CONTENTS page

CONTENTS page

PREFACE

During revision of the International Classification of Impairments Disabilities and Handicaps (ICIDH), the Dutch Collaborating Centre for the ICIDH (WCC) was asked to develop proposals for the revision of the classification of disabilities of the ICIDH (ICIDH-D), including the revision of the *severity scale* of the ICIDH-D. This process of developing a revised severity scale comprises four successive phases. A report on the first two phases of the project was published in 1995 (Hopman-Rock & Miedema, 1995). The current report describes phase 3 of the project, in which we generate a preliminary proposal for the revision of the severity scale by a quantitative analysis of a variety of existing data sources. Disability measures were analysed for the domains: Personal Care, Locomotor disability, Body disposition and Dexterity.

This project was carried out by TNO Prevention and Health, Division of Public Health and Prevention in Leiden, the Netherlands. Part of the research of this division is directed toward prevention of the consequences of illness and ageing. This project was financially supported by a grant from the WHO Collaborating Centre for the ICIDH in the Netherlands.

ACKNOWLEDGEMENTS

We would like to thank Dr. M.W. De Kleijn-de Vrankrijker for initiating this project and for her continued support and helpful comments during the course of phase 3 of this project. She was chairperson of the committee of experts (The Classification Development Group on Disablement) for the project. In addition, we would like to thank all those researchers who responded to our request to make their data bases available for our research. In this report, data files were analysed from: Dr. M.H. Liang (Department of Medical Rheumatology, Brigham and Women's Hospital, Boston MA, USA), Dr. T.P.B.M. Suurmeyer (Northern Centre for Health Care Research, University of Groningen, The Netherlands), Drs. M. Hopman-Rock (TNO Prevention and Health Leiden, in collaboration with the Rotterdam study, Dr. E. Odding, Erasmus Medical School, Rotterdam, The Netherlands), and Mrs. C. Molleman (Higher Institute of Labour, Leuven, Belgium). Data from the Netherlands Health Interview Survey (Netherlands Central Bureau of Statistics, The Hague, The Netherlands) were also used, which were obtained with the help of the Wetenschappelijk Statistisch Agentschap NWO.

SUMMARY

Within the context of the revision of the classification of disabilities of the ICIDH, this project aims to develop a proposal for the revision of the *severity scale* for disabilities in Personal Care, Locomotion, Body Disposition and Dexterity; ICIDH-D codes 30-69 of the original ICIDH (1980, reprint 1993). The project is divided into four phases. The first two phases (including the selection of disability measures) have been described in an earlier TNO report (Hopman-Rock & Miedema, 1995). The present text is the report of the *third phase*. The goal of this phase is to develop a working methodology that facilitates and enhances scale development by appropriate quantitative analyses of existing data. An important by-product of this phase is a preliminary proposal for a new Severity of Disabilities Scale (SDS) in the domains mentioned above.

The data used in this phase consisted of responses to different questionnaires and rating scales. The following instruments were included:

- Arthritis Impact Measurement Scale (AIMS);
- Functional Status Index (FSI);
- Sickness Impact Profile (SIP);
- Health Assessment Questionnaire (HAQ);
- Groningen Activity Restriction Scale (GARS);
- OECD long-term disability questionnaire (OECD);
- Physical Performance Test (PPT).

Six data sources containing data on these instruments were combined into two common data sets (total N=± 2500), one containing *walking* items (ICIDH-D 40) and one containing *dressing* items (ICIDH-D 35/36). Walking and dressing items were chosen because these were the disability items that occurred most frequently in the available data.

Different items were administered to different populations that have different levels of disability, which creates a daunting analysis problem. However, by carefully exploiting the overlap between the different data sets we were able to analyse the combined data by advanced psychometric methods. Polytomous Rasch analysis, multiple imputation and factor analysis were used to place different items onto a single common *interval scale*. A property of an interval scale is that the difference between scores 1 and 2 is equal to the difference between scores 2 and 3 (for example, Fahrenheit and Celcius

scales are both interval scales). For each transition between two successive response categories of the same item, the polytomous Rasch model contains a *threshold parameter*. This parameter positions response categories on an imaginary 'severity of disabilities' scale. The analysis allowed us:

- to compare individual response categories of the same item (in terms of severity);
- to compare response categories across different items (in terms of severity);
- to estimate the severity of disabilities per person and data source.

We found that the order of the response categories as obtained by the Rasch model corresponded quite well with our intuitive notions about the items. The same holds for the estimated levels of disability per data source. A detailed technical validation of the model did not reveal clear conceptual discrepancies between the mathematical model and the data. We obtained slight evidence that the time to perform a certain task could be an independent component of the concept of severity. The present data set is, however, too limited to study this aspect in sufficient detail.

Using the threshold parameter estimates, we found seven levels of severity for walking and dressing disability. For walking, the items could be roughly grouped as follows (in increasing order of severity):

0 If walking disability is not present, if a person is able to walk 15 meters in less than 20 seconds;

1 Inside walking with mild pain, walking inside and outdoors with some or mild difficulty, walking more slowly;

2 Much difficulty walking outdoors, moderate difficulty walking inside, often moderate pain is present, only short distances can be walked;

3 Walking with the use of an aid (cane, crutches, artificial limbs, walking frame, etc.);

4 Walking outdoors is only possible with the help of someone else, and inside with much difficulty;

5 Walking is only possible with help, unable to walk outdoors.

Dressing items could be grouped as:

0 No problems at all, can put a coat on and take it off in less than 10 seconds;

1 Mild difficulty with underpants and shoes, slower with the coat;

2 Dressing with some difficulty, mild problems with buttoning, again slower with the coat;

3 Buttoning with moderate difficulty and trouble with shoes;

4 Dressing and putting shoes on with much difficulty, very slow and unable to button clothes;

5 Dressing with the aid of special devices, help of someone else or with severe difficulty, dressing only with the help of someone else.

These lists showed that substantial empirical differences in severity exist between several levels of the current SDS category ‘difficulty in performance’. Based on this result and taking into account earlier criticisms of the SDS, we tentatively propose a new SDS. Table *i* contains our proposal for a new SDS in the domains of Personal Care, Body Disposition, Locomotor and Dexterity.
The term ‘difficulty’ is an abstraction that subsumes matters such as ‘pain involved’, ‘time taken’, ‘number of errors’, ‘clumsiness’, and so on. It will be clear that, for a given type of disability, each category of the SDS needs an *operational definition* that describes the specific category in terms of a number of observable characteristics. The walking and dressing lists just given are examples of such

Table i *Preliminary proposal for the Severity of Disabilities Scale of the ICIDH*

1980 code	Proposed code	Label	Description
0	0	Not disabled	individual can perform the activity or sustain the behaviour unaided without any difficulty
1	1	Some or mild difficulty in performance	individual can perform the activity or sustain the behaviour unaided but only with some or mild difficulty
1	2	Moderate difficulty in performance	individual can perform the activity or sustain the behaviour unaided but only with moderate difficulty
1	3	Much or severe difficulty in performance	individual can perform the activity or sustain the behaviour unaided but only with much or severe difficulty
2	4	Aided performance	individual can perform the activity or sustain the behaviour only with a physical aid or appliance
3,4,5	5	Assisted performance	individual can perform the activity or sustain the behaviour only with assistance of another person
6	6	Complete inability	individual cannot perform the activity or sustain the behaviour
8	8	Not applicable	
9	9	Severity unspecified	

definitions. Likewise, ‘mild’, ‘moderate’ and ‘severe’ are gradations in difficulty that only get a precise meaning through the operational definition.

A controversial issue is whether severity of disabilities should be defined, measured and interpreted *with* or *without* aids and appliances. The analyses done to date indicate that respondents usually considered ‘performance with difficulty’ as a less severe disability than ‘aided performance’, which in turn is considered less severe than ‘assisted performance’. This suggests that in practice severity of disabilities is more likely to be interpreted and measured as the severity *without* aids and appliances. We have therefore preserved the ‘aided performance’ category in our preliminary proposal. The number of items on which this conclusion is based is small however.

A unique aspect of our approach is its empirical basis and its strong emphasis on items that are actually applied in the field. The interpretation and application of the proposed scale might therefore be easier than the current scale. The finer grain on the lower end of the proposed SDS makes it more suitable for applications in public health and prevention. The techniques we use in this report provide keys to conversion issues. It is possible to translate the current SDS into the proposed SDS, to convert the severity as measured by existing disability items into the proposed SDS, or to convert existing items into other (existing or novel) disability items. Such possibilities will preserve much valuable work. Finally, because of the strict mathematical basis of the model, formal tests on aspects of reliability and validity of the scale become available.

Since the present work was a first-time application of the Rasch model to this field, compromises and limitations were inevitable. First, only two types of disabilities (walking and dressing) have been used to generate the proposed SDS. Second, our coverage of instruments that measure disability is far from complete. Third, our sample contains very few severely disabled people. Fourth, because of the sheer incompleteness of the data, only rigid models with strong assumptions could be applied. Fifth, the threshold estimates are very variable because many item categories contain only a few observations. Sixth, we were forced to assume equality between two dressing items in order to get a linkage between them. Seventh, we did not take into account any differences in the mode of data collection (self report, interview, observation).

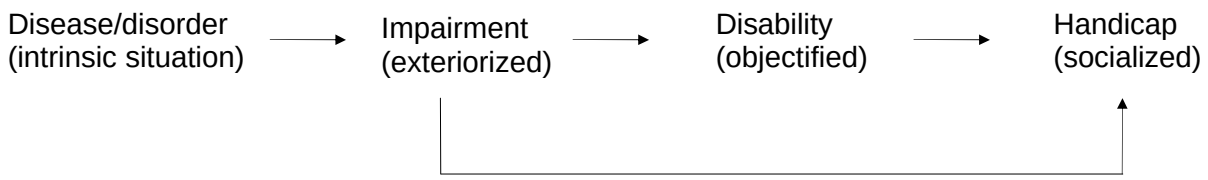
Especially the first four of these points could seriously affect the outcome. Since these problems are mostly of organizational and technical nature, many, if not all, of these problems can be overcome in a dedicated effort. However, before embarking on such a journey we would appreciate to obtain feedback from the forum of experts, particularly with respect to the relevance and clarity of the present approach. Armed with these insights, we could then set course for a SDS that can count on ample support.

1. INTRODUCTION

1.1 General introduction

In 1980 the World Health Organization published the International Classification of Impairments, Disabilities and Handicaps (WHO, 1980; reprint 1993). This manual classifies the *consequences* of diseases, disorders and injuries. The ICIDH can be used in several settings such as clinical health care, rehabilitation, nursing homes, vocational rehabilitation, population surveys, etc. figure 1.a shows the concepts used in the ICIDH.

Figure 1.A Concepts of the ICIDH (1980, 1993)



The underlying concept of the ICIDH is that diseases may lead to impairments, which, in turn, may induce handicaps either directly or indirectly, via disability. In practice, the situation is much more complex than is suggested in figure 1.a (Brandsma et al., 1995). Disability may result from a handicap (for example a walking disability after a person has been confined to bed) or an impairment may result from a disability (for example an impairment of mood following a physical disability). Moreover, it is possible to be impaired without being disabled and disabled without being handicapped. This is one of the issues which will be changed in the revised ICIDH (to be published in 1999).

The following definitions are used in the ICIDH (1980, 1993):

Impairment in the context of health experience, an impairment is any loss or abnormality of psychological, physiological, or anatomical structure or function;

Disability in the context of health experience, a disability is any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being;

Handicap in the context of health experience, a handicap is a disadvantage for a given individual, resulting from an impairment or a disability, that limits or prevents the

fulfillment of a role that is normal (depending on age, sex, and social and cultural factors) for that individual.

These definitions are currently being modified according to recent insights. In addition to the classification of disabilities the ICIDH has two additional scales: ‘severity’ and ‘outlook’. The *severity* of a disability reflects the degree to which an individual’s ability to perform an activity is restricted. The *outlook* reflects the likely course of the individual’s disability status. In this report we focus on the severity scale of the ICIDH.

1.2 Severity of Disabilities Scale of the ICIDH

Whereas the ICIDH classification of disabilities emphasizes *the restriction or lack of ability* to perform an activity, the severity scale of disabilities is meant to reflect *the extent* to which an individual’s ability to perform is restricted. The ICIDH (1980, 1993) currently classifies the severity of a disability in one of seven categories, ranging from ‘difficulty in performance’ to ‘complete inability’. table 1.a contains the current classification.

The categories of severity are related to intervention goals. ‘Enhancement’ (when performance is difficult) is often needed in category 1, ‘supplementation’ (when a aid is needed for performance) is a typical action in categories 2-4 and ‘substitution’ (when no performance is possible, even with aid) may occur in categories 5 and 6. Yet, many exceptions and special situations exist in which this general relation is not adequate. A summary of criticisms of the severity or disabilities scale (SDS) of the ICIDH is given in the report on the first two phases of the project (Hopman-Rock & Miedema, 1995).

Table 1.A Current Severity of Disabilities Scale of the ICIDH

Code	Label	Includes
0	Not disabled	no disability present (the individual can perform the activity or sustain the behaviour unaided and on his own without difficulty)
1	Difficulty in performance	difficulty present (the individual can perform the activity or sustain the behaviour unaided and on his own but only with difficulty)
2	Aided performance	aid and appliance necessary (the individual can perform the activity only with a physical aid or appliance)
3	Assisted performance	the need for a helping hand (the individual can perform the activity or sustain the behaviour, whether augmented by aids or not, only with some assistance from another person)
4	Dependent performance	complete dependence on the presence of another person (the individual can perform the activity or sustain the behaviour, but only when someone is with him most of the time). Excludes: inability
5	Augmented inability	activity impossible to achieve other than with the help of another person, the latter needing an aid or appliance to enable him or her to provide this help (for example, the individual cannot get out of bed other than by the use of a hoist); behaviour can be sustained only in the presence of another person and in a protected environment
6	Complete inability	activity or behaviour impossible to achieve or sustain (for example, an individual who is bed-bound is also unable to transfer)
8	Not applicable	
9	Severity unspecified	

1.3 Summary of results of phase 1 and phase 2

In the last decade, several methods have been used to assess the presence and severity of disabilities. Some investigators have used (parts of) the ICIDH-D, with or without its severity scale, while others have included questions about disability in population surveys or patient questionnaires. A method for indicating the severity of disabilities that has become very popular is the use of standardized measures (often patient questionnaires), in which the degree of restriction in performing certain activities is assessed by using an ordinal scale. Often, an overall (severity of) disability score can be calculated. A large number of these measures have been developed, especially in the field of (instrumental or extended) activities of daily living (ADL). However, these measures can often be used only for certain research projects, research groups or fields of research. In most cases there is no direct relationship between the disability score (item or sum) obtained with these measures and the severity scale of the ICIDH-D (1980, 1993).

In the *phase 1* of the project an inventory was made of 96 measures, currently used in different (research) areas such as population surveys and statistics, rehabilitation, vocational assessment,

nursing homes, etc. The severity of disabilities could be graded with 77 of these measures. On the basis of the definition/formulation of severity, these measures could be grouped into seven clusters (categories). For each measure, we assessed the way in which severity of disabilities was defined, the reliability, validity, relation to ICIDH (on 2-digit level), and the cluster concerned. In addition, we asked researchers and clinicians in various fields of research and patient care to give their opinion about the measures that they use and about the current severity scale of the ICIDH-D.

In the *phase 2* of the project, we analysed the relationship between a selection of the most popular, valid and reliable disability instruments and the categories of the ICIDH-D (on the 3-digit level). The selected instruments were (see Table 1): Health Assessment Questionnaire, Functional Status Index, OECD Long-Term Disability Questionnaire, Pediatric Evaluation of Disability Inventory (PEDI) part I, Revalidatie Activiteiten Profiel (RAP), Physical Abilities Scale, Osteoporosis Functional Disability Scale, Barthel ADL Index, Northwick Park Index of Independence in ADL, Arthritis Impact Measurement Scale, Groningen Activity Restriction Scale, Children's HAQ, Rivermead ADL Scales, Functional Independence Measure, OARS activities of daily living, ADL Rating Scale, Pediatric Evaluation of Disability Inventory (PEDI) part II, OPCS disability scales, Sickness Impact Profile and the Physical Performance Test. Note that these instruments also contain items for measuring impairments and handicaps. We related each of the relevant items in the various instruments to the corresponding category of the ICIDH-D. Disability categories of the ICIDH-D that were measured most frequently were 'walking', 'dressing', 'disability in transfer to the toilet', 'bathing', 'other personal hygiene', 'feeding', 'climbing stairs', 'transfer' and 'subsistence'.

The participants of the ICIDH Revision meeting (Voorburg, the Netherlands, 1994) strongly recommended that this project be continued with phase 3 and phase 4. The aim of phase 3 will be described in the next section. Phase 4 will validate the results of phase 3.

Table 1.B Selected instruments of severity of disabilities in phase 2 (Hopman-Rock & Miedema, 1995)

Cluster	Type of scale	Selected scale or instrument	Number of relevant ICIDH categories
1	ordinal	Health Assessment Questionnaire (HAQ)	15
		Functional Status Index	12
		OECD Long-Term Disability Questionnaire	11
		Pediatric Evaluation of Disability Inventory (PEDI), part I	14
2	ordinal	Revalidatie Aktiviteiten Profiel	15
		Physical Abilities Scale	18
		Osteoporosis Functional Disability Scale	16
3	ordinal	Barthel ADL Index	11
		Northwick Park Index of Independence in ADL	14
		Arthritis Impact Measurement Scale	16
		Groningen Activity Restriction Scale	12
		Children's HAQ	15
4	ordinal	Rivermead ADL Scales	15
		OARS Activities of Daily Living	11
		Functional Independence Measure	12
		ADL Rating Scale	12
		Functional Status Index*	13
		Pediatric Evaluation of Disability Inventory (PEDI), part II*	12
5	weighted items	OPCS Disability Scales	20
		Sickness Impact Profile	17
6	time scores	Physical Performance Test	8

* = see also cluster 1

1.4 Aim of phase 3

The aims of the *entire* project (phases 1 to 4) are:

1) to develop a new severity scale for disabilities in the domains of Personal Care, Body Disposition, Locomotion, and Dexterity;
2) to investigate if the results obtained with currently used disability measures can be used to assess the severity of disabilities indicated in the ICIDH;
3) to relate results obtained with the proposed new severity of disabilities scale concerning particular items of the ICIDH to the results obtained with comparable items of several currently used disability instruments.

For simplification in this stage of the revision process, measures concerning behavioural communicational and situational disabilities were excluded.

In *phase 3* of this project, our aims were:

1) to develop a working methodology that facilitates and enhances scale development by appropriate quantitative analyses of existing data;
2) to make a preliminary proposal for a new severity of disabilities scale, based upon the results of phase 1 and 2 of the project and the results of the analyses included in phase 3 and taking into consideration criticisms of the current SDS.

We first developed a strategy to make it possible to compare the responses to items of different disability measures. Chapter 2 gives a global description of our approach. We collected data files which included different disability instruments for items concerning the same type of disability. Chapter 3 describes these data files. Items in the area of ‘walking’ and ‘dressing’ were most suitable for inclusion in the analyses. Detailed results of the analyses are given Chapters 4 and 5. On the basis of these results we developed a new severity of disabilities scale. Chapter 6 presents this scale. Chapter 6 also contains recommendations concerning phase 4.

2. METHODS

One of the main objectives of phase 2 was to compare different items for measuring the same disability. Our strategy to achieve this goal was to combine existing sources of data for different groups of respondents. Our approach consisted of the following steps:

1 Choose relevant instruments that measure disability;
2 Obtain data sources that contain responses on at least two of these instruments;
3 Select all items that belong to the same ICIDH-D category;
4 Construct a linked data set;
5 Estimate the severity per item category by polytomous Rasch analysis;
6 Order item categories according to severity;
7 Construct a new SDS scale.

Below, we treat some of these steps in more detail.

2.1 Selection of data sources

In phase 2 we developed a list of instruments that contained enough items relevant to ICIDH codes 30-60. In order to compare the items of these instruments, we needed raw data at an item level. To obtain such data, we searched the literature and contacted colleagues. Our criterion to include a study was that it should contain at least two of the selected disability instruments, and data about sex and age. We wrote to several authors of articles which described a relevant data file (see Appendix I). namely, Dr. Liang (Liang et al., 1985; Liang et al., 1990), Dr. Thompson (Kidd et al., 1995), and Dr. Suurmeijer (Suurmeijer et al., 1994; Kempen et al., 1996). We also used some of our own data (Hopman-Rock et al., 1996; Van Hell & Hopman-Rock, 1995; Hopman-Rock, 1994; Odding et al., 1995), and public microdata from the Netherlands Health Interview Survey 1994 (CBS, 1995). In addition, we received a data base with disability items that were scored according to the existing SDS from Mrs. Molleman (Higher Institute of Labour, Leuven, Belgium).

2.2 Data linkage

It is not useful to compare the responses for two items A and B if these items have been administered to different groups, because differences in scores may be due to differences between groups, or to

differences between items. If, however, a third item, C, is measured in both groups, and if item C measures the same construct as items A and B, then items A and B can be compared through the common item, C. In that case, items A and B are said to be linked by item C (Vale, 1986). As an example, consider the walking items SIP1 (“I walk shorter distances or often stop for a rest”) and GARS9 (“Can you, fully independently, walk outdoors (if necessary, with a cane)?”). The SIP1 item has been administered in the ERGOPLUS study, the GARS9 item is part of the EURIDISS study. Appendix II contains the frequency counts on both items. Since both the samples *and* the items differ, there is no sensible way of comparing these distributions. However, both studies also administered the HAQ8 item (“Are you able to walk outdoors on flat ground ?”). It is easy to see (in Appendix II) that the EURIDISS sample has more walking disabilities than the ERGOPLUS sample. The amount of severity that is measured by the SIP1 and GARS9 items can now be compared through the GARS9 item. It is said that SIP1 and GARS9 are linked by HAQ8.

A *linkage diagram* visualizes how items are distributed over different data sets. Figure 3.1 is an example of such a diagram. An important use of the diagram is to infer which items are linked. Items are *linked* if there is a path that connects them. In practice, one often needs to permute the rows and columns of the table to identify a connecting path. Items that are linked can be expressed on a common scale without requiring that the disability scores of distinct data sets are comparable.

An *item catalogue* is a systematic description of the individual items consisting of 1) a generic name identifying the item, 2) the source data set, 3) the name of the variable within this data set, 4) the exact wording of the item, 5) the response categories and 6) the frequency distribution of the responses. For easy reference, items are ordered in the same way as in the linkage diagram (from left to right, within columns from top to bottom). Appendices II and IV contain the item catalogues of the walking (ICIDH 40) and dressing items (ICIDH 35, 36) that we used in our study.

Because we obtained data from different investigators using different software and formats, we had to convert some data in order to construct useful common data sets that we could analyse. The first step was to convert the data into a common format, in our case into a SAS data set (SAS Institute, 1990). Next, appropriate variables were selected from the individual data sets, and the response categories were coded into a common scheme, starting with 1 through the number of categories. Finally, all items pertaining to the same ICIDH-D code were combined into a common data set that could be analysed further.

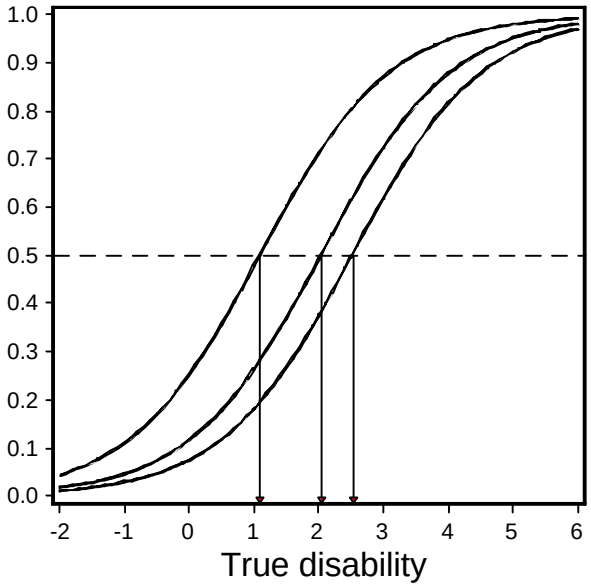
2.3 Polytomous Rasch analysis

The construction and validation of a common scale of ICIDH-D items can be accomplished by applying models of the Item Response Theory (IRT). The goal of IRT is to describe and test the characteristics of individual items, rather than those of complete tests, as occurs in classic psychometric test theory. IRT models are primarily used by educational testing organizations like Educational Testing Service (Princeton, US) and CITO (Arnhem, The Netherlands), but are slowly gaining acceptance as a generally useful model for the construction of questionnaires in public health and related fields.

The Rasch model is the basic IRT model and is particularly suited to construct common scales and to equate information from different sources. A good but technical reference is Molenaar & Fischer (1995). A particular advantage of the Rasch model is that it can produce valid disability estimates even if different subjects have been tested with different, possibly non-overlapping, sets of items. This is precisely what is needed for working with items that come from different sources. Also, the Rasch model can be used to test the unidimensionality of a scale, that is, whether all items measure the same underlying construct.

The Rasch model is restricted to items with only two categories, for example, items with only yes-no answers. However, many disability items used in practice have three or more response categories. For example, the OECD Walking item has four ordered categories. Recoding such polytomous responses into yes-no categories wastes potentially relevant information and may affect the validity of the scale (Roskam & Jansen, 1989). Several versions of the Rasch model have been developed to handle polytomous items (= items with more than two categories). The Graded Response model (Samejima, 1969) is the most useful one for our purpose because it accommodates response categories that are ordered.

Figure 2.A *Graded Response Functions of the OECD Walking item*



The Graded Response model stipulates that the response probability for each of the K categories varies *only* with the true (but unknown) level of disability. More specifically, for each successive pair of categories, the model assumes the existence of a *Graded Response Function* (GRF). figure 2.a displays GRF's of the OECD Walking item. The Graded Response functions are logistic functions of disability. The GRF gives, for each disability level, the probability that a response is in category k or higher (k=1,K). For example, the probability of obtaining a response in category 2 or higher when disability is equal to '-1' is 0.10. Disability is scaled here with mean 0 and variance 1. In the sequel, we will also use an alternative scaling in which the least and the most disables individuals are scored 0 and 100, respectively. For K categories, K-1 GRFs exist. The location parameter of curve k (for k=2,.,K) is called the *threshold parameter* and gives the point on the disability scale at which the probability that the response is in category k or higher is greater than 50%. For the OECD item, we found threshold parameters of 1.09, 2.03 and 2.51. We used the simplest form of the model, in which only the threshold of the logistic function depends on disability. The model can shift the curves along the x-axis, but does not alter any other characteristics of the curve such as the slope or the starting probability. We did not model differences in item discrimination because there were few data. The relevant computations were done using MULTILOG 6.03 (Thissen, 1991).

2.4 Validity and reliability

The Graded Response model assumes that items measure the same construct, which may not be true. The standard approach to this validity problem is to apply factor analysis to the item-item correlations, and then to test whether the correlation pattern among items can be explained by one underlying factor. A complication in our case was is that the matrix of item-item correlations was grossly incomplete because data were missing for many item-item combinations. Most software eliminates incomplete observations from computations, which would discard all data. Also, we could not use PRINCALS with the passive missing data option (Gifi, 1990) because the solution was dominated by the linkage structure and not by the disability scale. As an alternative, we used multiple imputation (Rubin, 1987) to create two versions of a complete data matrix, using an ignorable multivariate normal model (Schafer, 1996). These computation were programmed in SAS/IML as a Gibbs sampler. Details of this procedure are given in van Buuren et al. (1996). Subsequently, we applied maximum likelihood factor analysis to both imputed data sets, using SAS PROC FACTOR (SAS Institute, 1990).

Traditional measures of reliability are item-scale correlations and Cronbach's α. For many individuals, MULTILOG estimates the disability on the basis of only two or three items. This may deceptively raise the raw item-total correlations. Because of missing data, it is complicated to compute the conventional corrections. As an alternative, we estimated Cronbach's α from those item-item correlations that were observed as $\alpha = mr / (1 + (m\text{-}1)r)$, where r is the average correlation and m is the number of items. Assuming that unknown correlations do not systematically differ from observed ones, this coefficient α can be interpreted as the conventional reliability measure when all items have been administered to the entire sample.

3. DATA

3.1 Description of data sources

The following data files were collected and are described here.

LIANG

This data set included five health status instruments that were administered, in random order, to 50 patients with arthritis before and after total joint arthroplasty (Liang et al., 1985). The age of the patients was 50 to 80 years and they had a diagnosis of rheumatoid arthritis or osteoarthritis. Subjects with cognitive impairments, language barriers or visual or hearing deficits were excluded. Four of the five health status instruments (the FSI, the HAQ, the AIMS and the SIP) were on our list of selected disability measures from phase 2 of our project. After one year, a follow-up study was carried out (Liang et al., 1990) on the same group (response n=38). Only data for the last mentioned group were available at item level. The mean age was 67.4 years and 58% of the subjects were women. Eighty-seven per cent of the sample suffered from osteoarthritis and 13% from rheumatoid arthritis (Liang et al.,1990).

ERGOPLUS

This file included data from the Rotterdam study (Hofman et al., 1991) (in Dutch, the ERGO study; ERGO=Erasmus Rotterdam Health Research on Elderly People). The HAQ was used in a study (n=2,895) by Odding (Odding et al., 1995) in 1991, and the SIP was used in 1993 by Hopman-Rock in a sub-sample of 306 persons (Hopman-Rock et al., 1996). All persons were aged 55 to 75 years and lived independently. Four subgroups were identified: chronic pain in the hip and/or knee (reported pain on two occasions in 1991 and one in 1993, n=59), episodic pain (pain on two occasions, n=74), sporadic pain (pain on one occasion, n=101) and a reference group without pain (n=72). The mean age of the sub-sample was 64.8 years (SD=5.6), and 68% of the subjects were women.

EURIDISS

In this data set of 242 patients with recently diagnosed rheumatoid arthritis, the HAQ and the GARS were combined. EURIDISS (European Research on Incapacitating Diseases and Social Support) is an international longitudinal study of patients with recently diagnosed rheumatoid arthritis (Suurmeyer et al, 1994). The mean age of the sample was 53.9 (SD=11.8) and 64% of the subjects were women.

CBS-GE

This file is a public microdata file of health survey interviews conducted in 1994 by the Netherlands Central Bureau of Statistics (CBS) for persons aged 16 years and older. Items on long-term disability (parts of the OECD disability indicator: items 1,2,4,5,6,9,10 and 12) and items on Activities of Daily Living linked the data set of EURIDISS and GOW (see next section). These parts of the health interview were only completed by persons aged 55 years and over (N=2,113). The ADL questions were not from any existing list and included:

- eating and drinking
- getting in and out of a chair
- getting in and out of bed
- dressing and undressing
- moving towards another room on the same floor
- walking up and down the stairs
- leaving and entering the house
- moving along outside the house
- washing the face and hands
- washing the entire body.

The answer categories were: without difficulty, with some difficulty, with great difficulty and only with help from others. In the context of phase 3 of the project, we used the items ‘moving towards another room on the same floor’, ‘moving along outside the house’ (both ICIDH 40 walking) and ‘dressing and undressing’. These questions were similar to some of the items of the GARS. The mean age of the sample was 66.6 years (SD 8.8), and 55% of the subjects were women.

GOW

This data set (Goed Oud Worden=Ageing Well) contains information about disability using the OECD disability indicator and with the Physical Performance Test (PPT). Fifty healthy persons aged 75 to 85 years (mean age 78.7 years, SD 3.0) took part, 58% of whom were women (van Hell & Hopman-Rock, 1995). All persons lived independently in the city of Leiden and were participants of a course called “Ageing Well”. The pre-test measures of disability were used.

DETER

This data set (Determinants of immobility and physical activity) contained information about 30 older people (age 75 years and over) living in the city of Leiden and waiting for home care (on a

waiting list). In a small pilot study (Hopman-Rock, 1994) all people were visited and asked about their diseases/disorders, their complaints and their physical activity. Their attitude about exercise and their self-efficacy in exercising were also measured. One of the functional tests was the PPT. The OECD indicator was also used. The mean age of the sample was 78.6 years, (SD 3.2) and 70% of the subjects were women.

VFSIPH

In this data set (survey of by C. Molleman, project leader E. Samoy, Belgium, 1995), disability was assessed in a survey among a population of mentally handicapped by using items similar to those of the Functional Impairment Measurement (FIM). The answer categories, however, were in accordance with the existing SDS of the ICIDH. The questions were asked in three different settings, which are (in increasing order of disability): day-care centres (N=2,302; mean age 34 years), homes for working people (N=1,001; mean age 38 years; 20% needed help from other persons), and homes for non-working people (N=5,909; mean age 38 years, 75% needed help with ADL from other persons).

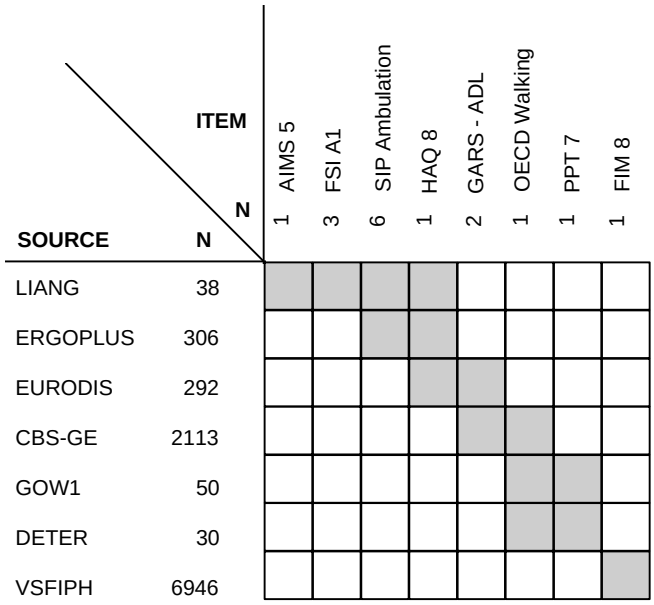
The exact wording of the disability items that we used is given in Appendices II and III.

3.2 Linkage of walking items (ICIDH-D 40)

figure 3.a is the linkage diagram for the walking items. The rows of the diagram contain the name of each data source and the number of observations in that source. The columns contain item blocks that are identified by name and the number of items it contains. There were 16 different walking items, and six of these were items of the SIP. Cells are colored if the item(s) was(were) present in the data source. Thus, item AIMS5 (“Are you unable to walk ...”, see Appendix II) was used in Liang’s investigation only and not in the other sources. figure 3.a displays a path from items AIMS5 through PPT7, so these items are linked. In contrast, item FIM8 was not linked and could be not included in the analysis.

Based on the item catalogue of Appendix II, a data set was constructed using items AIMS5 through PPT7. Data that were missing as a result of the linkage structure were coded as zeroes, and treated by MULTILOG as missing. Observations that had missing data for other reasons, for example because subjects did not respond to the item, were left out of the data set. About 12 percent of the observations were deleted for this reason. Item SIP6 (‘I do not walk at all’) was left out because no

Figure 3.A Item-source linkage diagram of walking items (ICIDH-D 40).



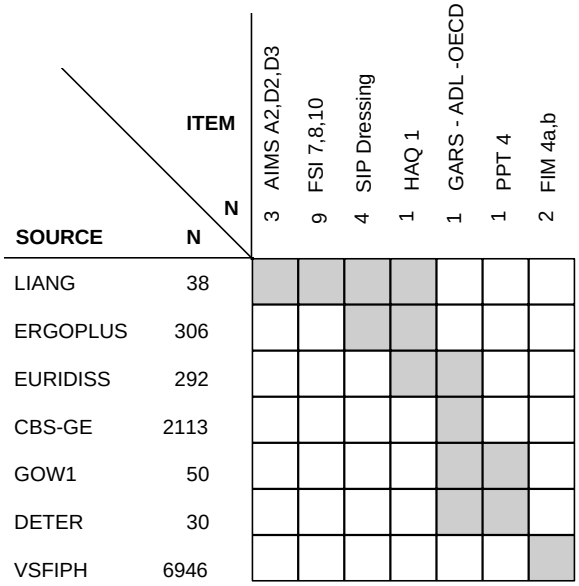
one responded to the 'yes' category. The remaining data set consisted of the scores of 2,484 individuals for 14 linked items. This data set contained 76% missing data.

3.3 Linkage of dressing items (ICIDH-D 35,36)

Both ICIDH-D codes 35 and 36 refer to dressing disability. It was difficult to assign dressing items to either code 35 or code 36 in a coherent way. For this reason we decided to analyse all dressing items together, to see if different ICIDH codes could be identified a posterior. Appendix III contains the exact descriptions of the individual items.

The linkage diagram in figure 3.b shows that there were 21 dressing items. The overall linkage structure was similar to that for the walking diagram, but with one important difference. Because the OECD dressing item was not part of the CBS-GE survey, the link between EURIDISS and GOW1 was broken. We restored the link by assuming that the GARS, ADL and OECD items were identical questions, even though there were some differences in wording. For example, it could be that the OECD item is slightly easier (i.e. measures less disability) than the GARS because it misses the expression 'geheel zelfstandig' (Dutch for 'fully independent'). The combination of items was called 'GAO'. In interpreting the results, the reader should be aware that the GAO item is a compromise.

Figure 3.B *Item-source linkage diagram of dressing items (ICIDH-D 35,36)*



The data set consisted of items AIMS A2 through PPT4. Some items (AIMS A2, FSH2) were deleted because they had zero frequencies. About 6 percent of the observations were deleted because they had missing values. The remaining data set contained data for 2670 individuals and 17 items. Unfortunately, because of a programming error, items FSG3, FSH3 and FSJ3 were replaced by the data of FSG4, FSG4 and FSJ4 respectively. This was discovered only after all computations were done, and no time was available to re-run the analyses. The analyses given below therefore apply to 14 instead of 17 items.

4. RESULTS FOR ICIDH-D 40: WALKING

4.1 Threshold parameters of walking items

We applied the MULTILOG to the walking data under the postulate that all observations derive from one common distribution, which is normal with zero mean and unit variance. The scale in which the results are expressed is arbitrary. To aid interpretation, we scaled all results such that the individuals with the lowest and the highest disability were scored as 0 and 100, respectively. The transformation to do this was Y = 9.1 + 28.7 X, where X represents the MULTILOG output. Note that the end points of this scale depend on the specific sample being used.

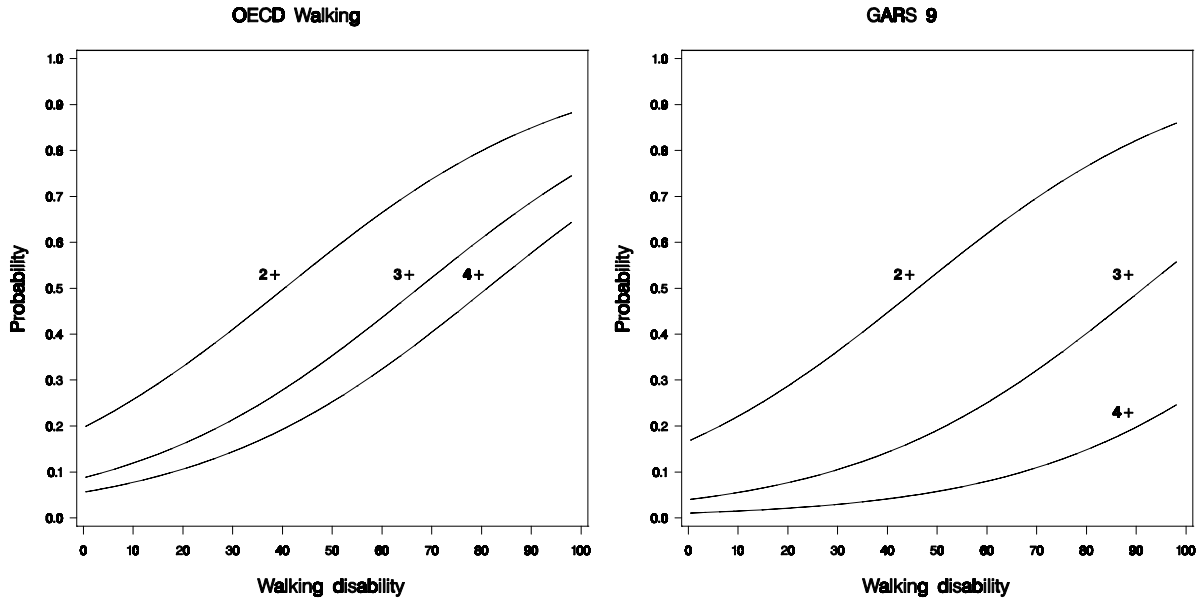
table 4.a lists the estimates for the location (or threshold) parameter of the GRFs. The first column of table 4.a presents the raw *threshold* estimates of the Graded Response model. The threshold is the level of disability at which 50% of the people will respond in the higher category. The notation 'a:b' is used to indicate category b of item a. For category GARS7:4, we find a threshold of 6.83, which means that about half of a sample that has an average disability of 6.83 (which is extremely high) will have a score in category 4.

The results are sorted by the threshold parameter. Therefore, the categories in the rows represent a gradual decrease in the severity of disabilities. So, category GARS7:4 measures more severe disability than HAQ8:4, GARS9:4 or OECD:4. The table clearly demonstrates that items with high thresholds are usually associated with inability to walk or walking only with help from others. Difficulty in walking inside is considered to be more severe than difficulty in walking outdoors (c.f. GARS7 and GARS9). The PPT7 item (a Dutch item about how long it takes to walk 15 meters) measures only mild forms of disability. The ordering produced by the Rasch analysis seems to be appropriate and matched our expectations.

Table 4.A Estimated threshold parameters of walking items

Threshold parameter of the Graded Response Model	Normalized threshold parameter	Item:Upper Category	Number of respondents in upper category	Description of the upper category
6.83	205	GARS7:4	4	inside: only with help
6.41	193	HAQ8:4	2	outdoors: unable
5.41	164	SIP8:2	2	only walk with help
4.72	145	GARS7:3	26	inside: much difficulty
4.22	130	GARS9:4	49	outdoors: only with help
3.96	123	FSI_H:2	1	inside: used cane, etc.
3.84	119	SIP11:2	9	use frame, crutches, etc.
3.27	103	SIP7:2	15	limp, wobble, etc.
2.91	93	PPT7:5	6	cannot walk 15 m
2.87	91	GARS9:3	106	outdoors: much difficulty
2.74	88	FSI_D:3	3	inside: moderate difficulty
2.73	88	FSI_P:3	3	inside: moderate pain
2.73	88	HAQ8:3	57	outdoors: much difficulty
2.67	86	AIMS5:2	3	unable unless assisted
2.51	81	OECD:4	174	cannot walk 400 m
2.44	79	GARS7:2	178	inside: some difficulty
2.19	72	SIP1:2	37	shorter distances
2.03	67	OECD:3	72	400 m: much difficulty
1.59	55	PPT7:4	10	15 m: > 25 sec.
1.38	49	FSI_D:2	4	inside: mild difficulty
1.29	46	GARS9:2	312	outside: some difficulty
1.28	46	SIP12:2	71	more slowly
1.09	40	OECD:2	209	400 m: some difficulty
1.09	40	HAQ8:2	116	outdoors: some difficulty
1.06	40	PPT7:3	6	15 m: 20-25 sec.
.93	36	FSI_P:2	6	inside: mild pain
-.22	3	PPT7:2	20	15 m: 15-20 sec.

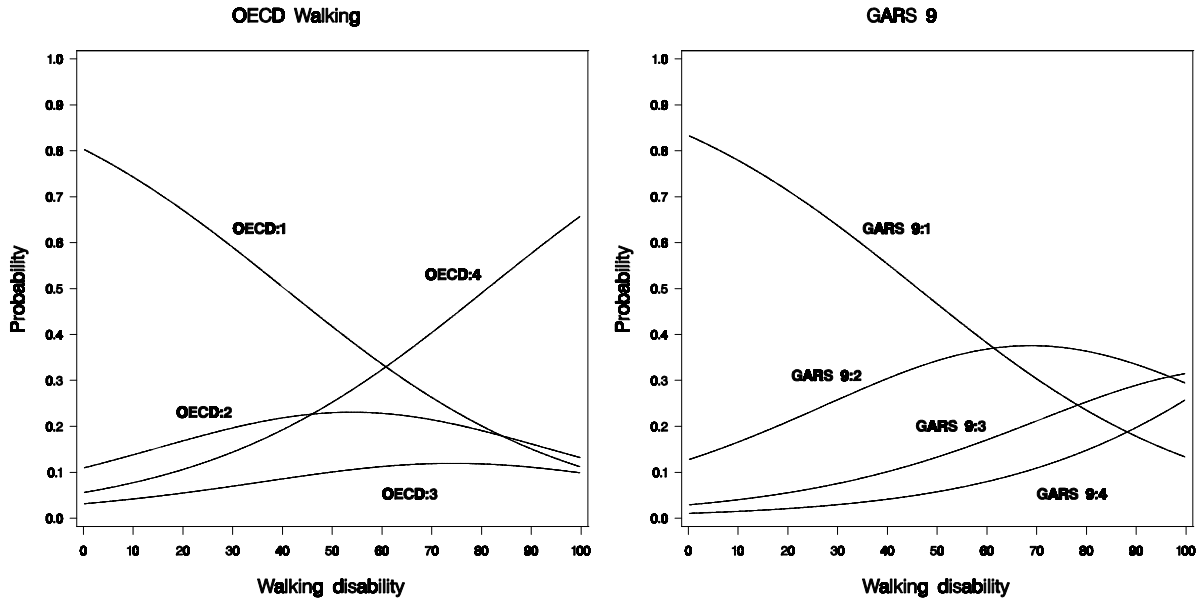
Figure 4.A *Graded response (ICC) curves of OECD Walking and GARS9 items. The lines indicate the probability of response in category 2 and higher, 3 and higher, and 4 as a function of disability.*



A property of the Graded Response Model is that larger category numbers indicate greater severity. As an illustration of this property, figure 4.a plots the GRFs of the items OECD Walking and GARS9. Both items contained four categories, and therefore three response functions were estimated per item. The curve labelled '2+' shows how the probability of getting the answer '2 or higher' is affected by the severity of the disability. So, for zero disability the response probability of categories OECD:2, OECD:3 and OECD:4, taken together, was equal to p=0.20, while the probability for OECD:1 was 0.80. The midpoints of the GRFs were located at DL40, DL67 and DL81 respectively, (notation DLxx is used as an abbreviation of 'Disability Level xx'). The midpoint gives the region of the disability axis where the category pair has optimal discrimination.

Response functions can be used to determine which items (and response categories within items) are most sensitive at a given level of disability. Suppose that our task is to group the four response categories of the OECD in two categories. The optimal way to do this depends on the population to which the item is administered. If we know that the average disability is about 40 then, the most efficient choice is to group together OECD:2, OECD:3 and OECD:4, because the curve varies substantially between say DL20 and DL60. For DL80 it would be wiser to group OECD:1, OECD:2 and OECD:3. However, if the functions '1:2' and '3:4' are not response functions of categories of one polytomous item, but are response functions for two different dichotomous items A and B, then it would be more informative to choose A if the disability is below or about DL40, and B if disability is

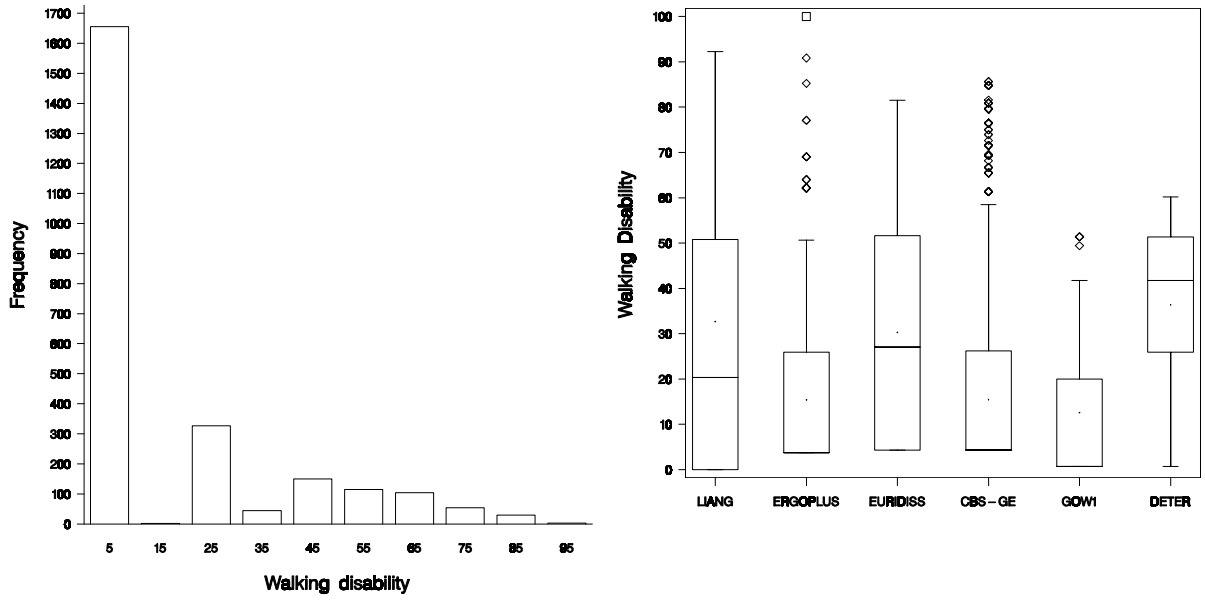
Figure 4.B *Category functions of OECD Walking and GARS 9. The lines indicate the response probability of categories 1, 2, 3 and 4 at different levels of disability.*



very high. Thus, the response function provides a rational basis for choosing the most efficient items for a given sample.

It is straightforward to transform the GRFs of figure 4.a into *category functions*. A category function describes the response probability of each category separately as a function of disability. Probabilities add up to 1 for all levels of disability, and can simply be found by subtraction of GRFs. The behaviour of individual categories could be compared through their category function. For example, figure 4.b shows that for zero disability, the probability of obtaining a response in OECD:1, OECD:2, OECD:3 and OECD:4 was 0.8, 0.11, 0.06, and 0.03 respectively. Likewise, the maximal probability of obtaining a response OECD:2 could be found at DL55. Note that the answer categories OECD:2 and OECD:4 were used more often than OECD:3 for all levels of disability. Thus, according to the model, respondents preferred the answer category OECD:2 and OECD:4 to OECD:3 irrespective of their disability, which is not what a good response system should do (one might try to improve the item by combining OECD:2 and OECD:3 into one single category, or by eliminating OECD:4). By comparison, the characteristics of the GARS9 item were more favourable. The curves are spaced more equally over the disability axis, and no answer category was entirely dominated by others though respondents apparently found it difficult to distinguish between the response categories GARS 9:3 and GARS 9:4.

Figure 4.C Distribution of walking disability. Left: All sources combined (n=2484). Right: per source

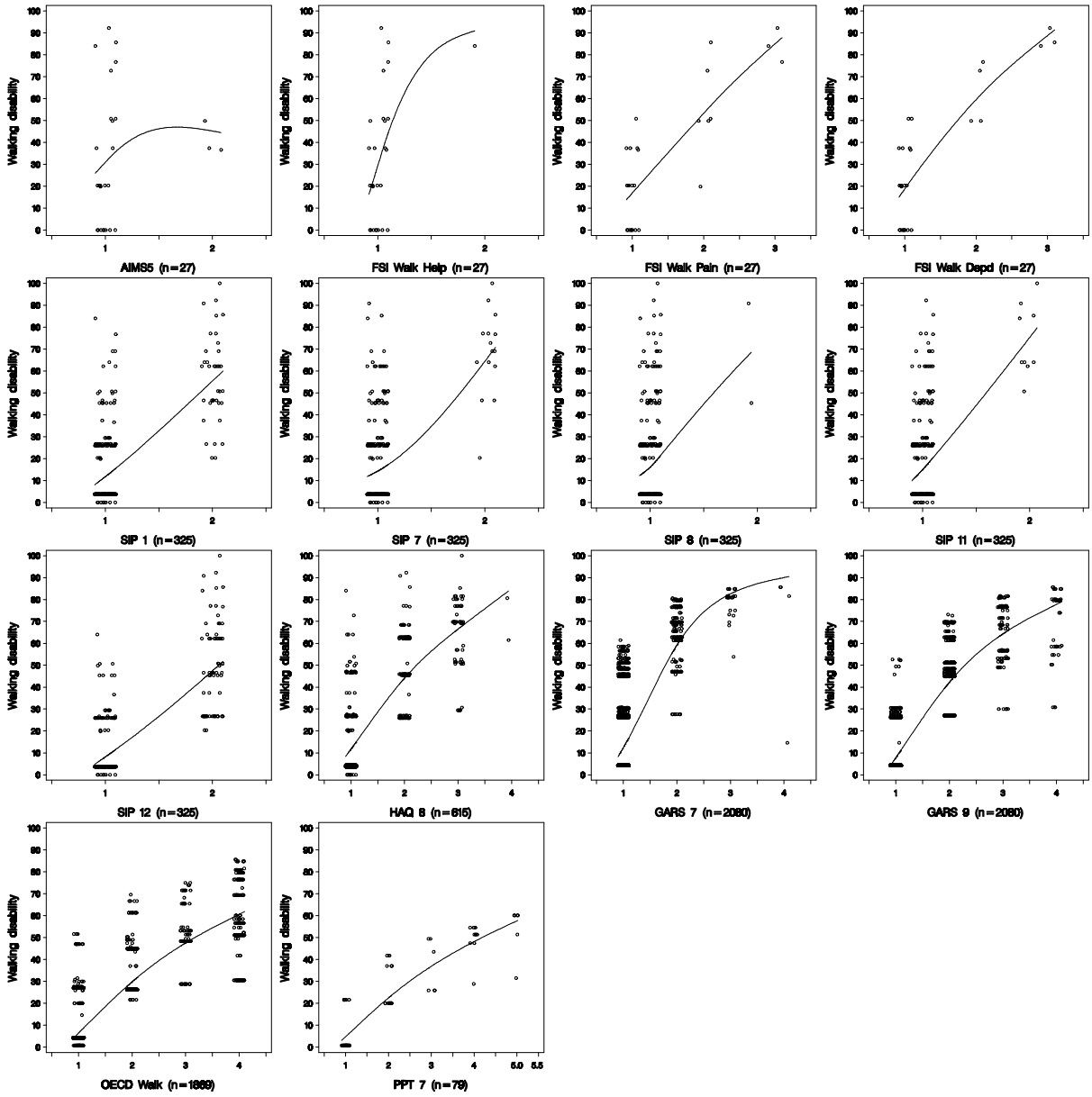


4.2 Distribution of walking disability

figure 4.c displays the distribution of disability with regard to walking, when all sources were combined. The distribution is very skewed. The tall bars on the left side represent the healthy part of the combined sample that reported no disability whatsoever. People that did not report any form of walking disability were located within the interval 0-10. The plot on the right side displays the same data, but now as a box plot per data source. The mean is indicated by the stars, the median is denoted by a line, and the boxes enclose observations between the 25th and 75th percentile. In some cases (ERGOPLUS, CBS-GE, GOW1), the median coincided with the 25th percentile. There were substantial differences in disability between the various sources. The order of mean walking disability among sources from high to low was DETER, LIANG, EURIDISS, CBS-GE, ERGOPLUS and GOW1. This ranking agreed with prior notions about differences in ability between these samples (In fact, one of the authors predicted an almost identical ordering before seeing the actual results).

figure 4.d displays walking disability as a function of all walking items. In order to make the plot more informative, category numbers were jittered by adding uniform random noise between -0.1 and 0.1 to each observation. Also, the relationship between category score and disability was smoothed by the SM70 scatterplot smoother (SAS Institute, 1990). As expected, disability scores generally

Figure 4.D Distribution of walking disability, per response category for each walking item.



increased with the answer category number. Effective items had large differences between categories relative to differences within categories. In this sense, item FSI-A4 was quite good, while item AIMS5 was not. The GARS9 item was preferable to the HAQ8 because there is less variation within the categories. Note also that OECD (a question) and PPT7 (a performance test) had similar smoothing lines, and therefore were sensitive to similar levels of disability.

4.3 Reliability and validity

Appendix IV contains the item-scale and item-item correlations for walking. Most items had substantial correlations with disability, which pointed to adequate reliability of the scale. However, as noted, these correlations were inflated due to the small number of items. The upper triangle of the table contained item-item correlations. The lower triangle gives the number of observations that were used to compute the correlation. Note that many correlations could not be calculated because the variables were never jointly observed. Three regions have been marked, each corresponding to a different set of items. Assuming that comparable values would have been found for the missing correlations, average correlations were computed for each region separately. Cronbach's α is equal to 0.88, so the reliability of the measure is satisfactory.

Multiple imputation was used to construct two completed data matrices. In order to avoid multicollinearity problems, items FSI-A2, SIP8 and SIP11 (which all had very low marginal frequencies) had to be eliminated from the analysis. To aid interpretation, the correlation matrix has been divided into three areas. The middle area contains correlations for which at least some data were available. Both outer areas, indicated by gray cells, gives correlations between items that were never jointly observed. The latter correlations depend entirely on extrapolation from existing data, and are therefore very variable.

In a factor analysis, both the 'eigenvalue-larger-than-1' and the 'elbow' rule suggested a two-factor solution. table 4.b contains the (unrotated) loadings of a one-factor and a two-factor solution, replicated over both imputations. The solution was quite stable within replications. The first factor was also found in the two-factor solution. Items that appeared to fit less well were PPT7, SIP1 and SIP7 a finding that was consistent over both replications. Most variability between replications was found for FSI-D and PPT7. This is not surprising since both items were observed only in small samples and thus are subject to uncertainty. The first factor was somewhat weaker in the second replication.

Factor 1 could be interpreted as a walking disability factor. Interpretation of factor 2 was less straightforward. Items that contained an element of performance time (PPT7, OECD) loaded somewhat higher on factor 2. However, the ordering of items on this factor was similar to the ordering in the linkage diagram. Thus, factor 2 could also be a methodological artifact of the linkage

Table 4.B *Factor loadings for walking. One- and two-dimensional solution. Two replications (imputation 1 and 2).*

	IMPUTATION 1		**IMPUTATION 2**	
	1-Factor solution		**1-Factor solution**	
	FACTOR 1		FACTOR 1	
AIMS5	0.51		AIMS5	0.46
FSI-P	0.77		FSI-P	0.70
FSI-D	0.90		FSI-D	0.57
SIP1	0.32		SIP1	0.30
SIP7	0.30		SIP7	0.26
SIP12	0.45		SIP12	0.55
HAQ8	0.94		HAQ8	0.87
GARS7	0.45		GARS7	0.43
GARS9	0.55		GARS9	0.46
OECD	0.33		OECD	0.45
PPT7	0.04		PPT7	0.38
Eigenvalue	3.59		Eigenvalue	3.04

	2-Factor solution (unrotated)		**2-Factor solution (unrotated)**		
	FACTOR1	FACTOR2		FACTOR1	FACTOR2
AIMS5	0.50	-0.24	AIMS5	0.44	0.06
FSI-P	0.79	-0.43	FSI-P	0.69	-0.31
FSI-D	0.88	-0.17	FSI-D	0.61	-0.58
SIP1	0.31	-0.13	SIP1	0.29	-0.36
SIP7	0.30	-0.29	SIP7	0.26	-0.36
SIP12	0.44	-0.24	SIP12	0.56	-0.40
HAQ8	0.91	-0.07	HAQ8	0.82	-0.02
GARS7	0.48	0.39	GARS7	0.46	0.40
GARS9	0.63	0.56	GARS9	0.52	0.48
OECD	0.42	0.74	OECD	0.56	0.71
PPT7	0.08	0.76	PPT7	0.41	0.37
Eigenvalue	3.70	2.06	Eigenvalue	3.20	1.93

structure. If so, the same factor would be expected to show up in the analysis of a different type of disability, e.g. dressing.

5. RESULTS FOR ICIDH-D 35,36: DRESSING

5.1 Threshold parameters of dressing items

The normalizing transformation for scaling dressing disability between 0 and 100 was equal to Y = 22.8 + 22.6 X, where X represents the MULTILOG output. table 5.a shows that at the most severe disability level, individuals required help from others for dressing, with the (decreasing) severity of disabilities thereafter being reflected by the use of special devices, and much difficulty, some/moderate difficulty and, finally, mild difficulty in performing a task or action. Problems with buttons and zips, or underpants were generally considered more severe disability than trouble with shoes. As before, the physical performance test PPT4 (time taken to put on and take off a coat) measured only mild disability.

5.2 Distribution of dressing disability

figure 5.a displays the distribution of dressing disability. People not reporting any form of dressing disability are now located in interval 20-30, and not near zero. The reason for this was the presence of the PPT4 item (put on and take off a coat within 10 seconds), which appeared to be difficult even for healthy people. In some sense, dressing scores within the interval 0-20 do not presented disability, but rather over-ability. For this reason, scores on the dressing scale cannot be directly compared with those on the walking scale, that is, DL50 on the dressing scale differs in intensity from DL50 on the walking scale. If cross-comparability of scales is desired, then the distribution of both scales must be anchored in some way. The order of mean dressing disability per source was from high to low, EURIDISS, DETER, LIANG, CBS-GE, ERGOPLUS and GOW1. This order was the same as for walking, except that EURIDISS was promoted from rank 3 to 1. Two distributions were so skewed that their 0^{th} and 75^{th} percentiles were identical (ERGOPLUS, CBS-GE).

Table 5.A Estimated threshold parameters of dressing items

Threshold parameter of the Graded Response Model	Normalized threshold parameter	Item:Upper Category	Number of respondents in upper category	Description of the upper category
6.24	164	SIP35:2	1	dress only with someone's help
4.96	135	GAO:4	28	dress and undress: only with help
4.58	126	FSH4:4	1	buttoning clothes: severe difficulty
4.57	126	SIP31:2	5	require help with buttons, zips etc.
4.50	125	FSG2:2	1	underpants: special device
4.50	125	FSJ2:2	1	shoes: special device
4.39	122	HAQ1:4	16	dress myself: cannot
4.11	116	GAO:3	35	dress and undress: much difficulty
3.69	106	FSG4:3	1	underpants: moderate difficulty
3.68	106	FSJ4:4	2	shoes: severe difficulty
3.67	106	PPT4:5	3	coar on and off: > 20 sec.
3.44	101	SIP34:2	14	dress myself very slowly
3.03	91	AIMS2:2	3	cannot button articles
3.01	91	HAQ1:3	37	dress myself: much difficulty
2.67	83	SIP29:2	27	trouble with shoes
2.53	80	FSJ4:3	2	shoes: moderate difficulty
2.21	73	AIMS3:2	5	cannot easily tie shoes
2.14	71	FSH4:3	4	buttoning clothes: moderate diff
2.12	71	GAO:2	277	dress and undress: some difficulty
1.79	63	FSH4:2	1	buttoning clothes: mild difficulty
1.64	60	PPT4:4	13	coat on and off: 15-20 sec.
1.37	54	HAQ1:2	110	dress myself: some difficulty
.83	42	FSG4:2	8	underpants: mild difficulty
.73	39	PPT4:3	12	coat on and off: 10-15 sec.
.29	29	FSJ4:2	8	shoes: mild difficulty
-.82	4	PPT4:2	24	coat on and off: < 10 sec.

Figure 5.A *Distribution of dressing disability. Left: All sources combined (n=2670). Right: per source*

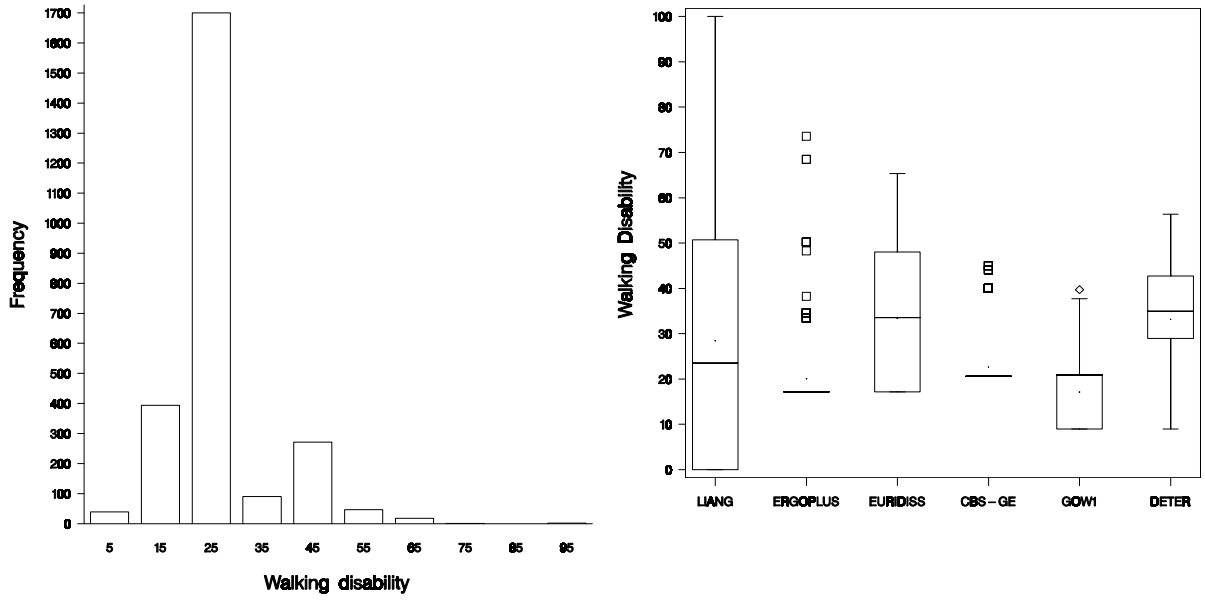
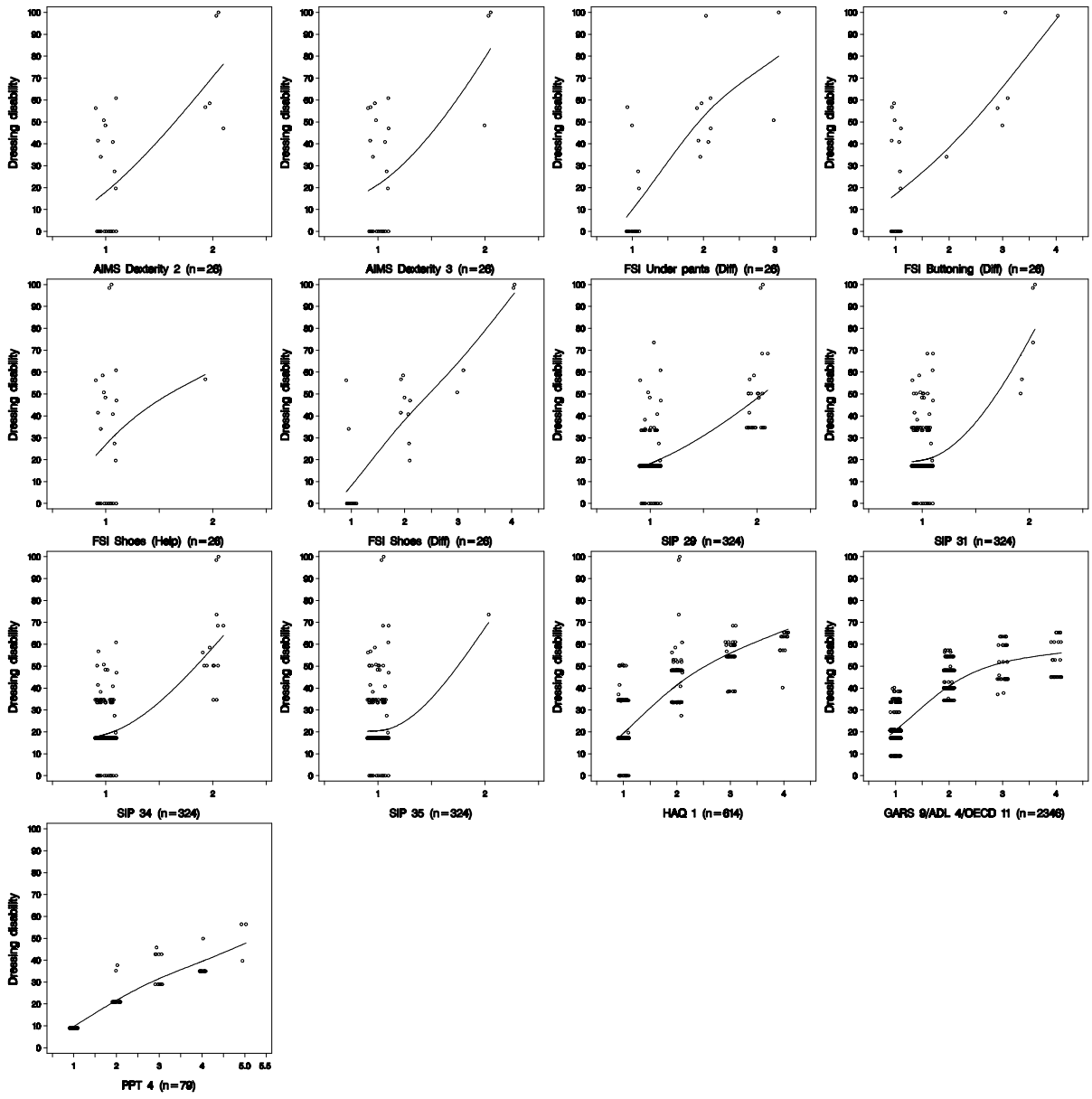


figure 5.b displays dressing disability as a function of all the dressing items. The fact that disability scores increased with the answer category number attested the face validity of the disability scale. The upper end of the scale was largely determined by two individuals from Liang's study who had extreme scores on nearly all items. This effect was also the result of the small sample size and the large number of items of Liang's study.

Figure 5.B *Distribution of dressing disability, per response category for each dressing item.*


Dressing disability
AIMS Dexterity 2 (n=26)
AIMS Dexterity 3 (n=26)
FSI Under pants (Diff) (n=26)
FSI Buttoning (Diff) (n=26)
FSI Shoes (Help) (n=26)
FSI Shoes (Diff) (n=26)
SIP 29 (n=324)
SIP 31 (n=324)
SIP 34 (n=324)
SIP 35 (n=324)
HAQ 1 (n=614)
GARS 9/ADL 4/OECD 11 (n=2346)
PPT 4 (n=79)

5.3 Reliability and validity

Item-scale and item-item correlations for dressing are given in Appendix V. Again, a number of correlations could not be calculated because the variables were never jointly observed. The correlations were all positive, and generally quite high. The average correlation was 0.49, which corresponds to a Cronbach's α of 0.91.

As for walking, multiple imputation was applied to the subject by item matrix to estimate the missing correlations. Items FSG2, FSJ2 and SIP35 were removed because their low marginal frequencies made the imputation algorithm unstable. In the factor analysis the 'eigenvalue-larger-than-1' and the 'elbow' criteria indicated three factors. table 5.b contains the (unrotated) loadings of a one-factor and a two-factor solution, replicated over both imputations. The three-factor solution is not reported because it is difficult to interpret. Considerable differences between imputation occurred for items with small samples (AIM3, FSH4, PPT4). Factor 1 was definitely a dressing disability factor. Factor 2 was not consistent across imputations. No linkage ordering was found, as was the case for walking. Given these results, it seems unlikely that factor 2 was a linkage artifact, but is not clear how factor 2 should be interpreted. Satisfactory explanations were not obtained by looking at various rotations of the loadings, or by investigating the third and higher factors.

Table 5.B *Factor loadings for dressing. One- and two-dimensional solution. Two replications (imputation 1 and 2).*

	IMPUTATION 1		**IMPUTATION 2**	
	1-Factor solution		**1-Factor solution**	
	FACTOR 1		FACTOR 1	
AIM2	0.30	AIM2	0.26	
AIM3	-0.02	AIM3	0.61	
FSG4	-0.19	FSG4	0.06	
FSH4	0.63	FSH4	0.06	
FSJ4	0.50	FSJ4	0.41	
SI29	0.54	SI29	0.70	
SI31	0.66	SI31	0.72	
SI34	0.72	SI34	0.67	
HAQ1	0.84	HAQ1	0.80	
GAO	0.46	GAO	0.27	
PPT4	0.19	PPT4	-0.22	
Eigenvalue	2.97	Eigenvalue	2.85	

	2-Factor solution (unrotated)		**2-Factor solution (unrotated)**		
	FACTOR1	FACTOR2		FACTOR1	FACTOR2
AIM2	0.35	-0.46	AIM2	0.37	0.67
AIM3	-0.09	0.77	AIM3	0.61	-0.48
FSG4	-0.29	0.84	FSG4	0.12	-0.45
FSJ4	0.48	0.19	FSJ4	0.59	0.71
SI29	0.52	0.33	SI29	0.67	-0.19
SI31	0.64	0.10	SI31	0.70	-0.04
SI34	0.72	0.42	SI34	0.65	-0.08
HAQ1	0.81	0.18	HAQ1	0.78	- 0.40
GAO	0.45	0.12	GAO	0.22	-0.42
PPT4	0.25	-0.48	PPT4	-0.29	-0.41
Eigenvalue	3.07	2.25	Eigenvalue	2.98	2,24

6. REVISION OF THE SEVERITY OF DISABILITIES SCALE

This chapter summarizes the results of the analyses in Chapters 4 and 5 and describes their role in our proposal of a new SDS.

6.1 Summary and interpretation of the results

The major task in Chapter 3 was to combine information from different data bases, collected by different researchers, on different samples, using different items on the same topic. Aided by the linkage diagram and the item catalogue, we constructed two linked data sets, one for 'walking' (ICIDH-D 40) and one for 'dressing' (ICIDH-D 35/36). Polytomous Rasch analyses of these data yielded a sensible ordering of categories, both for 'walking' and 'dressing' (table 4.a and table 5.a). Reliability in terms of Cronbach's α was satisfactory. Factor analysis, which was only possible after multiple imputation, suggested two (walking) and three (dressing) factors. The first factor was clearly a disability dimension. Higher factors were difficult to interpret.

The items in Tables 4.1 and 5.1 are ordered in sequence of severity. Using the treshold parameter estimates to classify items into different severity clusters, we found six levels of severity of walking or dressing disabilities. Extended with a category 'complete inability' gives seven severity levels. For walking, the items can be roughly grouped as follows (in increasing order of severity):

0 If walking disability is not present, if a person is able to walk 15 meters in less than 20 seconds;

1 Inside walking with mild pain, walking inside and outdoors with some or mild difficulty, walking more slowly;

2 Much difficulty walking outdoors, moderate difficulty walking inside, often moderate pain is present, only short distances can be walked;

3 Walking with the use of an aid (cane, crutches, artificial limbs, walking frame, etc.);

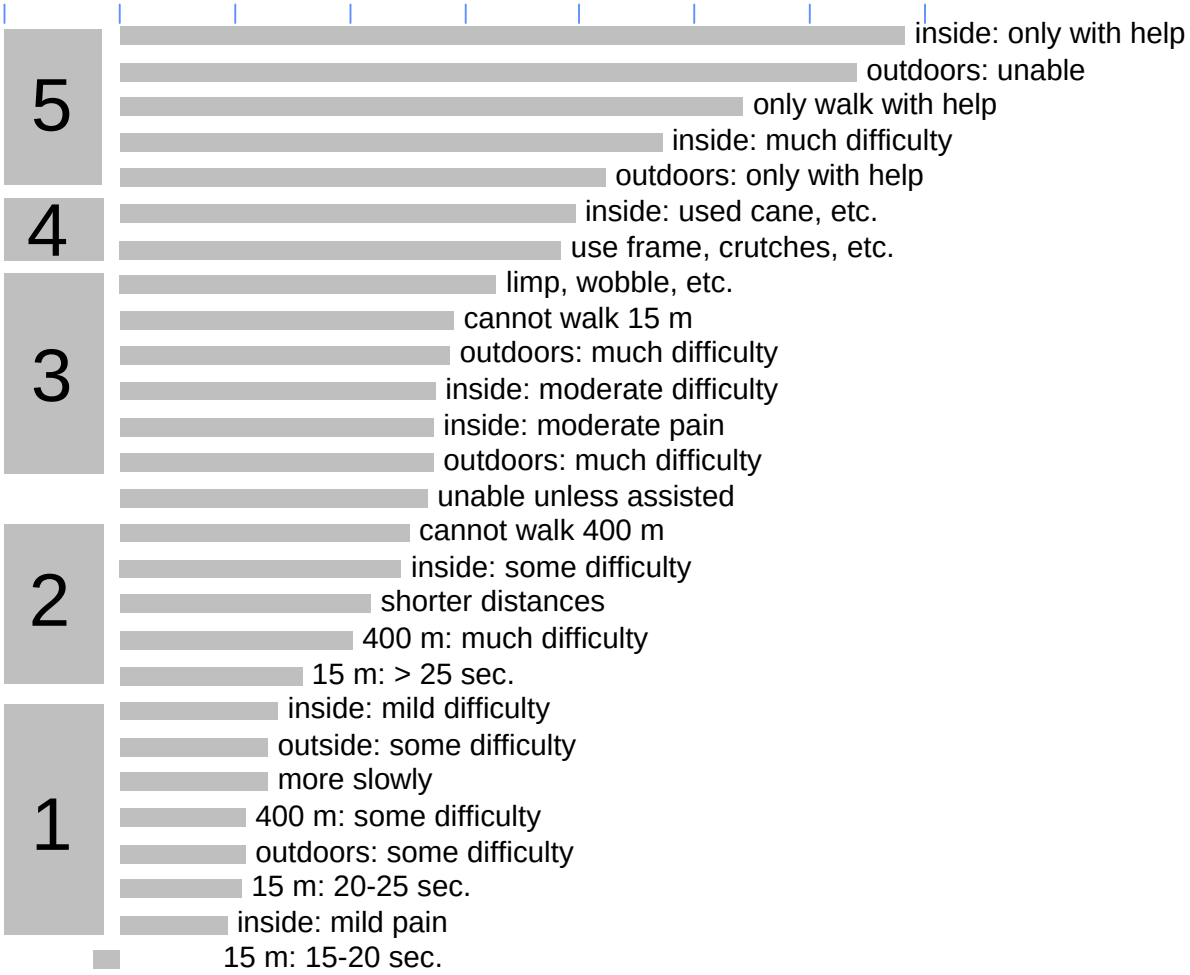
4 Walking outdoors is only possible with the help of someone else, and inside with much difficulty;

5 Walking is only possible with help, unable to walk outdoors.

This subdivision has been represented graphically in figure 6.a.

In the same way the items about dressing disability can be grouped:

Figure 6.A *Treshold parameters of the walking items, divided into five clusters*



0 No problems at all, can put a coat on and take it off in less than 10 seconds;

1 Mild difficulty with underpants and shoes, slower with the coat;

2 Dressing with some difficulty, mild problems with buttoning, again slower with the coat;

3 Buttoning with moderate difficulty and trouble with shoes;

4 Dressing and putting shoes on with much difficulty, very slow and unable to button clothes;

5 Dressing with the aid of special devices, help of someone else or with severe difficulty, dressing only with the help of someone else.

The detail of the categories ‘with difficulty’ for both walking and dressing is greater than that of the existing SDS scale. We found three different clusters of increasing difficulty, whereas the SDS offers only one category ‘difficulty in performance’. On the other hand, the current SDS includes several categories for performance with the help of someone else (‘assisted performance’, ‘dependent performance’ and ‘augmented inability’), whereas we found only one cluster of this kind.

We found for both walking and dressing disability that ‘performing with difficulty’ was, according to the parameters, considered a less severe disability than ‘performing with aids’, which in turn was considered less severe than ‘performing with someone’s help’ (Table 4.1 and 5.1). For walking disability, we found slight indications for a second dimension which included the performance time of a task. The present data set is, however, too limited to study this aspect in sufficient detail.

6.2 Preliminary proposal for a new Severity of Disabilities Scale

As a starting point for a new SDS, we took the current SDS and studied in which aspects this measure differed from other, frequently used, disability instruments that we had studied in phase 1 and 2. In the current SDS, severity of disabilities was classified by categories about ability to perform a certain activity, use of physical aids and need for help from other persons. Whereas in many disability measures difficulty in performance is divided into several categories (e.g. no difficulty, some difficulty, much difficulty, not able), in the current SDS only one level of severity relates to difficulty in performance. The next level of the current SDS concerns use of physical aids, a topic that is not present in many other disability measures. Levels 3 to 5 of the current SDS refer to the need for help from another person. In our analysis it was difficult to assess the level of specification that is needed for this topic because the data files we used did not include populations with severe disabilities. We felt that three levels for ‘need for help from another person’ was too many, especially in comparison with only one level for ‘difficulty’.

Table 6.1 contains our proposal for a new SDS in the domains of Personal Care, Body Disposition, Locomotor and Dexterity.

Table 6.1 Preliminary proposal for the Severity of Disabilities Scale of the ICIDH

1980 code	Proposed code	Label	Description
0	0	Not disabled	individual can perform the activity or sustain the behaviour unaided without any difficulty
1	1	Some or mild difficulty in performance	individual can perform the activity or sustain the behaviour unaided but only with some or mild difficulty
1	2	Moderate difficulty in performance	individual can perform the activity or sustain the behaviour unaided but only with moderate difficulty
1	3	Much or severe difficulty in performance	individual can perform the activity or sustain the behaviour unaided but only with much or severe difficulty
2	4	Aided performance	individual can perform the activity or sustain the behaviour only with a physical aid or appliance
3,4,5	5	Assisted performance	individual can perform the activity or sustain the behaviour only with assistance of another person
6	6	Complete inability	individual cannot perform the activity or sustain the behaviour
8	8	Not applicable	
9	9	Severity unspecified	

The term ‘difficulty’ is an abstraction that subsumes matters such as ‘pain involved’, ‘time taken’, ‘number of errors’, ‘clumsiness’, and so on. It will be clear that, for a given type of disability, each category of the SDS needs an *operational definition* that describes the specific category in terms of a number of observable characteristics. The walking and dressing lists just given are examples of such definitions. Likewise, ‘mild’, ‘moderate’ and ‘severe’ are gradations in difficulty that only get a precise meaning through the operational definition.

A controversial issue is whether severity of disabilities should be defined, measured and interpreted with or without aids and appliances. The analyses done to date indicate that respondents usually considered ‘performance with difficulty’ as a less severe disability than ‘aided *performance’*, which in turn is considered less severe than ‘assisted performance’. This suggests that in practice severity of disabilities is more likely to be interpreted and measured as the severity without aids and appliances. We have therefore preserved the ‘aided performance’ category in our preliminary proposal. The number of items on which this conclusion is based is small however.

In the data of Molleman (see Appendix II and III), we can see that, in a population of severely disabled persons, the old category ‘1’ was overfilled. The counts in categories 4, 5, and 6 of the old

SDS strike us as being relatively high compared with our results. According to the old SDS, category 6 contains people who are totally bed-bound and who are, even with assistance, unable to walk and dress (e.g. patients in a coma). According to the data for ‘homes for the non-working’, this group amounts to 16 percent of the total population, which seems rather high compared with our results.

6.3 Discussion

It should be emphasized that our proposal for a new SDS is based upon analyses of only two forms of disability, ‘walking’ and ‘dressing’, and that our coverage of the ICIDH is far from complete. Furthermore, the spectrum of severity of disabilities of the populations that were included in the data files was perhaps too narrow and excluded very disabled populations in, for instance, nursing homes. Thus it will be necessary to validate the new SDS for other populations and in relation to disabilities other than ‘walking’ and ‘dressing’.

The statistical analyses used in Chapters 4 and 5 are relatively demanding because information has to be combined from very different sources. On some occasions the link between items was very thin, and marginal counts were quite low. This caused problems during parameter estimation, and sometimes resulted in unexpected behaviour of the MULTILOG software. Correlations could not be computed for many combinations of items. To get around this problem, we applied multiple imputation. Because about 80% of the data were missing, imputations were unstable and thus results based on these data (here, factor analysis on the correlation matrix) should be interpreted with caution. This is perhaps an inevitable consequence of the post-hoc nature of using existing data. In a planned prospective study such undesirable factors can be eliminated, for example, by applying an appropriate balanced design. Moreover, the disability distribution should be a more robust estimate of the true population distribution, so that it can serve as a reference distribution. However, the use of existing data bases is perhaps the only way to get a quick overview of this broad and complex field.

One point of criticism of the current SDS (see Hopman-Rock & Miedema, 1995) is that severity is mainly determined by an individual’s dependence on aids or another person (physical independence is a form of a handicap). In the proposed SDS, emphasis is on ‘difficulty’, and ‘dependence’ plays a minor role. A criticism of the current SDS, that it is not possible to reach an improvement of performance with, for instance, a new physical aid or a person’s help, is still valid for the proposed SDS. A solution would perhaps be to look at the performance time of a task (a possible second

dimension of ability). However, it is possible that a task is performed better and faster with one form of aid than another, even though the severity of the disability (aided performance) remains the same. This is also true when a task can only performed very slowly (severe difficulty in performance), but much better (=faster) with the help of another person. A problem with performance time is the absence of a reference point that signals ‘normal’. A topic for phase 4 would be to investigate the possibility of measuring the performance time of a task in a reliable and valid way. One way to meet the criticism that the duration of a disability is not taken into account is to use the outlook scale of the disability section of the ICIDH. A person with a broken leg definitely has a better outlook than someone with a chronic disabling disease has, although the severity of disabilities can be the same now.

A unique aspect of our approach is its empirical basis and its strong emphasis on items that are actually applied in the field. The interpretation and application of the proposed scale might therefore be easier than the current scale. The finer grain on the lower end of the proposed SDS makes it more suitable for applications in public health and prevention. The techniques we use in this report provide keys to conversion issues. It is possible to translate the current SDS into the proposed SDS, to convert the severity as measured by existing disability items into the proposed SDS, or to convert existing items into other (existing or novel) disability items. Such possibilities will preserve much valuable work. Finally, because of the strict mathematical basis of the model, formal tests on aspects of reliability and validity of the scale become available.

Since the present work was a first-time application of the Rasch model to this field, compromises and limitations were inevitable. First, only two types of disabilities (walking and dressing) have been used to generate the proposed SDS. Second, our coverage of instruments that measure disability is far from complete. Third, our sample contains very few severely disabled people. Fourth, because of the sheer incompleteness of the data, only rigid models with strong assumptions could be applied. Fifth, the threshold estimates are very variable because many item categories contain only a few observations. Sixth, we were forced to assume equality between two dressing items in order to get a linkage between them. Seventh, we did not take into account any differences in the mode of data collection (self report, interview, observation).

6.4 Future phase 4

We propose to include the new SDS in a questionnaire, together with several other frequently used disability measures, and then to use this questionnaire on various populations, structured according to an appropriate study design. The sample should at least cover the general population, out-patient and in-patient populations and populations in nursing homes and rehabilitation centres. In this way a more representative data set will be available to validate the proposed SDS and to compare the items of several disability instruments with each other, and with the new SDS. An additional advantage of such a study would be that the results of past and future studies that use the disability instruments included in phase 4 could be linked to the severity of disabilities as measured with the proposed SDS of the ICIDH. A phase 4 study should also pay attention to the existence of a possible second dimension, such as time it takes a person to perform a certain task.

REFERENCES

ANDREWS G, PETERS L, TEESSON M. The Measurement of Consumer Outcome in Mental Health. A report to the National Mental Health Information Strategy Committee. Sydney, 1994.

BRANDSMA JW, HEERKENS YF, VAN RAVENSBERG CD. The international classification of Impairments, Disabilities, and Handicaps (ICIDH): a manual of classification relating to the consequences of disease? Journal of Rehabilitation Sciences, 1995; 8(1): 2-7.

BOWLING A. Measuring health. A review of quality of life measurement scales. Buckingham: Open University Press, 1992.

CBS. Vademecum Gezondheidsstatistiek 1994. Voorburg: CBS, 1995.

GIFI A. Nonlinear multivariate analysis. New York: Wiley, 1990.

HELL van L, HOPMAN-ROCK M. Ontwikkeling en evaluatie van het programma 'Goed Oud Worden': de testfase. Leiden: TNO PG, 1995, publ.nr 95.040.

HOFMAN A, GROBBEE DE, DE JONG PTVM, VAN DEN OUWELAND FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 7:403-22, 1991.

HOPMAN-ROCK M, MIEDEMA HS. The development of a proposal for revision of the Severity of Disabilities Scale of the ICIDH. Personal Care, Body Disposition, Locomotor and Dexterity Disabilities: Phase 1 and Phase 2. Leiden: TNO Prevention and Health, 1995. Publ.nr. 95.043.

HOPMAN-ROCK M, ODDING E, HOFMAN A, KRAAIMAAT FW, BIJLSMA JWJ. Physical and psychological disability in elderly subjects in relation to pain in the hip and/or knee. J Rheumatol 1996; 23: 1037-44.

HOPMAN-ROCK M. Determinanten van immobiliteit en fysieke activiteit. Een pilotstudie onder zelfstandig wonende ouderen die op de wachtlijst voor thuiszorg staan. Leiden: TNO PG 1994. Publ. nr. 94.007.

KIDD D, STEWART G, BALDRY J et al. The Functional Independence Measure: a comparative validity and reliability study. Disability and rehabilitation 1995; 17:10-14.

KEMPEN GIJM, MIEDEMA I, ORMEL J, MOLENAAR W. The assessment of disability with the Groningen Activity Restriction Scale. Conceptual Framework and Psychometric properties. Social Science & Medicine 1996, in press.

KOLEN MJ & BRENNAN RL. Test equating: Methods and practices. New York: Springer, 1995.

KONIJN G, HILDEBRANDT VH. Vragenlijsten over het functioneren van personen met aandoeningen van het bewegingsapparaat in ADL-en arbeidssituaties: een inventariserend literatuuronderzoek. Leiden: NIPG-TNO, 1993. Pub. nr 93.028.

KÖNIG-ZAHN C, FURER JW, TAX B. Het meten van de gezondheid. Deel 1: Algemene Gezondheid. Assen: van Gorcum, 1993.

LIANG MH, LARSON MG, CULLEN KE, SCHWARTZ JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis and Rheumatism 1985, 28:542-7.

LIANG MH, FOSSEL AH, LARSON MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.

MIEDEMA HS, HOPMAN-ROCK M. The development of a proposal for revision of the severity of disabilities scale of the ICIDH. Paper for the ICIDH revision conference on November 1994 in Voorburg, The Netherlands.

MOLENAAR IW, FISCHER GH (Eds.). Rasch models. Berlin: Springer-Verlag, 1995.

ODDING E, VALKENBURG HA, ALGRA D, VANDENOUWELAND FA, GROBBEE DE, HOFMAN A. Association of locomotor complaints and disability in the Rotterdam study. Ann Rheum Dis 1995; 54: 721-5.

RAAD VOOR GEZONDHEIDSRESEARCH. Internationale classificatie van stoornissen, beperkingen en handicaps 1980. Voorburg: Werkgroep Classifacatie en Coderingen, 1981.

ROSKAM EE, JANSEN PGW. Conditions for Rasch-dichotomizability of the unidimensional polytomous Rasch model. Psychometrika 54; 317-332: 1989.

RUBIN, DB. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.

SAMEJIMA F. Estimation of latent ability using a response pattern of graded responses. Psychometrika; 34: Monograph 17, 1969.

SAS Institute Inc. SAS/STAT User's Guide, Version 6, Fourth Ed. Cary, NC, 1990.

SCHAFER JL. Analysis of incomplete multivariate data. London: Chapman & Hall, 1996.

SUURMEIJER TPBM, DOEGLAS DM, MOUM T, et al. The Groningen Activity Restriction Scale for measuring disability: its utility in international comparisons. Am J Publ Health 1994; 84:1270-3.

THISSEN D. MULTILOG User's Guide. Chicago: Scientific Software, Inc., 1991.

VALE CD. Linking item parameters onto a common scale. Applied Psychological Measurement 1986, 10: 333-44.

VAN BUUREN S, BOSHUIZEN HC, KNOOK DL. Multiple imputation of missing blood pressure covariates in survival analysis. Manuscript, TNO Prevention and Health, 1996.

WHO. International Classification of Impairments, Disabilities and Handicaps. Geneva: World Health Organization, 1980. Reprint 1993.

WHO & NCBS, Third consultation to develop common methods and instruments for Health interview surveys. Report on a WHO meeting, Voorburg, the Netherlands, 22-24 September 1992, Voorburg: Netherlands Central Bureau of Statistics, 1993.

APPENDICES

Page

APPENDIX I

Letter to Dr. Liang

Dr. M.H. Liang
Department of Med. Rheumatol
Grigham and Women's Hospital
75 Francisstreet
Boston MA 02115
USA

MH/KH/4009/6300

Revision ICIDH

Dear Dr. Liang,

Last year, my colleague Dr. Harald Miedema and I were involved in the revision process of the International Classification of Impairments, Disabilities and Handicaps (ICIDH). The WHO Collaborating Centre in the Netherlands asked us to start a project with the aim to develop a proposal for the revision of the 'Severity of Disability Scale (SDS)' of the ICIDH.

This project is divided into four phases. Enclosed you will find the report about the first two phases which produced a list of 20 widely used disability measures.

In the third phase of the project we aim to relate some measures of this list to each other, to trace the most important dimensions of severity of disabilities. Therefore we need existing databases including two or more of these disability measures.

From your article in Arthritis and Rheumatism (Vol.28, May, 1995) we know that you were involved in research with 50 arthritis patients using the FSI, HAQ, AIMS and SIP. All these disability measures are on our list. We are very interested in the possibility to use this dataset for secondary analyses in our phase 3 revision project. We kindly request you to place this dataset at our disposal (including age, sex and answers on FSI, HAQ, AIMS and SIP on itemlevel). Please can you let us know if this is possible?

Yours sincerely,

Drs. Marijke Hopman-Rock
Project manager

APPENDIX II

Catalogues of walking items

WALKING ITEMS: ICIDH-D 40

AIMS Am 5 **LIANG** **AI10**

Are you unable to walk unless you are assisted by another person or by a cane, crutches, artificial limbs, or braces ?

no = 1	34
yes = 2	3
? = 9	1

FSI A1 HELP **LIANG** **FSA2**

Walking inside

0 = no help use	34
1 = used a cane, special equipment or other device	1
2 = used someone's else's help	0
3 = used devices and someone else's help	0
4 = unable to do the activity	0
9 = ?	2

FSI A1 PAIN **LIANG** **FSA3**

Walking inside

0 = no pain	20
1 = mild pain	7
2 = moderate pain	5
3 = severe pain	0
4 = extreme pain	0
9 = ?	5

FSI A1 DIFFICULTY **LIANG** **FSA4**

Walking inside

0 = none	21
1 = mild difficulty	4
2 = moderate difficulty	4
3 = severe difficulty	0
4 = extreme difficulty	0
9 = ?	8

SIP Am Amb 6 **LIANG** **SI74**

I do not walk at all

no = 1	38
yes = ?	0

SIP Ned 81 **ERGOPLUS** **LOOP06**

Ik loop helemaal niet

nee = 0	304
ja = ? (komt niet voor)	0

SIP Am. Amb 8 **LIANG** **SI76**

I only walk with help from someone else

no = 1	38
yes = 2	0

SIP Ned 83 **ERGOPLUS** **LOOP04**

Ik loop alleen maar als iemand mij erbij helpt

nee = 0	302
ja = 88	2

SIP Am Amb 11 **LIANG** **SI79**

I get about only by using a walking frame, crutches, stick, walls, or hold on to furniture

no = 1	36
yes = 2	2

SIP Ned 84 **ERGOPLUS** **LOOP05**

Ik verplaats me alleen maar m.b.v. een drie- of vierpoot, krukken, stok, de wanden, of door me aan meubels vast te houden

nee = 0	296
ja = 79	8

SIP Am Amb 1 **LIANG** **SI69**

I walk shorter distances or often stop for a rest

no = 1	25

yes = 2 13

SIP Ned 85 **ERGOPLUS** **LOOP06**

Ik loop kleinere afstanden of sta vaak stil om te rusten

nee = 0 276
ja = 48 28

SIP Am Amb 7 **LIANG** **SI75**

I walk by myself but with some difficulty; for example I limp, wobble, stumble, or I have a stiff leg.

no = 1 30
yes = 2 8

SIP Ned 89 **ERGOPLUS** **LOOP10**

Ik loop weliswaar zonder hulp maar wel met enige moeite, bijv. ik loop kreupel, ik waggel, ik strompel, ik heb een stijf been

nee = 0 294
ja = 55 10

SIP Am Amb 12 **LIANG** **SI80**

I walk more slowly

no = 1 20
yes = 2 18

SIP Ned 91 **ERGOPLUS** **LOOP12**

Ik loop langzamer

nee = 0 244
ja = 35 60

HAQ 8 **LIANG** **HQ13**

Are you able to walk outdoors on flat ground ?

0 = without any difficulty 30
1 = with some difficulty 7
2 = with much difficulty 0
3 = unable to do 0
9 = ? 1

HAQ 8 **ERGOPLUS** **AI1-10**

Are you able to walk outdoors on flat ground ?

0 = without any difficulty 242
1 = with some difficulty 43
2 = with much difficulty 15
3 = unable to do 0

HAQ 8 **GH-ITEMS** **WAL**

Are you able to walk outdoors on flat ground ?

0 = without any difficulty 178
1 = with some difficulty 68
2 = with much difficulty 42
3 = unable to do so 2
. = ? 2

GARS Ned. 7 **GH-ITEMS** **MOVG**

Kunt u geheel zelfstandig rondlopen in huis (eventueel) met stok) ?

Can you, fully independently, get around in the house (if necessary, with a cane)?

1 = Ja, dat kan ik geh. zelfst, zonder enige moeite 207
2 = Ja, dat kan ik geh. zelfst., maar met enige moeite 78
3 = Ja, dat kan ik geh. zelfst., maar met veel moeite 7
4 = Nee, dat kan ik niet zelfst., maar met hulp v.a. 0

ADL 5 **GE94-OUD** **bvr41e**

Zich verplaatsen naar een andere kamer op dezelfde verdieping

1 = zonder moeite 1825
2 = met enige moeite 120
3 = met grote moeite 21
4 = alleen met hulp 5
. = ? 142

GARS Ned. 9 **GH-ITEMS** **OUT**

Kunt u geheel zelfstandig buitenshuis rondlopen (eventueel met stok)? Can you, fully independently, walk outdoors (if necessary, with a cane)?

1 = Ja, dat kan ik geh. zelfst, zonder enige moeite	145
2 = Ja, dat kan ik geh. zelfst., maar met enige moeite	110
3 = Ja, dat kan ik geh. zelfst., maar met veel moeite	29
4 = Nee, dat kan ik niet zelfst., maar met hulp v.a.	8

ADL 8 **GE94-OUD** **bvr41h**

Zich verplaatsen buitenshuis

1 = zonder moeite	1606
2 = met enige moeite	229
3 = met grote moeite	86
4 = alleen met hulp	52
. = ?	140

OECD lopen **GE94_OUD** **OECD12**

Kunt u 400 meter aan een stuk lopen zonder stil te staan? (zo nodig met stok)

0 = niet ingevuld	179
1 = ja, zonder moeite	1480
2 = ja, met enige moeite	204
3 = ja, met grote moeite	70
4 = neen, dat kan ik niet	169
9 = ?	11

OECD lopen **GOW1** **LOPEN**

Kunt u 400 meter aan een stuk lopen zonder stil te staan.

1 = ja, zonder moeite	40
2 = ja, maar met enige moeite	6
3 = ja, maar met grote moeite	2
4 = nee, dat kan ik niet	2

OECD lopen **DETER** **LOPEN**

Kunt u 400 meter aan een stuk lopen zonder stil te staan.

1 = ja, zonder moeite	11
2 = ja, maar met enige moeite	7
3 = ja, maar met grote moeite	2
4 = nee, dat kan ik niet	10

PPT 7 Nederlands **GOW1** **METER**

(15 meter lopen)

0 = kan niet	1
1 = > 25 sec.	2
2 = 20.5-25 sec.	1
3 = 10.5-15 sec.	14
4 = <= 15 sec.	32

PPT 7 Nederlands **GOW1** **METER2**

(15 meter lopen, 2e meting)

0 = kan niet	0
1 = > 25 sec.	1
2 = 20.5-25 sec.	4
3 = 10.5-15 sec.	15
4 = <= 15 sec.	25
. = ?	5

PPT 7 Nederlands **DETER** **METER**

(15 meter lopen)

0 = kan niet	5
1 = > 25 sec.	8
2 = 20.5-25 sec.	5
3 = 10.5-15 sec.	6
4 = <= 15 sec.	5
. = ?	1

FIM 8 **DCHOP** **B8**

50 meter gaan op vlak terrein

0 = geen beperking	1646
1 = moeite met de uitvoering	292
2 = uitvoering met hulpmiddelen	35
3 = uitvoering met hulp	83
4 = afhankelijke uitvoering	35
5 = ernstig onvermogen	52
6 = volledig onvermogen	133
8 = niet van toepassing	0
9 = ernst niet nader omschreven	0
. = ?	26

FIM 8 **TNWHOP** **B8**

50 meter gaan op vlak terrein

0 = geen beperking	3652
1 = moeite met de uitvoering	581
2 = uitvoering met hulpmiddelen	107
3 = uitvoering met hulp	187
4 = afhankelijke uitvoering	111
5 = ernstig onvermogen	139
6 = volledig onvermogen	765
8 = niet van toepassing	0
9 = ernst niet nader omschreven	0
. = ?	101

FIM 8	**TWHOP**	**B8**
	50 meter gaan op vlak terrein	
	0 = geen beperking	924
	1 = moeite met de uitvoering	52
	2 = uitvoering met hulpmiddelen	9
	3 = uitvoering met hulp	2
	4 = afhankelijke uitvoering	3
	5 = ernstig onvermogen	0
	6 = volledig onvermogen	7
	8 = niet van toepassing	0
	9 = ernst niet nader omschreven	0
	. = ?	4

APPENDIX III

Catalogue of dressing items

DRESSING ITEMS: ICIDH-D 35/36

AIMS Activities 2 **LIANG** **AI33**

How much help do you need in getting dressed?

1 = no help at all	38
2 = only need help tying shoes	0
3 = need help getting dressed	0

AIMS Dex 3 **LIANG** **AI17**

Can you easily tie a pair of shoes?

1 = yes	33
2= no	5

AIMS Dex 2 **LIANG** **AI18**

Can you button articles of clothing ?

1 = yes	29
2 = no	8
9 = ?	1

FSI 7 HELP **LIANG** **FSG2**

Putting on underpants

0 = no help used	34
1 = used a cane, special equipment or other device	1
2 = used someone's else's help	0
3 = used devices and someone else's help	0
4 = unable to do the activity	0
9 = ?	2
. = ?	1

FSI 7 PAIN **LIANG** **FSG3**

Putting on underpants

0 = no pain	25
1 = mild pain	6
2 = moderate pain	1
3 = severe pain	0
4 = extreme pain	0
9 = ?	5
. = ?	1

FSI 7 DIFFICULTY **LIANG** **FSG4**

Putting on underpants

0 = no difficulty	17
1 = mild difficulty	9
2 = moderate difficulty	3
3 = severe difficulty	1
4 = extreme difficulty	0
9 = ?	7
. = ?	1

FSI 8 HELP **LIANG** **FSH2**

Buttoning clothes

0 = no help used	33
1 = used a cane, special equipment or other device	0
2 = used someone's else's help	0
3 = used devices and someone else's help	0
4 = unable to do the activity	0
9 = ?	4
. = ?	1

FSI 8 PAIN **LIANG** **FSH3**

Buttoning clothes

0 = no pain	26
1 = mild pain	1
2 = moderate pain	3
3 = severe pain	1
4 = extreme pain	0
9 = ?	6
. = ?	1

FSI 8 DIFFICULTY **LIANG** **FSH4**

Buttoning clothes

0 = no difficulty	21

1 = mild difficulty	1
2 = moderate difficulty	4
3 = severe difficulty	2
4 = extreme difficulty	0
9 = ?	8
. = ?	1

FSI 10 HELP **LIANG** **FSJ2**

Putting on shoes or slippers

0 = no help used	34
1 = used a cane, special equipment or other device	1
2 = used someone's else's help	0
3 = used devices and someone else's help	0
4 = unable to do the activity	0
9 = ?	2
. = ?	1

FSI 10 PAIN **LIANG** **FSJ3**

Putting on shoes or slippers

0 = no pain	24
1 = mild pain	4
2 = moderate pain	1
3 = severe pain	1
4 = extreme pain	0
9 = ?	7
. = ?	1

FSI 10 DIFFICULTY **LIANG** **FSJ4**

Putting on shoes or slippers

0 = no difficulty	15
1 = mild difficulty	9
2 = moderate difficulty	2
3 = severe difficulty	3
4 = extreme difficulty	0
9 = ?	8
. = ?	1

SIP Am BCM 31 **LIANG** **SI22**

I have trouble putting on my shoes, socks, or stockings (tights)

1 = no	30
2 = yes	8

SIP Ned 36 **ERGOPLUS** **LICHVR19**

Ik heb moeite met het aantrekken van schoenen, sokken en kousen

0 = nee	282
57 = ja	22
. = ?	2

SIP Am BCM 31 **LIANG** **SI24**

I do not fasten my clothing; for example, I require help with buttons, zips or shoe laces

1 = no	34
2 = yes	4

SIP Ned 36 **ERGOPLUS** **LICHVR20**

Ik maak mijn kleren niet vast, ik heb iemands hulp nodig, bijv. met knopen, ritssluiting en schoenveters

0 = nee	301
74 = ja	3
. = ?	2

SIP Am BCM 34 **LIANG** **SI27**

I dress myself but do so very slowly

1 = no	32
2 = yes	6

SIP Ned 38 **ERGOPLUS** **LICHVR22**

Ik kleed mezelf wel aan maar het gaat erg langzaam.

0 = nee	294
43 = ja	10
. = ?	2

SIP Am BCM 35 **LIANG** **SI28**

I only get dressed with someone's help

1 = no	38
2 = yes	0

SIP Ned 37 **ERGOPLUS** **LICHVR21**

Ik krijg mijn kleren alleen maar aan als iemand mij helpt

0 = nee	303
88 = ja	1
. = ?	2

HAQ 1 **LIANG** **HQ6**

Are you able to dress yourself, including tying shoelaces and doing buttons?

0 = without any difficulty	25
1= with some difficulty	9
2 = with much difficulty	3
3 = unable to do	0
9 = ?	1

HAQ 1 **ERGOPLUS** **AI1_04**

Are you able to dress yourself including handling of clusures (buttons, zippers, snaps)?

0 = without difficulty	283
1 = with some difficulty	14
2 = with much difficulty	3
3 = unable to do	0
. = ?	6

HAQ 1 **GH-ITEMS** **DREHA**

.. u zelf aan te kleden

0 = without any difficulty	154
1 = with some difficulty	87
2 = with much difficulty	33
3 = unable to do	16
. = ?	2

GARS Ned. 9 **GH-ITEMS** **DRE**

Kunt u zich geheel zelfstandig aan- en uitkleden?

1 = Ja, dat kan ik geh. zelfst., zonder enige moeite	160
2 = Ja, dat kan ik geh. zelfst., maar met enige moeite	102
3 = Ja, dat kan ik geh. zelfst., maar met veel moeite	17
4 = Nee, dat kan ik niet zelfstandig, maar met hulp v.a	13

ADL 4 **GE94-OUD** **BVR41D**

Aan- en uitkleden

1 = zonder moeite	1778
2 = met enige moeite	168
3 = met grote moeite	16
4 = alleen met hulp	15

OECD aankleden **GOW1** **KLEDEN**

Kunt u zichzelf aan- en uitkleden?

1 = ja, zonder moeite	48
2 = ja, maar met enige moeite	1
3 = ja, maar met grote moeite	1
4 = nee, dat kan ik niet	0

OECD aankleden **DETER** **KLEDEN**

Kunt u zichzelf aan- en uitkleden?

1 = ja, zonder moeite	22
2 = ja, maar met enige moeite	7
3 = ja, maar met grote moeite	1
4 = nee, dat kan ik niet	0

PPT 4 Nederlands **GOW1** **JAS**

Jas aan- en uittrekken

0 = kan niet	1
1 => 20 sec	1
2 = 15.5-20 sec	4
3 = 10.5-15 sec	20
4 = <=10 sec	24

PPT 4 Nederlands **DETER** **JAS**

Jas aan- en uittrekken

0 = kan niet	2
1 => 20 sec	12
2 = 15.5-20 sec	8
3 = 10.5-15 sec	4
4 = <=10 sec	3

FIM 4 **DCHOP** **B5**

Bovenlichaam aan- en uitkleden

0 = geen beperking	1189
1 = moeite met de uitvoering	377
2 = uitvoering met hulpmiddelen	19
3 = uitvoering met hulp	339

4 = afhankelijke uitvoering	164
5 = ernstig onvermogen	53
6 = volledig onvermogen	139
. = ?	22

FIM 4 **TNWHOP** **B5**

Bovenlichaam aan- en uitkleden

0 = geen beperking	2371
1 = moeite met de uitvoering	739
2 = uitvoering met hulpmiddelen	43
3 = uitvoering met hulp	831
4 = afhankelijke uitvoering	543
5 = ernstig onvermogen	142
6 = volledig onvermogen	915
. = ?	59

FIM 4 **TWHOP** **B5**

Bovenlichaam aan- en uitkleden

0 = geen beperking	931
1 = moeite met de uitvoering	50
2 = uitvoering met hulpmiddelen	3
3 = uitvoering met hulp	7
4 = afhankelijke uitvoering	4
5 = ernstig onvermogen	2
6 = volledig onvermogen	1
. = ?	3

FIM 4b **DCHOP** **B6**

Onderlichaam aan- en uitkleden

0 = geen beperking	1190
1 = moeite met de uitvoering	376
2 = uitvoering met hulpmiddelen	17
3 = uitvoering met hulp	331
4 = afhankelijke uitvoering	158
5 = ernstig onvermogen	57
6 = volledig onvermogen	147
. = ?	26

FIM 4 **TNWHOP** **B6**

Onderlichaam aan- en uitkleden

0 = geen beperking	2350
1 = moeite met de uitvoering	719
2 = uitvoering met hulpmiddelen	45
3 = uitvoering met hulp	825
4 = afhankelijke uitvoering	537
5 = ernstig onvermogen	150
6 = volledig onvermogen	958
. = ?	59

FIM 4 **TWHOP** **B6**

Onderlichaam aan- en uitkleden

0 = geen beperking	918
1 = moeite met de uitvoering	61
2 = uitvoering met hulpmiddelen	7
3 = uitvoering met hulp	5
4 = afhankelijke uitvoering	3
5 = ernstig onvermogen	3
6 = volledig onvermogen	1

APPENDIX IV

Item-total, item-item correlations and Cronbach's α (walking)

Item-item correlations (upper triangle =correlation, lower triangle = n), and item-total correlations.

CORR/N	AIMS 5	F_A 2	F_A 3	F_A 4	SIP 1	SIP 7	SIP 8	SIP 11	SIP 12	HAQ 8	GARS7	GARS9	OECD	PPT7	*TOTAL*
AIMS5	•	-.07	-.05	-.02	-.25	-.16		-.07	-.11	.13					*.10*
FSI_A2	27	•	.45	.47	-.14	-.09		1.00	.20	-.09					*.34*
FSI_A3	27	27	•	.85	.11	.53		.45	.46	.39					*.81*
FSI_A4	27	27	27	•	.19	.59		.47	.46	.44					*.83*
SIP1	27	27	27	27	•	.24	.10	.23	.44	.32					*.67*
SIP7	27	27	27	27	325	•	-.02	.23	.35	.24					*.53*
SIP8	27	27	27	27	325	325	•	.23	.05	.11					*.19*
SIP11	27	27	27	27	325	325	325	•	.23	.21					*.45*
SIP12	27	27	27	27	325	325	325	325	•	.28					*.77*
HAQ8	27	27	27	27	325	325	325	325	325	•	.52	.61			*.77*
GARS7	0	0	0	0	0	0	0	0	0	290	•	.59	.46		*.70*
GARS9	0	0	0	0	0	0	0	0	0	290	2080	•	.68		*.87*
OECD	0	0	0	0	0	0	0	0	0	0	1790	1790	•	.65	*.87*
PPT7	0	0	0	0	0	0	0	0	0	0	0	0	79	•	*.89*
TOTAL	*27*	*27*	*27*	*27*	*325*	*325*	*325*	*325*	*325*	*615*	*2080*	*2080*	*1869*	*79*	•

region	mean	m	alpha
I	.281	14	.85
II	.321	10	.83
III	.585	5	.88

Item-item correlations after multiple imputation (upper triangle = first imputation, lower triangle = second imputation), n = 2673. Gray cells contain correlations of variables that were never jointly observed.

OORR	AIMS 5	F_A 2	F_A 3	F_A 4	SIP 1	SIP 7	SIP 8	SIP 11	SIP 12	HAQ 8	GARS7	GARS9	OECD	PPT 7
AIMS5	•		.45	.42	.10	.02			.20	.66	.27	.32	-.20	-.23
FSI_A2		•												
FSI_A3	.47		•	.92	.20	.31			.43	.79	.06	.23	.13	-.31
FSI_A4	.21		.64	•	.18	.29			.43	.79	.35	.38	.40	-.12
SIP1	.10		.19	.33	•	.28			.47	.34	.27	.07	-.10	.04
SIP7	-.21		.28	.39	.29	•			.40	.24	.08	.12	-.07	-.33
SIP8							•							
SIP11								•						
SIP12	.06		.53	.59	.47	.38			•	.37	.08	.13	-.04	.00
HAQ8	.61		.60	.51	.29	.23			.33	•	.41	.52	.25	.13
GARS7	.28		.19	.12	-.06	.01			-.02	.39	•	.59	.49	.30
GARS9	-.05		.10	.05	.08	.04			.16	.48	.58	•	.70	.47
OECD	.30		.25	-.14	-.04	-.09			.06	.35	.47	.69	•	.60
PPT7	.18		.04	.05	-.18	.01			.35	.29	.29	.38	.59	•

APPENDIX V

Item-total, item-item correlations and Cronbach's α (dressing)

DRESSING	n	Item-total
AIMS2	26	.63
AIMS3	26	.69
FSI-G2	26	.18
FSI-G3	26	.78
FSI-G4	26	.78
FSI-H3	26	.73
FSI-H4	26	.73
FSI-J2	26	.18
FSI-J3	26	.88
FSI-J4	26	.88
SIP29	324	.67
SIP31	324	.56
SIP34	324	.69
SIP35	324	.24
HAQ1	614	.83
GAO	2346	.88
PPT	79	.92

CORR	AIM2	AIM3	FSG4	FSH4	FSJ4	SIP29	SIP31	SIP34	HAQ1	GAO	PPT4
AIM2	•	.43	.31	.76	.65	.44	.62	.51	.37		
AIM3	26	•	.41	.30	.59	.75	.74	.60	.67		
FSG4	26	26	•	.44	.71	.41	.31	.53	.31		
FSH4	26	26	26	•	.60	.30	.49	.62	.45		
FSJ4	26	26	26	26	•	.59	.65	.50	.55		
SIP29	26	26	26	26	324	•	.32	.49	.25		
SIP31	26	26	26	26	324	324	•	.34	.34		
SIP34	26	26	26	26	324	324	324	•	.35		
HAQ1	26	26	26	26	324	324	324	324	•	.68	
GAO	0	0	0	0	0	0	0	0	290	•	.27
PPT4	0	0	0	0	0	0	0	0	0	79	•

The item-item average correlation is 0.49, which corresponds to a Cronbach's α of 0.91.