ISCoS. International perspectives on spinal cord injury. Malta: World Health Organization; 2013.
 33. Stucki G, Kostanjsek N, Ustün B, Cieza A. ICF-based classification and measurement of functioning. *Europ J Phys Rehabil Med* 2008; 44: 315.