Authors:

Christine Fekete, PhD
Christine Boldt, MSc
Marcel Post, PhD
Inge Eriks-Hoogland, MD
Alarcos Cieza, PhD, MPH
Gerold Stucki, MD, MS

Affiliations:

From the Swiss Paraplegic Research (SPF), Nottwil, Switzerland (CF, CB, MP, IE-H, AC, GS); Centre of Excellence in Rehabilitation Medicine, Rehabilitation Center De Hoogstraat and Rudolf Magnus Institute for Neuroscience, University Medical Center, Utrecht, the Netherlands (MP); Institute for Health and Rehabilitation Sciences, Ludwig-Maximilian University, Munich, Germany (AC); ICF Research Branch in cooperation with the WHO Collaborating Centre for the Family of International Classifications in Germany (at DIMDI), Nottwil, Switzerland (AC, GS); and Department of Health Sciences and Health Policy, University of Lucerne, Lucerne and SPF, Nottwil, Switzerland (GS).

Correspondence:

All correspondence and requests for reprints should be addressed to: Christine Fekete, PhD, Swiss Paraplegic Research, Guido A. Zäch Strasse 4, 6207 Nottwil, Switzerland.

Disclosures:

The responsibility for the content of this publication lies with the ICF Research Branch.
Financial disclosure statements have been obtained, and no conflicts of interest have been reported by the authors or by any individuals in control of the content of this article.

0894-9115/11/9011(Suppl)-0S29/0
American Journal of Physical Medicine & Rehabilitation
Copyright © 2011 by Lippincott Williams & Wilkins

DOI: 10.1097/PHM.0b013e318230fe41

INVITED REVIEW

How to Measure What Matters

Development and Application of Guiding Principles to Select Measurement Instruments in an Epidemiologic Study on Functioning

ABSTRACT

Fekete C, Boldt C, Post M, Eriks-Hoogland I, Cieza A, Stucki G: How to measure what matters: development and application of guiding principles to select measurement instruments in an epidemiologic study on functioning. Am J Phys Med Rehabil 2011;90(suppl):S29–S38.

The purpose of this article was to describe and to apply a comprehensive set of guiding principles in the selection of measurement instruments for a longitudinal epidemiologic study focusing on functioning using the International Classification of Functioning, Disability, and Health (ICF) as reference framework. Based on the literature, the ICF linkage rules, and the definition of ICF categories to be measured, the following guiding principles for selecting measurement instruments are defined: redundancy, efficiency, level of detail of information, comparability, feasibility, and truth and discrimination. Examples illustrate that the application of guiding principles allows for a systematic and reasoned process of measurement instrument selection and thus offers a potential solution for the multifaceted challenges that one encounters in the selection of measurement instruments. It is transparently demonstrated how the ICF linkage rules enable researchers to address issues such as efficiency, comparability, and redundancy and how the definition of a set of ICF categories to be measured allows assessing inefficiencies in measurement instruments. Because of the ICF linkage rules and the definition of ICF categories to be measured, new guiding principles for selecting measurement instruments emerge. The main challenges lie in the difficulty to quantify and prioritize the applicability of guiding principles and the fact that they strongly interact. Notably, the prioritization and application of guiding principles must be defined considering the specific characteristics and aims of the specific study.

Key Words: Methodology, Epidemiology, Measurement Instruments, ICF, Spinal Cord Injury

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

In the planning of an epidemiologic study of health and, more broadly, the lived experience of people with a health condition such as spinal cord injury (SCI), the use of a suitable framework, and the selection of appropriate measurement instruments are critical.

The International Classification of Functioning, Disability, and Health (ICF)[1] provides this suitable framework.[2] It allows researchers to determine "what to measure" in a top-down, theory-informed manner as described in an adjunct article of this special issue.[2] This approach ensures that all potentially relevant aspects of functioning are systematically considered for inclusion as study variables. This is all the more important than considering the variation in aspects covered by main current longitudinal studies.[3]

More concretely, based on the ICF, it is possible to define so called domain sets of ICF categories.[2] Because ICF categories are often not detailed enough, they often do not provide the necessary specificity to guide the selection of measurement instruments. Therefore, it is crucial to specify the ICF categories of a domain set as relevant for the study population. The domain set, together with the category specifications, determines "what to measure."[2]

Once one knows "what to measure," the challenge is to identify suitable measurement instruments. The selection of suitable measurement instruments is a complex process that involves a range of considerations. They include decisions regarding study design, such as response mode (e.g., self-report *vs.* expert assessment), timing of the measurement (e.g., acute phase *vs.* community setting), and the usefulness of information in relation to study purposes (e.g., description, discrimination, and correlation).

To guarantee a scientifically sound and a transparent selection of measurement instruments, guiding principles must be defined and applied. The development of guiding principles offers the opportunity to systematically transfer the definition of "what to measure" into the selection of measurement instruments and to facilitate an informed decision process. The definition of guiding principles mainly relies on three pillars:

1. We can rely on published "guiding principles" for selecting measurement instruments. For example, Outcome Measures in Rheumatology (OMERACT), an international informal network, has suggested a practical approach including three criteria: truth (issues of validity; is the measure truthful; does it measure what is intended?), discrimination (issues of reliability and sensitivity to change; does the measure discriminate between situations of interest?), and feasibility (can the measure be applied easily given constraints of time, money, and interpretability?).[4] This approach has also been successfully used not only in rheumatology but also in rehabilitation medicine.[5] Although guiding principles from literature are useful, they do not address all challenges one faces when selecting measurement instruments for an epidemiologic study using the ICF as a framework. Most importantly, they do not address the issue of redundancy of different instruments covering not only different but also overlapping ICF concepts.
2. The use of the ICF as a reference framework and the availability of linking procedures[6–8] offer the opportunity to define additional guiding principles. Qualitative and quantitative approaches allow mapping the world of measurement instruments to the ICF. The qualitative mapping of measurement instruments to the ICF relies on linkage rules,[6,7] the quantitative approach relies on transformations using the Rasch model.[8] Thanks to the mapping to the ICF,[6–8] redundancy issues can now be made transparent and can be considered in the selection of measurement instruments. Furthermore, there is now much experience on using the ICF as a reference framework for the comparison and selection of measurement instruments.[9]
3. With the approach to first define "what to measure" through the definition of a domain set and category specifications, one also needs to further consider the guiding principle of efficiency when using a certain measurement instrument. It would not be efficient to use a measurement instrument that covers some domains or category specifications but that, in addition, covers a wide range of concepts that are not defined in the domain set.

The objective of our article was to describe and apply a comprehensive set of guiding principles for the selection of measurement instruments for a longitudinal epidemiologic study focusing on functioning, using the ICF as reference framework. To illustrate the application of these principles, we use examples from an epidemiologic study on functioning of persons with SCI, namely the Swiss Spinal Cord Injury (SwiSCI) Study.[10]

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

METHODS

Definition of Guiding Principles to Select Measurement Instruments

The order of presentation of these guiding principles does not represent a prioritization because the relative importance of these principles largely depends on aspects specific to the study under consideration. The starting point for the SwiSCI Study was the domain set, which is the pile of ICF categories to measure in the study as described by Cieza et al.[2] elsewhere in this issue.

Redundancy

Redundancy can be defined as the overlap between measurement instruments with respect to underlying ICF categories. This principle implies that the items of selected measurement instruments should overlap as little as possible regarding the ICF categories that they address. The mapping methodology[6,7] can be used to study both the degree to which all ICF categories of the study are measured by the candidate measurement instruments, and the overlap between measurement instruments. By mapping all items of candidate measurement instruments to the ICF, one can identify how many ICF categories are measured by more than one measurement instrument. Many measurement instruments have already been linked to the ICF to facilitate such a comparison[11] and a number of publications can be taken into account concerning the issue of linking procedures.[12–14] Redundancy can be quantified by dividing the overlapping ICF categories through the total of assessed ICF categories in two specific measurement instruments. The closer this ratio comes to the value of 0, the lower the redundancy of the two considered measurement instruments. However, it is not possible to give a uniformly valid cutoff point for acceptable redundancy; comparison of redundancy of different combinations of measurement instrument might identify the most efficient combination.

Efficiency

A measurement instrument or combination of measurement instruments should contain as few items as possible that measure variables outside the domain set. Item efficiency can be quantified as the number of items that address the domain set in relation to the total number of items. The closer this ratio comes to the value of 1, the higher the efficiency of the measurement instrument or combination of measurement instruments is.

$$\textit{Item efficiency} = \frac{\textit{Number of items that cover defined ICF categories in the Domain Set}}{\textit{Total number of items}}$$

A second aspect of efficiency refers to the number of items assigned to a single ICF category to be covered. A highly efficient instrument in this context is an instrument that uses as few items as possible per defined ICF category. One can easily quantify the efficiency of items to cover an ICF category by looking at the number of items that address a given ICF category. Efficiency decreases with an increasing number of items used to cover the considered ICF category. The closer this ratio comes to the value of 1, the higher the efficiency with which the ICF category is measured.

$$\textit{Item efficiency per category} = \frac{1}{\textit{Number of items to cover one ICF category}}$$

Level of Detail of Information

Level of detail of information is another aspect of efficiency and involves mainly two issues: the number of items assessing a single ICF category and the response scale (e.g., metric *vs.* dichotomous). The level of detail of information is dependent on the objective of a specific study. The score should correspond to the level of the ICF category that one intends to measure. The desired level of detail of information is specified by selecting the ICF categories in the domain set. For example, one might want to measure several third-level categories of walking in a study but only the second-level categories in another study. Sometimes, however, more detailed information will be required than is provided in the ICF. The decision on the level of detail of information is largely given by the reasoning on "what to measure," that is, on the definition of the domain set and the category specification.[2]

Comparability

Using well-validated and often-used measurement instruments is usually recommended. Therefore, one may prefer the selection of measurement instruments that are widely used in similar studies in different regions in the world. The use of such measurement instruments allows the pooling and comparison of data. In addition, interpretation and information to the scientific community

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

used to certain measurement instruments may be facilitated or enhanced. Alternatively, the use of an ICF-based approach enhances the comparability of studies and, most importantly, the results of an ICF-based study can be reported as an ICF categorical profile irrespective of the specific measurement instruments used.[8] In addition, it has been shown that items from different patient-reported instruments can be integrated into a psychometrically sound ICF category interval scale to operationalize single ICF categories.[8] Based on the created scale, patients' scores on the original items can easily be transformed into the response options of the ICF qualifier. For example, if we know the response of a person to one of the vitality items of the 36-item Short-Form Health Survey (SF-36),[15] this can be transformed into one of the response options of the qualifier for ICF category b130 Energy and drive functions.[8]

Feasibility

Feasibility concerns, as already suggested by OMERACT,[4] include, on one hand, issues important for researchers (e.g., applicability in the community, costs, legal aspects) and, on the other hand, issues relevant for participants (e.g., acceptability, length, or comprehensibility). For example, it would not be appropriate to use the SF-36 item that asks whether health limits someone in walking more than 1 mile in a population of wheelchair users.[15] In some cases, it might be an option to adopt measurement instruments for a specific population, for example, by skipping or rephrasing questions on walking in a population of wheelchair users.[16] The burden for participants in terms of complexity and length of measurement instruments must be acceptable and when selecting measurement instruments, one has to consider the relation between the length of instrument and response rate.[17,18] To be feasible, measurement instruments must be understandable for all participants and have to be presented in a language that is familiar to them.[19] A high standard of translation and cultural appropriateness must be ensured.[19]

Truth and Discrimination

Measurement instruments should meet validity and reliability requirements that have been summarized as a truth and discrimination criterion by OMERACT.[4] Validity is the extent to which a measurement instrument measures what is intended. Content and face validity assess whether items adequately address the domain of interest. They are qualitative matters of judging whether an instrument is suitable for its proposed application, in our case, for the collection of relevant information to cover the defined domain set. Reliability concerns whether an instrument is internally consistent or reproducible, and it assesses the extent to which an instrument is free from measurement error. Sensitivity to change is another aspect of reliability that should be taken into account.[4] There are numerous approaches to assess these psychometric properties, and there exists extensive literature that describes these approaches in detail.[20,21]

However, validity and reliability should not be seen as fixed properties and must be assessed in relation to the specific population and measurement objectives. In case of use of established measurement instruments, validated and reliable measurement instruments for SCI research should be preferred. In case of the use of single items to assess specific ICF categories, content validity should be ensured. If the case occurs that single items cannot be adopted from other established studies such as the World Health Survey[22] because they do not exist, pretesting of the items is recommended.

RESULTS

Application of Guiding Principles to Select Measurement Instruments

Redundancy

First, we show an example to illustrate how redundancy concerns can be faced. The redundancy principle was encountered through linking eligible measurement instruments to the ICF. For example, the self-report World Health Organization's Disability Assessment Schedule, Version 2,[23] the ICF Measure of Participation and Activities questionnaire,[24] and the SF-36[15] were considered as measurement instruments. By linking these measurement instruments,[25] whenever possible, to an appropriate level of the ICF, overlaps of the measurement instrument can be detected (Table 1).

In addition to the simple demonstration of overlaps between the measurement instruments shown in Table 1, redundancy can be quantified by dividing the redundant ICF categories through the total of assessed ICF categories in two specific measurement instruments. An example of the procedure to quantify redundancy can be found in Table 2. Obviously, the redundancy between the World Health Organization's Disability Assessment Schedule, Version 2[23] and the SF-36[15] is lowest compared with other combinations of measurement instruments.

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

TABLE 1 Example of ICF linking to detect redundancy among the WHODAS II, IMPACT-S, and SF-36

ICF Code	ICF Category	Item in WHODAS II	Item in IMPACT-S	Item in SF-36
b130	Energy and drive functions			4
b140	Attention functions	1	1	
b144	Memory functions	1	1	
b152	Emotional functions	1		8
b210	Seeing functions		1	
b230	Hearing functions		1	
b280	Sensation of pain			2
d1	Learning and applying knowledge	1	1	
d175	Solving problems	1		
d210	Undertaking a single task		1	
d230	Carrying out daily routine	2		2
d240	Handling stress and other psychologic demands		1	
d3	Communication		2	
d310	Communicating with/receiving spoken messages	1	1	
d310-d329	Communicating/receiving		1	
d350	Conversation	2		
d410	Changing basic body position	1	1	2
d415	Maintaining basic body position	1	1	
d430	Lifting and carrying objects		1	2
d435	Moving objects with lower limbs		1	
d440	Fine hand use		1	
d445	Hand and arm use		1	1
d450	Walking	1		3
d455	Moving around		1	3
d460	Moving around in different locations	1		
d470	Using transportation		1	
d475	Driving		1	
d498	Mobility, other specified	1		
d510	Washing oneself	1	1	1
d520	Caring for body parts		1	
d530	Toileting		1	
d540	Dressing	1	1	1
d550	Eating	1		
d570	Looking after one's health		1	
d620	Acquisition of goods and services		1	
d630-d649	Household tasks	3	1	2
d650	Caring for household objects		1	
d660	Assisting others		1	
d7	Interpersonal interactions and relationships	1		
d710	Basic interpersonal interactions		1	
d730	Relating with strangers		1	
d740	Formal relationships		1	
d750	Informal social relationships	2	1	
d760	Family relationships		1	
d770	Intimate relationships	2	1	
d839	Education, other specified and unspecified	2		
d845	Acquiring, keeping, and terminating a job		1	
d850	Remunerative employment		1	2
d859	Work and employment, other specified and unspecified	2		2
d860-879	Economic life	1	1	
d910	Community life	1	1	
d920	Recreation and leisure	1	1	4
d930	Religion and spirituality		1	
d950	Political life and citizenship		1	
e4	Attitudes	1		

Items were coded on the second level of the ICF categories.

ICF, International Classification of Functioning, Disability, and Health; WHODAS II, World Health Organization's Disability Assessment Schedule Version 2 self-rater version[23]; IMPACT-S, ICF Measure of Participation and Activity[24]; SF-36, 36-item Short-Form Health Survey.[15]

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

TABLE 2 Example of a quantification of redundancy among the WHODAS II, IMPACT-S, and SF-36

	Measurement Instruments	Number of ICF Categories/ Redundancy
Total of assessed second-level ICF categories	WHODAS	26
	IMPACT-S	41
	SF-36	15
Total of redundant second-level ICF categories	WHODAS II and IMPACT-S	14
	WHODAS II and SF-36	9
	IMPACT-S and SF-36	9
Redundancy[a]	WHODAS II and IMPACT-S	0.264
	WHODAS II and SF-36	0.281
	IMPACT-S and SF-36	0.191

[a]Number of overlapping ICF categories in relation to total number of assessed ICF categories in the two considered measurement instruments. Values are between 0 and 1, where 0 indicates no overlap; 1, complete overlap.

ICF, International Classification of Functioning, Disability, and Health; WHODAS II, World Health Organization's Disability Assessment Schedule Version 2 self-rater version[23]; IMPACT-S, ICF Measure of Participation and Activity[24]; SF-36, 36-item Short-Form Health Survey.[15]

Efficiency

The application of the efficiency principle is demonstrated as we present two aspects of efficiency based on the formula introduced in "METHODS" that can be taken into account when evaluating the efficiency of measurement instruments. As an example, the domain set that is used as reference to calculate aspects of efficiency is defined as consisting of the 38 second-level ICF categories that are represented in the Brief ICF Core Set for SCI in the early postacute[26] and in the long-term contexts.[27] This exemplary exercise leads to the conclusion that the ICF Measure of Participation and Activities questionnaire scores higher in the evaluated efficiency concerns and therefore would be preferred (Table 3).

Level of Detail of Information

The decision on the level of detail of information is largely given by the reasoning on "what to measure," that is, on the definition of a domain set and the respective category specifications.[2] Briefly, the decision on the level of detail of information, that is, the category specifications for each ICF category in the domain set to be measured is based on the literature on each ICF category in the domain set. More specifically, the following steps should be taken for each ICF category: an electronic

TABLE 3 Example for the evaluation of efficiency in the WHODAS II and IMPACT-S

	WHODAS II	IMPACT-S
Total number of items	37[a]	28[a]
Total of assessed second-level ICF categories	26	41
Number of items that cover ICF categories in the domain set	6	9
Number of items that cover ICF categories not in the domain set	28	33
Number of ICF categories in the measurement instrument in the domain set	6	9
ICF categories not in the domain set	20	32
Item efficiency[b]	0.16	0.32 +
Item efficiency per category[c]	In 1 category: 0.33 In 6 categories: 0.5 In 19 categories: 1.0	In 1 category: 0.5 In 40 categories: 1 +

[a]Items that could not be linked to the ICF on a categorical level are excluded from calculation.

[b]Item efficiency = number of items that cover ICF categories in the domain set/Total number of items.

[c]Item efficiency per category = 1/number of items to cover one ICF category.

+, higher efficiency (not a statistical measure; evaluation is only valid for the specific comparison in this example).

ICF, International Classification of Functioning, Disability, and Health; WHODAS II, World Health Organization's Disability Assessment Schedule Version 2 self-rater version[23]; IMPACT-S, ICF Measure of Participation and Activity.[24]

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

literature search, abstract selection, extraction of the dependent and independent variables used in the extracted studies, and documentation of the variables and relationships studied. Based on this information and expert opinion, a final decision on the category specifications and therefore on the level of detail of information can be made.[2]

Comparability

An example to illustrate how comparability concerns can be faced is given in Table 4. We demonstrate a potential procedure to enhance comparability and use the example of the ICF category b152 Emotional functions. Whenever possible, one should prefer measurement instruments that have been widely applied in other epidemiologic studies on functioning, especially in SCI research. It is thus a valuable exercise to select studies in SCI research with comparable purposes and to examine what measurement instruments were used to assess a specific ICF category. In our example, four longitudinal epidemiologic studies on functioning in SCI were identified: the American Model Systems Database, the Australian Spinal Cord Injury Register, the European Multicenter Study on Spinal Cord Injury, and the Dutch research program "Physical strain, work capacity, and mechanisms of restoration of mobility in the rehabilitation of persons with spinal cord injuries." In addition, we involved two longitudinal epidemiologic studies on functioning in the general population, the Swiss Health Survey 2007 and the World Health Survey 2002. Afterward, we identified the measurement instruments that were used to assess the ICF category b152 in these selected studies (Table 4).

Although this example shows that, similar to many other ICF categories, there is no gold standard to assess b152 Emotional functions, such an overview can inform the selection of a measurement instrument. Nevertheless, the cross-calibration of the scores of items from different measurement instruments might be a solution to enhance comparability.[8]

Feasibility

The aspect of feasibility is at the core of the decision in the example of the SwiSCI Study because the amount of ICF categories and related specifications of these ICF categories[2] is way beyond the number of items contained in a generic instrument. Therefore, to keep the length of the measurement battery within reasonable limits, nonredundancy and efficiency are highly important criterion for selection as discussed previously.[17]

Truth and Discrimination

In case of use of established measurement instruments, validated measurement instruments for SCI research are preferred. An example for an instrument that has been tested in an SCI population is the Spinal Cord Independence Measure (SCIM, version III), an instrument that is used to assess independence in the performance of daily

TABLE 4 Overview of measurement instruments used in longitudinal epidemiologic studies on functioning in SCI or in the general population to assess the ICF category b152 emotional functions

Epidemiologic Study in SCI/General Population	Setting, Population	Measurement Instrument to Assess b152 Emotional Functions
American Model Systems Database	America, SCI population	9-item depression scale of the Patient Health Questionnaire–9[29]
Australian Spinal Cord Injury Register	Australia, SCI population	–
Dutch research program "Physical strain, work capacity, and mechanisms of restoration of mobility in the rehabilitation of persons with spinal cord injuries"	The Netherlands, SCI population	36-Item Short-Form Health Survey Mental Health Scale[15]
European Multicenter Study on Spinal Cord Injury	18 European countries, SCI population	Hamilton Depression Scale[30]
Swiss Health Survey 2007	Switzerland, general population	Wittchen-Scale[31]
World Health Survey 2002	World Health Organization World Region, general population	Single item: "Overall in the last 30 days, how much of a problem did you have with feeling sad, low, or depressed?"[22]

ICF, International Classification of Functioning, Disability, and Health; SCI, spinal cord injuries.

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

activities in persons with SCI.[28] If measurement instruments are not validated for an SCI population, ideally, a measure that has been validated in other populations should be chosen. For example, the ICF Measure of Participation and Activities questionnaire has not been validated in an SCI population but has been tested in a population of road accident victims.[24] If such a measure is also not available, or if feasibility requires the use of single items to measure a range of ICF categories, items taken from existing measures are to be preferred over newly developed items. In case that the development of new items is inevitable, at least cognitive testing is recommended.

DISCUSSION

Besides guiding principles from the literature, the transfer from "what to measure" to the measurement instrument selection implies new approaches that had not been shown before: the linkage rules enable us to assess efficiency, comparability, and redundancy, and the definition of a domain set allow an efficiency assessment of measurement instruments. We transparently showed that guiding principles offer potential tools to face the multifaceted challenges that one encounters in the selection of measurement instruments to cover specific domain sets. The development of guiding principles allows a systematic and reasoned process of measurement instrument selection that may serve as model in the planning of future epidemiologic studies on functioning.

Obviously, it is not possible that the selection of measurement instruments satisfies the entire range of guiding principles, but, ideally, the selection represents the best possible and feasible decision. Although we judge this process of measurement instrument selection based on guiding principles as valuable, researchers are confronted with several limitations concerning their application. Researchers may expect a quantification, prioritization, or uniformly applicable rules. However, this is not achievable for a series of reasons.

A methodologic limitation lies in the difficulty to quantify the applicability of guiding principles and to set cutpoints to assess whether a criterion is fulfilled or not. For example, it is possible to calculate indictors for efficiency, but it is not possible to give a universally applicable cut point to evaluate if efficiency is appropriate or not since efficiency concerns largely depend on the predefined category specification. Therefore, a transparent reflection of relevant aspects must be made from case to case to come up with an evaluation on whether a measurement instrument is in line with the guiding principle or not.

Another limitation concerns the prioritization of the guiding principles because they cannot be prioritized in terms of importance; instead, they should be considered in relation to the proposed application, that is, the specific characteristics and purpose of the planned study. The process of prioritization should be guided by "objective" and well-reasoned criterion that can be justified by researchers. For example, given the fact that the purpose of the SwiSCI Study is to cover a large amount of ICF categories defined in the domain sets,[2] feasibility and efficiency concerns were strong arguments to guide the selection of measurement instruments and were therefore prioritized. Concerning the prioritization of comparability, one has to decide on the envisioned comparison group that can be either a generic (e.g., general population) or specific (e.g., SCI community). The existence of norms in the interpretation of data might also enhance comparability with other study samples. It is important to mention that the linking of items to ICF categories indeed enhances comparability, but which category specifications are measured in comparable studies because ICF categories do not describe a priori the specifications of an ICF category need to be defined.[2]

The impossibility of prioritizing guiding principles leads to another limitation that concerns the fact that guiding principles strongly interact. Therefore, the decision for a specific measurement instrument often claims a trade-off between different guiding principles. For example, the detail of information is largely linked to efficiency, and therefore, researchers have to decide on the trade-off between efficiency and the detail of information that is most appropriate. A potential solution to face this challenge is to clearly define the focus of the study, that is, to clarify which domains are central or which ones need fewer specifications as it has been made in the decision concerning "what to measure."

Concerning comparability, we face the limitation that there is no gold standard in measurement instrument use, that is, that widely used measurement instruments must not be the best ones, do not meet our needs, or do not exist at all. Sometimes, it is worth introducing a new or less used measurement instrument that seems to fit better, that is, scores higher in terms of efficiency. In the case of lack of measurement instruments to address specific ICF categories, alternative specifications for relevant ICF categories are needed;

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

therefore, the principle of comparability might be violated.

CONCLUSIONS

The application of a new set of guiding principles presented in this paper allows a systematic and reasoned process of measurement instrument selection. The examples given in this article may serve as a methodologic tool for other studies because the presented procedures can be transferred to upcoming epidemiologic studies on functioning. Furthermore, this paper is a reference for future reporting of results and may serve as a justification for the selection of measurement instruments that are available to the public and the scientific community.

ACKNOWLEDGMENTS

We thank Dr. Martin Brinkhof, Dr. Reuben Escorpizo, Dr. Szilvia Geyh, Dr. Jan D. Reinhardt, and Dr. Erik von Elm for comments and suggestions.

REFERENCES

1. World Health Organization: *International Classification of Functioning, Disability and Health: ICF*. Geneva, Switzerland: WHO; 2001
2. Cieza A, Boldt C, Ballert C, et al: Setting up a Cohort Study on functioning: deciding what to measure. *Am J Phys Med Rehabil* 2011;90(suppl 2):S17–28
3. Eriks-Hoogland I, Cieza A, Post M, et al: Category specification and measurement instruments in large spinal cord injury studies: a comparison using the international classification of functioning, disability, and health as a reference. *Am J Phys Med Rehabil* 2011;90(suppl 2):S39–49
4. Boers M, Brooks P, Strand CV, et al: The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198–9
5. Stucki G, Boonen A, Tugwell P, et al: The World Health Organization International Classification of Functioning, Disability and Health: A conceptual model and interface for the OMERACT process. *J Rheumatol* 2007;34:600–6
6. Cieza A, Brockow T, Ewert T, et al: Linking health-status measurements to the International Classification of Functioning, Disability and Health. *J Rehabil Med* 2002;34:205–10
7. Cieza A, Geyh S, Chatterji S, et al: ICF linking rules: An update based on lessons learned. *J Rehabil Med* 2005;37:212–8
8. Cieza A, Hilfiker R, Boonen A, 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;62:912–21, 921.e1–3
9. Stucki G, Kostanjsek N, Ustun B, et al: ICF-based classification and measurement of functioning. *Eur J Phys Rehabil Med* 2008;44:315–28
10. Post MWM, Brinkhof MWG, von Elm E, et al: Design of the Swiss Spinal Cord Injury Cohort Study. *Am J Phys Med Rehabil* 2011;90(suppl 2):S5–16
11. Geyh S, Cieza A, Kollerits B, et al: Content comparison of health-related quality of life measures used in stroke based on the international classification of functioning, disability and health (ICF): A systematic review. *Qual Life Res* 2007;16:833–51
12. Coenen M, Cieza A, Stamm TA, et al: Validation of the International Classification of Functioning, Disability and Health (ICF) Core Set for rheumatoid arthritis from the patient perspective using focus groups. *Arthritis Res Ther* 2006;8:R84
13. Geyh S, Kurt T, Brockow T, et al: Identifying the concepts contained in outcome measures of clinical trials on stroke using the International Classification of Functioning, Disability and Health as a reference. *J Rehabil Med* 2004;44:56–62
14. Scheuringer M, Kirchberger I, Boldt C, et al: Identification of problems in individuals with spinal cord injury from the health professional perspective using the ICF: A worldwide expert survey. *Spinal Cord* 2010;48:529–36
15. Ware JE, Sherbourne CD: The MOS 36-item Short-form Health Survey (SF-36). A conceptual framework and item selection. *Med Care Instrument* 1992;30: 473–83
16. Lee BB, Simpson JM, King MT, et al: The SF-36 walk-wheel: A simple modification of the SF-36 physical domain improves its responsiveness for measuring health status change in spinal cord injury. *Spinal Cord* 2009;47:50–5
17. Edwards PJ, Roberts I: Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev*. 2009;3:MR000008
18. Jepson C, Asch DA, Hershey JC, et al: In a mailed physician survey, questionnaire length had a threshold effect on response rate. *J Clin Epidemiol* 2005; 58:103–105
19. Harkness J, Bloom A: *European Social Survey: Round 3 ESS Translation Guidelines*. 2006, Mannheim.
20. Fitzpatrick R, Davey C, Buxton MJ, et al: Evaluation patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:i–iv, 1–74
21. Furr M, Bacharach V: *Psychometrics: An Introduction*. Thousand Oaks, CA: Sage Publications; 2008
22. World Health Organization: WHO survey data centre: World Health Survey. 2007. Available at: http://www.who.int/healthinfo/survey/whsshortindividuala.pdf. Accessed November 8, 2010
23. World Health Organization: WHODAS II: Disability assessment schedule. 2000. Available at: http://

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

www.who.int/icidh/whodas/. Accessed empiric study. Accessed November 8, 2010

24. Post MW, de Witte LP, Reichrath E, et al: Development and validation of IMPACT-S, an ICF-based questionnaire to measure activities and participation. *J Rehabil Med* 2008;40:620–7
25. Cieza A, Stucki G: Content comparison of health-related quality of life (HRQOL) instruments based on the International Classification of Functioning, Disability and Health (ICF). *Qual Life Res* 2005;14:1225–37
26. Kirchberger I, Cieza A, Biering-Sorensen F, et al: ICF Core Sets for individuals with spinal cord injury in the early post-acute context. *Spinal Cord* 2010;48:297–304
27. Cieza A, Kirchberger I, Biering-Sorensen F, et al: ICF Core Sets for individuals with spinal cord injury in the long-term context. *Spinal Cord* 2010;48: 305–12
28. Catz A, Itzkovich M, Tesio L, et al: A multicenter international study on the Spinal Cord Independence Measure, version III: Rasch psychometric validation. *Spinal Cord* 2007;45:275–91
29. Spitzer RL, Kroenke K, Williams JB: Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;282:1737–44
30. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62
31. Wittchen HU: Die Studie 'Depression 2000,' Eine bundesweite Depressions-Screening-Studie in Allgemeinpraxen. *Fortschr Med* 2002;118:1–3

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.