TNO Prevention and Health

Division Public Health
Wassenaarseweg 56
P.O. Box 2215
2301 CE Leiden
The Netherlands

www.tno.nl

T +31 71 518 18 18
F +31 71 518 19 20
Info-VGZ@pg.tno.nl

TNO report

2001.097

Response conversion: A new technology for comparing existing health information

Date	June 2001
Authors	S van Buuren (TNO Prevention and Health) S Eyres (University of Leeds) A Tennant (University of Leeds) M Hopman-Rock (TNO Prevention and Health)

Financial support	EC DG SANCO/G3 - Health monitoring program
Grant	SI2.131854 (99CVF3-510)
Status	Final report
E-mail	S.vanBuuren@pg.tno.nl
Project number	011.40903
Number of pages	68
ISBN	90-6743-813-8

Summary

The present project is executed within the framework of the Health Monitoring Program (HMP) of the European Commission (EC). The goal of the HMP is to provide relevant and timely information about the health in each Member State. The present project focuses on methodological issues related to comparability of information. To avoid unnecessary duplication, the new health monitoring system will have to be fed by existing data. These data are collected by the individual member states, usually by the statistical office or by a public health institute. Incomparability of information is a major problem in this context. Each Member State has its own tradition in collecting and processing data, and changing established ways of working is not so easy.

This report contains the results of a pilot project. The goal of the project is to develop and demonstrate a new technology, called *response conversion*. More specifically, the project set outs to
- to demonstrate the response conversion methodology on a practical problem,
- to identify key problems, if any.

The method will be illustrated by applying it to two disability areas, walking and dressing disability, but the potential field of application is much broader.

The method consists of two steps. The first step involves the construction of a so-called *conversion key*. This is a relatively complex activity, but needs to be done only once. In the second step, one uses the conversion key to convert prevalence information from individual Member States into a common scale. This step is simple, and can be repeatedly done on a routine basis as new information arrives. The present report includes both steps.

The primary reason why the technique works is that it systematically exploits any overlap in existing information through a well-established statistical model. A *linkage map* is a systematic way of arranging overlapping information, and forms the basis for the statistical analyses. The statistical model relies on *item response theory*, which embraces sophisticated techniques (like Rasch analysis) that have been developed within educational research.

The technique only works if enough overlapping information in the existing information can be found. Therefore, the major danger in practical application of the technique is that linkage may not be possible. For walking and dressing disability, this situation did not arise, and a conversion key could be made. The properties of the statistical model are well known, but application of it in a new environment brings some fresh methodological problems. Important topics for further development are, e.g. how to measure the quality of the conversion key, how to properly account for the uncertainty and translation errors, how to assess the fit of the model, and so on.

The most important asset of the methodology is that it allows the expression of *existing* information onto a common scale. The values on the common scale can subsequently be used to compare and monitor health indicators of different countries. The method thus allows setting up a health monitoring system without the need to drastically change established ways of working.

Contents

Appendices

1 Introduction

1.1 Background

The Health Monitoring Program (HMP) was initiated in 1998 by the European Commission. The goal of the HMP is the "development and exchange of adequate, reliable and comparable indicators of public health, and the structures needed to exchange the relevant data" (EC, 1998). The HMP contributes to the "establishment of a Community health monitoring system that makes it possible to

- Measure health status, trends and determinants throughout the Community;
- Facilitate the planning, monitoring and evaluation of Community programmes and action;
- Provide Member States (MS) with appropriate health information to make comparisons and support their national health policies" (EC, 1998).

The EC signals that various organisations have contributed to the development based on their own specific policies, but that these initiatives have not always been co-ordinated in any major way. This has resulted in consequences like:

- Member States are reporting data to a number of bodies which implies multiple reporting;
- Unnecessary duplication of effort;
- Data and information are often of limited comparability between countries and sometimes of medium or poor quality;
- There are significant gaps in the data available on a number of important diseases (EC, 1998).

It is thus important to bring together the effort of the many different actors in European health monitoring in order to improve its quality and value. Also, it is clear that any future efforts in the field of European health monitoring must be based on the data and the expertise that are already available, in particular at national level but also at international level.

1.2 Dealing with incomparability

1.2.1 Some types of comparability problems

The prospective health monitoring system will bring together data collected in different Member States. It will be clear that any differences in data collection methodologies should be accounted for before these data can be used to provide comparative information across Member States. Incomparability may occur at different levels:

- Appropriate data may not be collected at all in some MS;
- Some MS collect appropriate data for specific subsamples, or with special designs;
- The definition of diseases may differ between MS, e.g. by using different classifications;
- The wording of the question or the formulation of the response categories can differ.

Each of there problems can seriously affect comparability, and so each of these needs to be adequately addressed before a meaningful comparison between MS can be made. The present report is primarily concerned with the last problem, i.e. with ways to cope with differences in wording and categories, and to a some extent with the third problem, that is, the problem causes by using different definitions and classifications.

To illustrate the type of problems that we deal with, suppose we want to get insight into the level of disability of the populations of different MS. Disability is often measured by questionnaire items in health surveys. Many MS conduct such surveys, but the precise way in which disability is measured could be quite different. For example, for walking disability, the U.K. health survey contains a question *How far can you walk without stopping/experiencing severe discomfort, on your own, with aid if normally used?* with response categories "can't walk", "a few steps only", "more than a few steps but less than 200yds" and "200yds or more". The Dutch health interview contains the question *Can you walk 400 metres without resting (with walking stick if necessary)?* with response categories "yes no difficulty", "yes minor difficulty", "yes major difficulty" and "no". Both items obviously intend to measure the ability to walk of the respondent, but it is far from clear how an answer on the U.K.-item can be compared with one on the Dutch item.

1.2.2 Pre-harmonisation and post-harmonisation

There are two broad strategies to deal with incomparability: pre-harmonisation and post-harmonisation.

Pre-harmonisation is the royal road to solve comparability problems. The idea is that, once and for all, all MS will start collecting comparable data. The major advantage is that comparability is guaranteed since every office works in the same way using the same instrument. As easy as this may sound however, it is not trivial to actually achieve this in practice. The national data collecting agencies of the individual MS will generally be very reluctant to change their sampling methods and instruments. Their major argument is that a change of the current practice will break the comparability to historic data. In that case, pre-harmonisation does not solve the problem, but puts it on a different level, that is, at the level of the national offices of the MS.

By its nature pre-harmonisation will only work for new, and not for existing data. In addition, even if done well, pre-harmonisation could still yield implausible results that will raise comparability issues. As an demonstration of this, consider the single question *How is your health in general?* and a five point Likert response scale “very good, good, fair, poor, very poor”. This question was posed (after translation) in 12 countries of the European Union, using the same survey and methods within the context of the 1994 European Community Household Survey (Eurostat, 1997). Figure 1.1 is taken from Sadana *et al*. (2000) and contains the age-sex-standardised proportions of the responses per country. Note that the category ‘very good’ health is reported by as much as 53% of the Danish and as little as 8% of the Portuguese population. Also, nontrivial differences occur for the bad and very bad categories. It is very unlikely that these results reflect real differences in subjective health. Maybe there is a bias because of cultural differences in the interpretation of the question. Whatever the explanation is, such differences raise suspicion that pre-harmonisation may not be enough to solve all comparability problems.

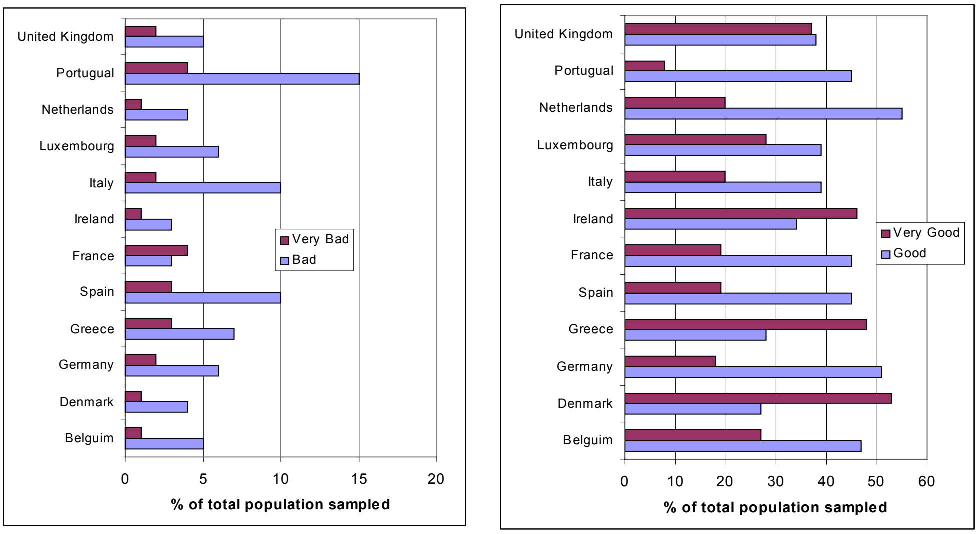


Figure 1.1 – Pre-harmonisation result: Response on the question *How is your health in general?* in 12 European countries. Source: Sadana et al. (2000).

Post-harmonisation is the murky way to solve comparability problems. The idea is that we can somehow transform incomparable data into a comparable version, and use the latter in our analyses. The big advantage is that we can use existing data. The disadvantage is that we often do not know what the transformation should be, and whether applying it will affect the results. In addition, it is sometimes simply impossible to transform the data into a comparable form without making strong, untestable assumptions. On the other hand, post-harmonisation is often the only option if we are to make any progress. Given that situation, we should try to use the best avail-able scientific technology to make post-harmonisation work. This implies that we should be explicit about the concepts, assumptions and limitations of the method.

1.3 Goal of the project

This project aims to develop and demonstrate a new post-harmonisation technology, called response conversion. This methodology is of potential value to the HMP for converting existing health information into community indicators. The present project was a pilot project that focused on the conversion of walking and dressing disability information (in terms of physical disability and activity limitations) from all Member States. The goals of the pilot project were:

- to demonstrate the response conversion methodology on a practical problem,
- to identify key problems, if any.

1.4 Contents

Chapter 2 introduces the new method by applying it to an intentionally simple problem. It describes the essential concepts and main assumptions of the method. Chapter 3 applies the method to walking disability, and chapter 4 addresses the problem of dressing disability. Chapter 5 concludes this reports, and discusses the usefulness of the methodology for use within the HMP.

2 The method of response conversion

This chapter introduces a new method for harmonising existing information. An essential element of the method is the conversion of actual responses to a common scale, and therefore the method will be termed *response conversion*. The text illustrates the main concepts by applying the method to a simple problem involving only a few questionnaire items and studies. Section 2.1 describes the comparability problem from a methodological perspective. Section 2.2 outlines three ways to attack the problem. Section 2.3 is a more detailed description of the statistical principles of the method. This description has been kept as non-technical as possible, though some technicalities could not be avoided.

2.1 Description of the problem

The objective of the Health Monitoring Program is to set up a system in which the health of people in different Member States of the European Union can be compared. This system will have to be based on *existing* population surveys. This requirement introduces new issues regarding the comparability of information across Member States. The present section outlines some complexities of the comparability problem in the case of two populations.

Suppose that we are interested in comparing two populations, and that we have access to one survey for each population. Each survey provides information on a sample of respondents. In the sequel, we will denote such surveys as *target studies*, that is, studies that contain the information that we want to compare. Survey instruments typically consist of a standardised set of *questionnaire items*. For a given field of health, we may be able to identify specific instruments or items that measure that particular aspect of health. In the sequel, we will call these *target instruments* and *target items*.

If both studies use equivalent instruments/items, there are (in principle at least) no problems regarding the comparability of content. The target studies could still differ in their sampling methods, in their ways for collecting data (e.g. interview, self-report), or in other ways. Those differences have to be accounted in any valid comparison, for example through differential weighting of sampling units. Though important, such problems are not the object of study of the present report.

This report concerns the problem that target studies may contain measurements of the same thing, but use different instruments or items. Let A and B denote two target items that measure the same characteristic. In general, responses on A and B can only be meaningfully compared if the scales on which they are measured have the same origin and the same unit. If A and B are different, it is not informative to directly compare their responses since differences in the response distribution of A and B may be due to

1 real differences between populations;
2 systematic differences between the target items;
3 a combination of both.

In practice, interest focuses on comparing (sub)populations, which presupposes that possibility 1 is true. Without any additional information or assumptions, it is however impossible to distinguish between the three possibilities. Thus, we generally do not

Table 2.1 Example of two studies measuring walking disability.

Item	Description	Response categories	Study ERGOPLUS n=306	EURIDISS n=292
SIP01	I walk shorter distances or often stop for a rest.	0 = No	276	
		1 = Yes	28	
GARS9	Can you, fully independent-ly, walk outdoors (if neces-sary, with a cane)?	0 = Yes, no difficulty		145
		1 = Yes, with some difficulty		110
		2 = Yes, with much difficulty		29
		3 = No, only with help from others		8

know whether differences between the responses on A and B reflect real population differences.

To illustrate this point further, we take an excerpt from the combined data on walking disability analysed by Van Buuren & Hopman (2001). The ERGOPLUS study (Odding et al, 1995) contains responses on the item SIP01 from the ambulation scale of the SIP (Sickness Impact Profile). Likewise, the EURIDISS study (European Research on Incapacitating Diseases and Social Support) contains responses on the item GARS9 with four response categories from the GARS instrument (Suurmeijer et al, 1994). The problem is to compare the amount of waking disability between both studies.

Table 2.1 contains the counts per response category in both studies. In the sequel, we will always code the response categories starting with zero and such that the lowest categories correspond to the lowest disability levels. It is obvious that both SIP01 and GARS9 measure some aspect of walking disability, but it is not clear how this information could be used to compare the amount of walking disability between the ERGOPLUS and EURIDISS studies.

2.2 Three strategies to address comparability

We distinguish three major strategies to address to comparability issues raised in Section 2.1. These are:

1. *by fiat*: Assume a common score system, recode the responses into a common system, and compare;
2. *link by item*: Identify additional items on walking disability (within both studies) that are common to both studies, and exploit the overlap to compare studies;
3. *link by study*: Look for other (third) studies that contain both items, and use the relationship between both target items in comparing both target studies.

We now discuss each of these in more detail.

At first sight, the first strategy (*by fiat*) may seem most appealing. If we would have a way to recode the responses on both items into a comparable system, then we can simply use the recoded data to gain insight into differences in walking disability between both samples. For example, in Table 2.1 we can postulate that categories 0 and 1 of GARS9 are equivalent to category 1 ("No") of SIP01, and that categories 2 and 3 of GARS9 are equivalent to category 1 ("Yes") of the SIP01. We can then recode GARS9 into two categories that are, by definition, comparable to SIP01. In the above case, this

Table 2.2 Example data with an additional bridge item.

Item	Description	Response categories	Study	
			ERGOPLUS	EURIDISS
			n=306	*n=292*
SIP01	I walk shorter distances or often stop for a rest.	0 = No	276	
		1 = Yes	28	
HAQ8	Able to walk outdoors on flat ground?	0 = Without any difficulty	242	178
		1 = With some difficulty	43	68
		2 = With much difficulty	15	42
		3 = Unable to do	0	2
GARS9	Can you, fully independent-ly, walk outdoors (if neces-sary, with a cane)?	0 = Yes, no difficulty		145
		1 = Yes, with some difficulty		110
		2 = Yes, with much difficulty		29
		3 = No, only with help from others		8

would yield 110 + 145 = 255 EURIDISS respondent in the "No"-category and 8+29 = 37 EURIDISS respondents in the "Yes"-category. We can now compute the Yes/No ratio's for both studies (37/255 = 0.15 for EURIDISS and 28/276 = 0.10 for ERGO-PLUS), and conclude that the EURIDISS sample is considerably more disabled than the ERGOPLUS sample. We have solved the comparability problem by "assuming away" any systematic differences that might exist.

Some comments are in order on this strategy. First, it is only possible to move into the direction of the item with the lowest number of response categories. This will inevitably lead to a loss of information for items that have more refined response systems. In principle, one could try to solve this problem by splitting a crude category into refined sub-categories. For example, one can divide the 28 "Yes"-respondent from the ERGOPLUS study over categories 2 and 3 of GARS9 (e.g. by assigning 8 respondents to category 3 and the remaining 20 respondent to category 2). It is however difficult to see how such splitting proportions should be chosen. The whole procedure relies on arbitrary and untested criteria, and could therefore generate considerable debate. There is no way of knowing whether the chosen cut-point is actually correct. The *by fiat* strategy should therefore only be chosen in cases where 1) the possibility of dispute is relatively small, 2) the response categories are finely grained, and 3) a clear authority can endorse the system.

Pursuing strategies 'link by item' and 'link by study' requires additional data. We first look at strategy 2 (*link by item*) in more detail. If both studies contain additional items on walking disability that are common to both studies, then this information provides a link between both studies. In this example, both studies also administered the HAQ8 item *Are you able to walk outdoors on flat ground?*. See Table 2.2. Such an item that connects two studies is called a *bridge item*.

The HAQ8 item provides a means to compare both studies. Simple visual inspection of the category frequencies for both studies tells us that, like before, the EURIDISS sample is more disabled than the ERGOPLUS sample. Note that thus far, we have done nothing new. We have simply replaced an incomparable set of items (SIP01 and GARS9) by a comparable item (HAQ8) that happened also to be administered in both

Table 2.3 Example data with an additional bridge study.

Item	Description	Response categories	Study		
			ERGOPLUS	BRIDGE	EURIDISS
			n=306	*n=300*	*n=292*
SIP01	I walk shorter distances	0 = No	276	215	
	or often stop for a rest.	1 = Yes	28	85	
GARS9	Can you, fully	0 = Yes, no difficulty		150	145
	independently, walk	1 = Yes, with some difficulty		105	110
	outdoors (if necessary,	2 = Yes, with much difficulty		34	29
	with a cane)?	3 = No, only with help from others		11	8

studies. Of course, we could have started with the HAQ8 right away, and not be concerned with either SIP01 or GARS9 at all.

Now imagine that we have two *new* studies, where the first contains only SIP01 (but not HAQ8) and the second contains only GARS9 (but not HAQ8). The interesting question then is: It is possible to use the information contained in Table 2.2 in such a way that we can validly compare the two new studies, even in the absence of bridge items? The answer is yes, given that both of the following assumptions hold:

- the bridge item measures the same characteristic as the target items;
- the bridge item is equivalent in both studies.

If true, it is possible to define a statistical model for converting observed scores into a comparable form. In later applications, this model can be used to convert information without the need of any bridge items. Section 2.3 describes the technique in more detail. More precise definitions of both assumptions will also given there.

The third strategy (*link by study*) is the logical complement of the second. Suppose a third study is available, that administers both target items to a third population. Such a study is called a *bridge study*.

Table 2.3 contains an example of observations from a hypothetical bridge study. The sample size (n=300) of the bridge study is chosen to be similar to the target studies for ease of comparison. Equality of sample sizes is not a requirement in actual application. The bridge study administers both SIP01 and GARS9 to a third population. The comparison of the score distributions on GARS9 suggests that the disability of the bridge population is almost equal to that in de EURIDISS study. In contrast, the difference on SIP01 with the ERGOPLUS study is substantial. Combining these two findings suggests that, like before, the ability level in ERGOPLUS is higher than in EURIDISS.

The validity of the link-by-study strategy depends on the following assumptions:

- the items in the bridge study are equivalent to those in the target studies;
- the relationship between both items does not depend on the ability level of the sample.

It is important to observe that it is not required that the ability level of the bridge study is comparable to one of the target studies. The second assumption implies instead that the relationship between the items is assumed to be the same in all studies. This condition is much weaker.

This section introduced three strategies to address comparability issues. No strategy is clearly superior to the others. In practice, one would mix all strategies. The classic *by fiat* relies on (arbitrary) re-scoring rules. It can work quite well if consensus about such rules can be obtained. Both other strategies consider the use of additional data, but differ in the precise type of information they need. Both aim to optimally place different items onto a common scale. The points raised in this section are related to a field that is known as *test equating*. The work of Vale (1986) on linking designs can be used as a useful starting point for further exploration of this field. Kolen and Brennan (1995) collected a large number equating techniques. In their terminology, the link-by-item and link-by-study approaches are special cases of the *nonequivalent linked group design*. The next section describes the statistical methodology that we use to exploit the overlap introduced by the additional information.

2.3 Statistical principles

Response conversion assumes the existence of a continuous *latent trait* θ that underlies all items. In the data of Tables 2.1-2.3, the latent trait θ can be interpreted as walking disability. A latent trait is a theoretical construct with some of the following properties. A latent trait varies continuously and can take on all values. The ability level of each person in the sample can be characterised by a position θ_i on the trait. The trait is latent, which means that it cannot be observed directly. So the "true value" of θ_i for person *i* is not known, and can only be observed through the manifest item responses.

2.3.1 Relation between disability and response probability

The main idea of response conversion is that the value of the latent trait governs the probability of responding in a specific response category. For low θ_i (e.g. no disability), the probability of answering in the most severe disability response categories is low. For example, a person without any walking restrictions is unlikely to respond in category 1 ("Yes") of SIP01, or in category 3 of GARS9. On the other hand, persons with severe restrictions (i.e. with high values of θ_i) have high probabilities to respond in those categories, and have low probabilities to respond in the lower categories. Figure 2.1 is an illustration of this idea for SIP01 and GARS9.

These plots are known as Category Characteristic Curves (CCC). One can make such a plot for each item. Such curves are the result of the fit of a statistical model on the appropriately linked data. Section 2.3.4 gives more detail about the precise choice of the model. We like to emphasise at this point that the position of each person on the latent trait is not made up by the investigators, but estimated from the observed data using techniques from item response theory. Section 2.3.2 provides more details on how this is done.

The linked data of Table 2.2 were used as input for the curves in Figure 2.1. The horizontal axis orders walking disability from no disability (left) to high disability levels (right). The horizontal axes in the different plots are identical. So, if we know the disability position θ_i of a person, then we can read off the response probabilities for every item. For example, someone with $\theta_i = -1$ has a probability of 0.62 of responding in category 0 of SIP01, and a probability of 0.38 of answering category 1. The same

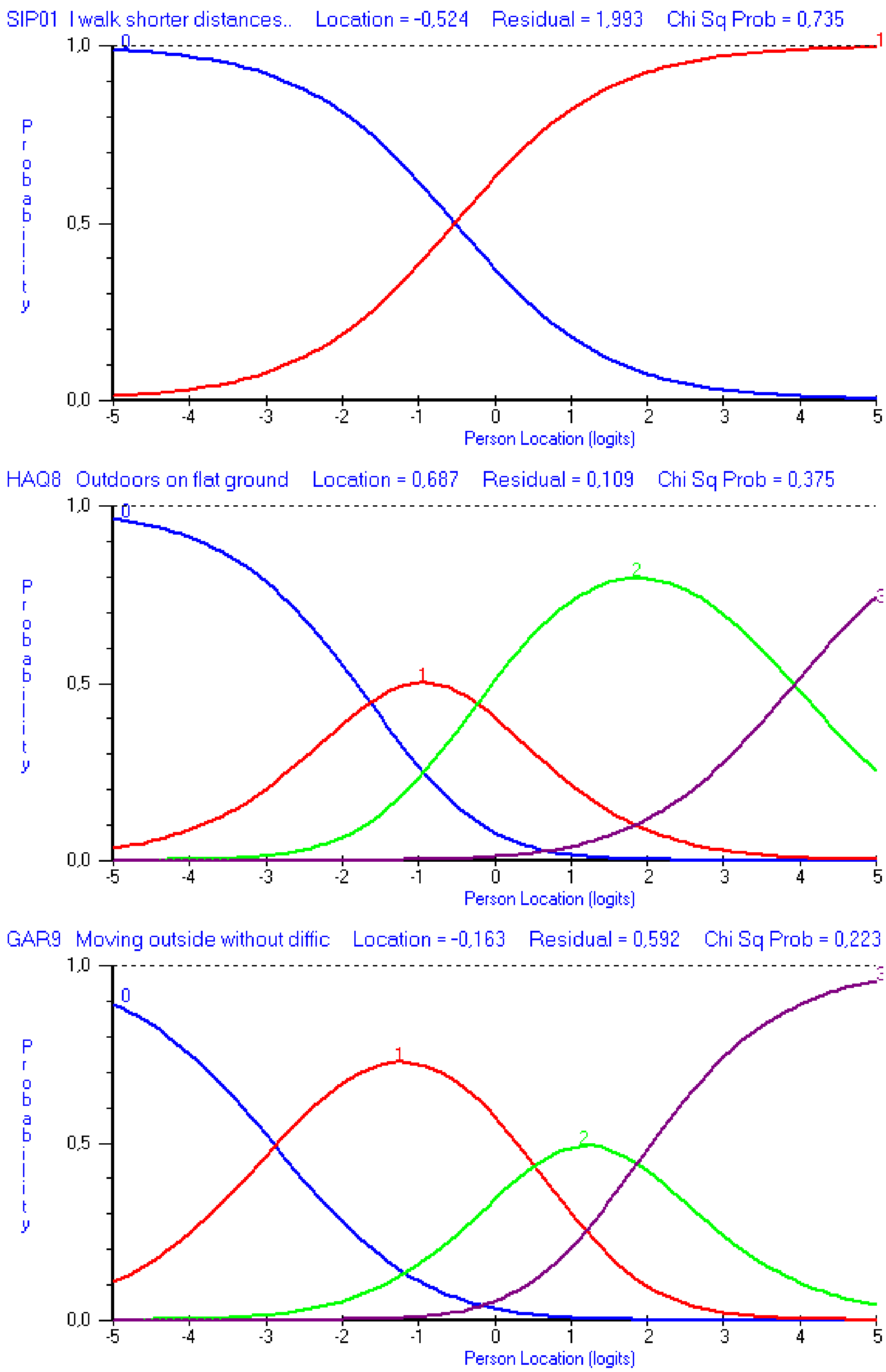


Figure 2.1 Category characteristic curves: Probability of responding in each category for a given level of disability.

person has probabilities of 0.27, 0.50, 0.23 and 0.00 to respond in respectively categories 0, 1, 2 and 3 of <u>HAQ8</u>. The response probabilities for <u>GARS9</u> are respectively 0.11, 0.72, 0.16 and 0.01. The end points of the scale are arbitrary. Figure

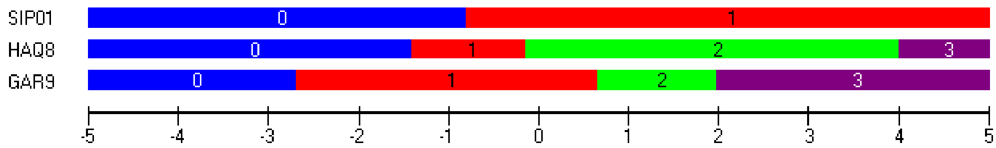


Figure 2.2 Threshold map for three items for measuring walking disability.

2.1 was scaled such that the thresholds have zero mean. If desired, one could linearly transform the scale (e.g. to a range of 0 to 100) without affecting the relationship between disability and response probability.

Figure 2.2 is known as a *threshold map*, and is a more compact way to represent the CCC's. The figure codes the category with the highest response probability at a given disability level as a coloured bar. The transition locations correspond to the intersection points in the CCC's in Figure 2.1. These points are known as thresholds. Knowledge of the thresholds is enough to reconstruct the CCC's. As we will see later, the threshold map is a graphic representation of a *conversion key*.

How are the locations of the thresholds determined? Suppose that a person with $\theta_i = -1$ responds to both SIP01 and HAQ8. In that case there are 2*4=8 possible combinations of categories, or *response patterns*. Now, what is the expected probability of each response pattern? The answer can be found by multiplying the separate probabilities and divide it by the number of items. For example, the expected probability of observing the combination (0,0) at $\theta_i = -1$ is equal to 0.62 * 0.27 / 2 = 0.084. We can compute this probability for all patterns, and the sum of probabilities over all patterns will automatically add up to one. The expected probability may of course differ from the actually observed probability in the data. Such differences can be minimised by appropriately placing the CCC's, a task that is typically done by a dedicated computer program. CCC's and their thresholds are estimated such that the probabilities of the observed and expected patterns agree as much as possible.

2.3.2 *Estimation of disability*

One thing has not yet been mentioned. How do we know the location θ_i of person *i*? It will be clear that θ_i will depend on the answers given by person *i*. The answer pattern (0,0) corresponds to lower disability levels than patterns (0,3), (1,0) or (1,3). In some circumstances, the sum score over all items can be used to estimate θ_i. In the general case however, estimation is a bit more complicated. It is beyond the scope of the report to discuss these matters, and we refer to Hambleton *et al.* (1991) for more detail. The important thing to remember here is that, for each person, the location θ_i is estimated from the observed response pattern. In the sequel, we will use the Bayesian EAP estimator (Bock & Mislevy, 1982) with a left-skewed lognormal prior with logmean 1 and logvariance placed on the interval –5 to +5.

Using this method, we can calculate the posterior disability distribution for each response pattern. Table 2.4 contains the mean of the posterior distributions for all patterns consisting of responses on exactly one item. Note that the mean disability corresponding to categories 0 and 1 of HAQ8 and GARS9 are similar. In contrast to this, the mean disability of categories 2 and 3 of HAQ8 is higher than in GARS9.

Table 2.4 Mean disability per category on the common scale for response patterns consisting of one item.

Item	Response category			
	0	1	2	3
SIP01	-2.44	-0.49		
HAQ8	-2.72	-1.71	0.06	2.68
GARS9	-2.89	-1.94	-0.22	2.00

The numbers in Table 2.4 can be used to estimate the mean disability level of a given study, a parameter that is useful to compare across studies. The mean disability level can be found by calculating the average of mean disabilities, weighted by response frequency of the study. For example, the mean disability level according to HAQ8 in ERGOPLUS is equal to ((–2.72*242)+(–1.71*43)+(0.06*15)+(2.68*0))/300 = –2.44. Likewise, in EURIDISS it is –2.05. The difference in mean ability (0.39) confirms earlier notions of higher disability levels in EURIDISS. Note that disability differences between both studies are now quantified.

2.3.3 *Comparison*

Let us now return to the central problem: How can we compare the amount of walking disability if we only have observed responses on SIP01 and GARS9? Given that we have an appropriate conversion key, comparison of the mean disability becomes quite simple. For ERGOPLUS, we calculate the mean disability level measured by SIP01 as ((–2.44*276)+(–0.49*28))/304 = –2.26, while the mean disability in EURIDISS based on GARS9 is equal to –2.13, so the difference is 0.13. Thus, even in the absence of a bridge item or bridge study, we see that the ERGOPLUS sample has on average fewer disabilities. The main progress that we've made is that information contained in different items is now expressed on a common scale. This common scale, or a linear transformation of it, can be used to compare the level of disability of the underlying samples.

Observe that the mean disability difference between target items is smaller than between bridge items (0.13 vs. 0.39). This is a general phenomenon, and it is related to overfitting (the model is fitted on HAQ8, and applied to SIP01 and GARS9). The magnitude of the effect is relatively large here because there is one bridge item and one target item per study. In some sense, this is the price for having incomplete data and crude measurements. This topic has some statistical subtleties, and we will come back to it in Chapter 5.

Application of the method to the situation where several items are available for disability estimation is straightforward. Using multiple items makes the disability estimate more reliable. In the example above, we compared *mean* disability levels. There is however, nothing in the method that restricts it to the mean. In fact, any aspect of the disability distribution (mode, variance, 95th percentile) can be used for comparing samples. For example, the 95th percentile of the disability distribution is more sensitive to changes in extreme disability, and thus could be a more useful indicator in specific applications. Figure 2.3 contains the disability distribution estimated from SIP01 and GARS9. The 95th percentiles are located at 0.00 (ERGOPLUS) and +0.72 (EURIDISS). This means that the 5 percent most disabled

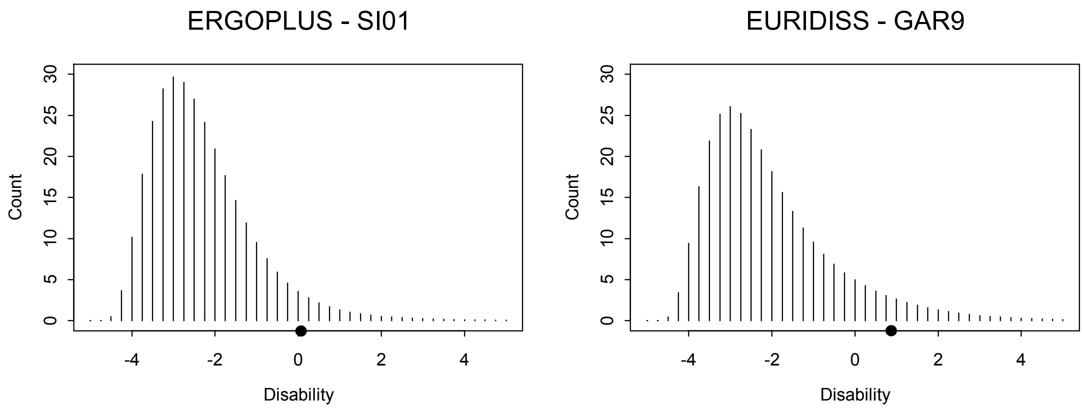


Figure 2.3 Posterior distributions (on the common scale) of the ERGOPLUS and EURIDISS samples. The left distribution is estimated from the SIP01, while the right panel is estimated from the GARS9 item. The dots on the horizontal axes indicate the position of the 95th percentiles.

people in the ERGOPLUS sample starts at $\theta = 0$, while in EURIDISS the 5 percent most disabled individual starts at a higher disability level $\theta = +0.72$.

2.3.4 *Choice of the model*

We like to point out that the precise relation between θ_i and the response probability (the CCC) can be specified in many different ways. Figure 2.1 represents just one way of doing it. Many models appear in the psychometric literature. The field is collectively known as *Item Response Theory* (IRT). Useful introductory text into IRT are Wright & Masters (1982) and Hambleton *et al.* (1991). More advanced works are Fischer & Molenaar (1995), Van der Linden & Hambleton (1997) and Boomsma *et al* (2001). Important theoretical distinctions between different models can be made, and different scientific schools stressing different aspects exist. From a practical point of view, the actual differences (when fitted to data) are usually not that large. In our experience, all models do more or less the same, but the results have different theoretical properties.

In this report we have chosen to use the logistic three-parameter full model with location, dispersion and skewness parameters developed by Andrich. This model should not be confused with the Birnbaum 3-parameter model that includes a guessing parameter. The model is very flexible in the sense that it can describe a large variety of relations between θ_i and the response probabilities. The method uses a pair-wise estimation method that can handle linked data quite well (cf. Andrich, 1988, p. 57-59). High-quality software for estimating and inspecting the model is commercially available (RUMM Laboratories, 2000).

2.4 Conclusion

Three strategies to achieve comparability can be distinguished: *by fiat*, *link by item* and *link by study*. In practice, one typically makes a combination of these options. Response conversion is a method that assists in the second and third strategy by systematically exploiting any information overlap between different studies. Overlap can occur in items, in samples, or in both, leading to different linked data matrices.

The major tasks in the practical application of response conversion consist of

1. Identification and construction of the linked data matrix;
2. Construction of a conversion key that place different items on a common scale;
3. Application of conversion key to estimate disability on a common scale.

Steps 1 and 2 need to be done only once, where step 2 results in a conversion key. A separate conversion key is needed for each topic. Once a conversion key is available, applying it to new data is cheap and easy, and can be done on a routine basis.

The next two chapters will apply these principles to harmonise disability surveys from the Member States of the European Union.

3 Walking disability

Chapters 3 and 4 deal with harmonisation of information about walking and dressing disabilities. Disability is headed under the functional and activity limitations (code 2.3) within the ECHI list of indicators (ECHI working group, 2000). Walking and dressing disability have been chosen because they are conceptually easy, have serious personal consequences in daily life, are being collected in many surveys. In addition, these topics allow us to build upon and extend earlier work (Van Buuren et al, 1996, 2001; Hopman-Rock et al, 2000).

Response conversion methodology consists of a number of steps:
1. Choose a specific area of disability (e.g. walking or dressing);
2. Identify the instruments and items that are used in each member state for measuring this type of disability;
3. Search the literature for *bridge studies* and *bridge items*;
4. Construct a *linkage diagram* that shows if, and how, prevalence items can be linked by means of bridge studies and bridge items;
5. Formulate explicit *equivalence assumptions* about which items can be considered equivalent;
6. Obtain *microdata* from bridge studies;
7. Construct a *linked data set* containing the combined data from the bridge studies;
8. Conduct *preliminary statistical analysis*;
9. Check equivalence assumptions;
10. Construct the *conversion key*;
11. *Express prevalence data* on a common scale using the conversion key.

Below, we address these steps in more detail for walking disability.

3.1 Type of disability

The chapter is restricted to instruments for measuring walking disability. A disability is "any restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being" (WHO, 1993). Walking disability is defined according to the ICIDH-D code 40. This includes ambulation on flat terrain, and excludes occasional steps in terrain, climbing stairs, other climbing and running disability. It is not always precisely clear how a specific item should be classified. It is sometimes difficult to distinguish between code 40 (Walking disability) and code 41 (Traversing disability). We have taken a fairly liberal approach with respect to the inclusion of items. Items that *explicitly* refer to occasional steps or to climbing stairs are excluded. However, activities such as 'walking 400 meters' or 'move around the house' are included.

3.2 Questionnaire items walking disability in the EC

We used a number of sources to identify walking and dressing items that were being collected throughout the European Union. These include Hupkens (1998), Rasmussen et al. (1999) and Robine et al (2000).

An item consists of a question and a set of response categories. Variations occur in the exact formulation of the question, as well as in the precise response categories that are

Table 3.1 Target item for measuring walking disability in the European Community.

Country	Study	Survey	Walking
Austria	A01	Microcensus Survey on Disabilities	DWELLING
Belgium	B01	Health Interview Survey	FAR1
			AFF1KM
Denmark	DK01	Danish Health & morbidity survey	W400B
			AFF1KM
			AFFS100M
			AFF1100M
Germany	D01	Health Interview Survey	AFF1KM
			LIM1BL
			LIMSEVB
Spain	E01	Disability, Impairment & Health Status survey	DIFOUT
	E02	Spanish Health Interview Survey	FAR10
Finland	FIN01	Health care survey	W400D
France	F01	National Disability Interview	FAR3
	F02	National Health Interview Survey	FAR4
Greece		- no items found -	
Ireland		- no items found -	
Italy	I01	Italian Survey on health conditions	FAR9
Luxembourg		- no items found -	
Netherlands	NL01	Health Interview Survey	W400C
			ADL8
			ADL5
Portugal	P01	Health Interview Survey	FAR5
Sweden	S01	Swedish Living Conditions Interview Survey	BRISK
United Kingdom	UK01	Health Survey for England	FAR2
	Non-EC member states		
Norway	N01	Health interview survey (Helseundersokelsen)	FAR4
Switzerland	CH01	Swiss health survey	FAR11
Canada	CAN01	Hals	W400A
			ROOM
New Zealand	NZ01	1996 Household Disability Survey	RUGBY

used. Table 3.1 lists the walking items that we could find for the EC member states, as well as some non-EC countries. For most countries, there is only one survey that is considered to be nationally representative, but Spain, and France have two such studies. In terms of Chapter 2, Table 3.1 contains the *target items* for walking disability, that is, the items that we want to compare. The target items are identified by item names. For walking, a total of 27 target items are currently being used.

Appendix A contains a description of each study. The full description of each item can be found in Appendix B, which a full alphabetical of all 81 different walking items that are used in this study. For clarity, target items within this set are indicated.

Target items may differ on many aspects. The most important differences relate to the concepts behind the item formulation. For example, some items ask how difficult it is to walk a *fixed distance* (often 400 metres), other concentrate on *how far* you can walk, or *how long* you can walk without difficulty, and still others focus on how *limited* your activities are. These *conceptual groups* can sometimes be traced back to a common ancestor. For example, fixed distance items derive from the OECD long-term disability questionnaire ("Can you walk 400 metres without resting?"). Items that use 'how far' in the question are variations on the WHO-Europe long-term disability questionnaire ("What is the furthest you can walk on your own without stopping and without severe discomfort?"). The 'are you limited'-group is similar to questions in the 36-Item Short-Form Health Survey (SF-36). Within conceptual groups, different variations occur in either the exact wording of the questions, in the response categories, or in both.

There appear to be two groups of countries. Belgium, Spain, France, Italy, Portugal and United Kingdom all use a variation on the "How far"-question. The other group includes Denmark, Finland, Netherlands, New Zealand and Canada, and uses a variation on the question "Can you walk 400metres without resting?". Other countries use still other formulations, or do not measure walking disability at all. Walking items could be found for 12 of the 15 EC member states. No information for Greece, Ireland and Luxembourg could be identified.

It will be clear that the comparison of walking disability across different member states will be hampered by this pluriformity. It will also be clear that this situation is not unique for walking disability. Similar problems occur for other types of health measurements.

3.3 Bridge studies and bridge items

Bridge studies and bridge items are needed in order to be able to link different items to a common scale. We searched the literature and used our networks to identify studies that collected data on two or more walking or dressing items. This yielded 14 additional studies, mainly from the United Kingdom and the Netherlands. These studies were used to identify overlapping information in walking disability items by means of a linkage diagram. Bridge studies are indicated in Appendix A. Most items in Appendix B function as bridge items.

3.4 Linkage diagram

A next step consists of the construction of the so-called linkage diagram. Figure 3.1 contains the linkage diagram based on the identified bridge studies and bridge items. The 'Y'-symbol in a cell indicates that the specific study-item combination occurs. Cells with the 'Y'-symbol are also coloured to make them easier to find. The precise meaning of the colour coding will become clear in Section 3.5. Items can be classified into a limited number of conceptual groups. As far as possible, the items in Figure 3.1 are sorted such that each item is located near the other members of its conceptual group.

The most important use of the linkage diagram is to see which studies are linked, that is, if there is a path that connects them. The existence of a link is a technical requirement for scaling different items on a common scale. In the present diagram, rather few items are directly linked. For example, FAR7 and MANAGE are linked by the bridge item AIDS1.

NAME	Block	CHO1	EO2	FO1	FO2	NO1	IO1	PO1	BO1	UKO1	UKO8	UKO2	UKO5	UKO4	DO1	DKO1	NZO1	NLO9	UKO3	NLO3	NLO2	UKO7	UKO6	AO1	CANO	EO1	NLO4	NLO6	NLO7	NLO1	EURO	FINO1	SO1
far11	B	Y	I	I	I	I	I	I	I	I	I	I	I	I																			
far10	B	I	Y	I	I	I	I	I	I	I	I	I	I	I																			
far3	B	I	I	Y	I	I	I	I	I	I	I	I	I	I																			
far4	B	I	I	I	Y	Y	I	I	I	I	I	I	I	I																			
far9	B	I	I	I	I	I	Y	I	I	I	I	I	I	I																			
far5	B	I	I	I	I	I	I	Y	I	I	I	I	I	I																			
far1	B	I	I	I	I	I	I	I	Y	I	I	I	I	I																			
far2	B	I	I	I	I	I	I	I	I	Y	I	I	I	I																			
far8	B	I	I	I	I	I	I	I	I	I	Y	I	I	I																			
far6	B	I	I	I	I	I	I	I	I	I	I	Y	I	I																			
far7	B	I	I	I	I	I	I	I	I	I	I	I	Y	Y																			
howlong												Y		Y																			
stand3												Y	Y	Y																			
stand1	L											Y	Y	Y				I															
stand10m	L											I	I	I				Y															
slowly	K											Y		Y						I	I												
sip12	K											I		I						Y	Y												
qmilehil														Y																			
furthest														Y																			
w200ws										Y																							
stand2											Y																						
limhmil	C								I					Y	I	I																	
aff1km	C								Y					I	Y	Y																	
lim100y	D								I					Y	I	I																	
lim1bl	D								I					I	Y	I																	
aff1100m	D								Y					I	I	Y																	
affs100m	E								Y						I	Y																	
limsevb	E								I						Y	I																	
qmilews												Y																					
qmilelev												Y																					
limmile														Y																			
house														Y																			
sip11	M												I	I			I	I	I	Y	Y												
aids1	M												Y	Y			I	I	Y	I	I												
aids2	M												I	I			Y	I	I	I	I												
aids3	M												I	I			I	Y	I	I	I												
rugby																	Y																
walk10m																		Y															
walk30m																		Y															
stand30m																		Y															
diff																		Y															
onlyind																						Y											
mobil																							Y										
euro																						Y											
pain																						Y											
london																						Y											
difout																										Y							
fsipain																					Y												
fsihelp																					Y												
walkindo	F																	Y						I									
dwelling	F																	I						Y									
walkoutd	J																	Y	I														
manage	J																	I	Y														
bart																							Y										
helpout	N																				I	Y											
aims5	N																				Y	I											
sip01																				Y	Y												
sip08																				Y	Y												
sip07																				Y	Y												
aims6																				Y	Y	Y											
wkatall																						Y											
mobind																						Y											
fsidiff	H																				Y				I	I					I		
moveins	H																				I				I	Y					I		
room	H																				I				Y	I					y		
adl8	I																										I			Y	I		
gars9	I																										Y			I	I		
outdoors	I																										I			I	Y		
w400a																									Y								
w400b	A															Y												Y	Y	I	Y	I	
w400c	A															I												I	I	Y	I	I	
w400d	A															I												I	I	I	I	Y	
ppt7																												Y	Y				
gars7	G																		I								Y			I			
adl5	G																		I								I			Y			
manshe	G																		Y								I			I			
haq8																				Y	Y						Y						
whoshe																			Y														
brisk																																	Y

Figure 3.1 Linkage diagram of items for measuring walking disability.

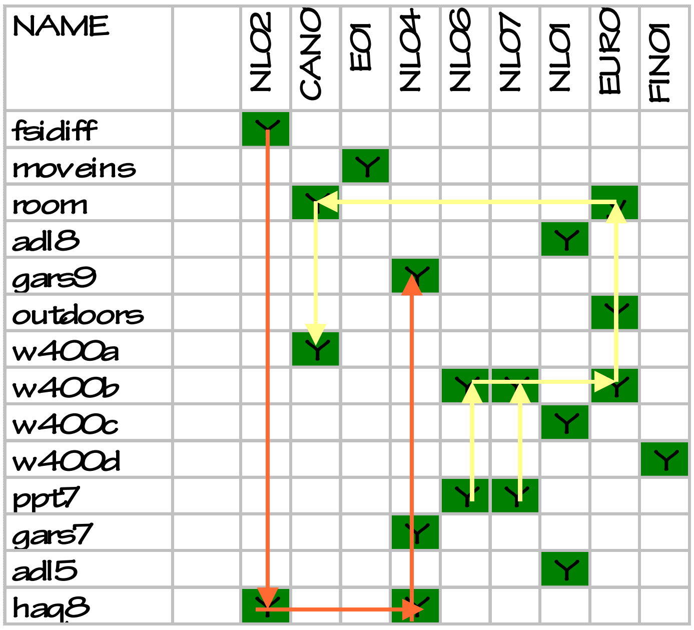


Figure 3.1 Available item-study combinations for a part of the linkage diagram. Arrows indicate how items are linked.

Other examples are illustrated in Figure 3.2, which is a part of the lower right corner of the linkage matrix. This diagram shows that FSIDIFF is linked to GARS9 and GARS7 through the bridge item HAQ8. Also, PPT7 is linked to W400A through bridge items W400B and ROOM. In the latter case, EUR01 is a bridge study.

It appears that such chains are relatively isolated features of the linkage diagram. Better linkage can be obtained by making explicit equivalence assumptions, i.e. by assuming that specific items measure the same.

3.5 Equivalence assumptions

Items FAR1 to FAR11 are all variations on the same idea. In order to be able to compare results from different studies, we often need simplifying assumptions with regard to the formulation of the question and with respect to the response categories. In this light one should question oneself: Do additions like 'on your own', 'on a level ground' or 'with a walking stick if needed' really affect the answer of the respondent, or are these variations by and large cosmetic? Can the response categories be easily coded into a common, perhaps cruder, coding system? The answers to the questions determine whether there is enough ground to equate the responses on two items, that is, to declare them as 'essentially identical'. Of course this process is a bit arbitrary, but if we are to make any progress on harmonisation, these steps are inevitable. An advantage of the process is that any assumptions must be made explicitly, thus providing a means for independent verification. The process is a form of the *by fiat*-strategy of Section 2.2.

We assume that the following items in Table 3.1 are equivalent, or can be made equivalent after appropriately rescoring the response categories. Denmark, Finland and The Netherlands use variations of the "400 metres" item with four response categories. Item W400A has two response categories, and so we decided that W400A could not be part of the *item block* A. The equivalence assumption implies that we can directly compare the responses on these items. Similarly, countries using the "how far" question

Table 3.2 Equivalence assumptions about items for measuring walking disability.

Block	Categories	Equivalent items
A	4	w400b = w400c = w400d
B	3	far1 = far2 = far3 = far4 = far5 = far6 = far7 = far8 = far9 = far10 = far11
C	3	limhmil = aff1km
D	3	lim100y = aff100m = lim1bl
E	3	affs100m = limsevb
F	3	walkindo = dwelling
G	4	gars7 = adl5 = manshe
H	2	fsidiff = moveins = room
I	4	adl8 = gars9 = outdoors
J	4	walkoutd = manage
K	2	slowly = sip12
L	2	stand1 = stand10m
M	2	sip11 = aids1 = aids2 = aids3
N	2	helpout = aims5

can be compared under equivalence. Some of the "How far" items have to be recoded to three categories. It is fairly obvious how this can be done. This includes Belgium, Spain, France, Italy, Portugal and United Kingdom. Combined with other equivalence restrictions in Table 3.2, the result is a linked data matrix. Figure 3.3 indicates equivalence assumptions by a pink colour.

Under equivalence, most walking items can be linked to each other (c.f. Figure 3.1). Some of the target items are still isolated though. For example, the Swedish item BRISK is not connected to any other item. The consequence of this is that, without any additional bridge items or bridge studies, it will not be possible to convert the Swedish walking disability into a common scale.

3.6 Obtaining data

Application of the method depends on the availability of two types of data. First, we need microdata (i.e. data at a person level) from bridge studies in order to be able to construct a conversion key. The conversion key is subsequently used to estimate the amount of disability in each MS. For this, we need a second type of data, prevalence data, for example in the form of response category frequencies by sex and age.

We approached statistical offices of the MS and investigators of bridge studies with a request for data. If a MS measured only item, we asked for a table of response frequencies, split according to age and sex. For bridge studies and for MS that administered more than one item on walking disability, we asked for the microdata on these items, or alternatively, for a multidimensional contingency tables of the items, also split according to age and sex.

It here became apparent that the documentation on which we based Table 3.1 was inaccurate. For example, our Danish contact assured us that items AFF1KM, AFFS100M and AFF1100M were never sampled in Denmark, while our documentation indicated otherwise. The same held for all German items. Other documentation errors

NAME		NL02	CAN0	E01	NL04	NL06	NL07	NL01	EURO	FIN01
fsidiff	H	Y	I	I					I	
moveins	H	I	I	Y					I	
room	H	I	Y	I					y	
adl8	I				I			Y	I	
gars9	I				Y			I	I	
outdoors	I				I			I	Y	
w400a	A		Y							
w400b	A					Y	Y	I	Y	I
w400c	A					I	I	Y	I	I
w400d	A					I	I	I	I	Y
ppt7						Y	Y			
gars7	G				Y			I		
adl5	G				I			Y		
haq8		Y			Y					

Figure 3.2 Linkage diagram with equivalence assumption added (in pink, coded I).

we found were a reference to the wrong institute ("Sorry, we have never heard of this survey") and discrepancies in the item identification numbers, resulting in the fact that we got the wrong items. Some offices never responded to our requests. Practical difficulties like these are of little help in speeding up the data collection process.

We managed to get appropriate prevalence information from Austria, Belgium, Denmark, Finland, Italy, Netherlands, United Kingdom, Norway, Switzerland and New Zealand. Estimation of the amount of disability on a common scale is thus only possible for these countries. In addition, we could obtain data from 14 bridge studies, mainly from the United Kingdom and the Netherlands.

3.7 Construction of the linked data set

The data thus obtained were combined into one data set for further analysis. Data were organised in the same structure as the linkage diagram in Figure 3.1, and all response categories were consistently recoded into the same direction, with zero indicating the category with the least disability. The total number of observations in the data set was equal to 141730. Table 3.3 presents a breakdown of the number of number observations by study, and indicates whether the study acted as bridge study, as prevalence study, or as both.

Item scores that are assumed to be equivalent were combined into a common column, labeled with one of the block names A to N. In this way, the number of different items to analyse reduces to 31. Observe that only records with scores on at least two items will contribute to the conversion key. For reasons of efficiency, we therefore selected records with at least two item scores. The total number of records available for the conversion key construction was thus 21487.

Table 3.3 Obtained linkage data set for walking disability.

Study	Country	Bridge study	Prevalence data	Frequency
A01	Austria		y	6085
B01	Belgium			6466
CH01	Switzerland		y	13004
DK01	Denmark		y	1081
EUR01	Various	y		2585
FIN01	Finland		y	7227
I01	Italy		y	62461
N01	Norway		y	1726
NL01	Netherlands	y	y	1790
NL02	Netherlands	y		38
NL03	Netherlands	y		306
NL04	Netherlands	y		292
NL06	Netherlands	y		50
NL07	Netherlands	y		30
NL09	Netherlands	y		4006
UK01	UK	y	y	19788
UK02	UK	y		11158
UK03	UK	y		1426
UK04	UK	y		301
UK05	UK	y		966
UK06	UK	y		681
UK07	UK	y		263
Total				141730

3.8 Preliminary analysis

The actual derivation of the conversion key requires three model-fitting steps. A preliminary statistical analysis of 31 items was done in order to have a starting estimate of the conversion key. Next, the appropriateness of equivalence assumptions was assessed by means of specific sub-analyses. Finally, the final model is formulated and estimated. The parameters of the final model define the conversion key.

Figure 3.4 is the item threshold map of the items in the preliminary analysis as calculated by RUMM2010. The threshold map depicts the most probable category of each item as a function of the common latent trait. Each colour transition is located at the threshold. As explained in Chapter 2, thresholds are optimally chosen under the unidimensional 3-parameter model. Items are ordered according to their location parameter. Items on top (e.g. K (walking more slowly) or LIMMI (limited in walking a mile) are "easy" in the sense that the probability of responding into the upper disability categories of these items is high for low levels of disability. At the other extreme, we find the item STAN3 ("can you stand at all?"). This item is answered "yes" only if the walking disability is very high. Other items fall in between these extremes, and their ordering is quite logical. Note that being unable to walk inside is more severe than being unable to walk outside. The results are close to previous analyses using different statistical models and other data (Van Buuren & Hopman, 2001).

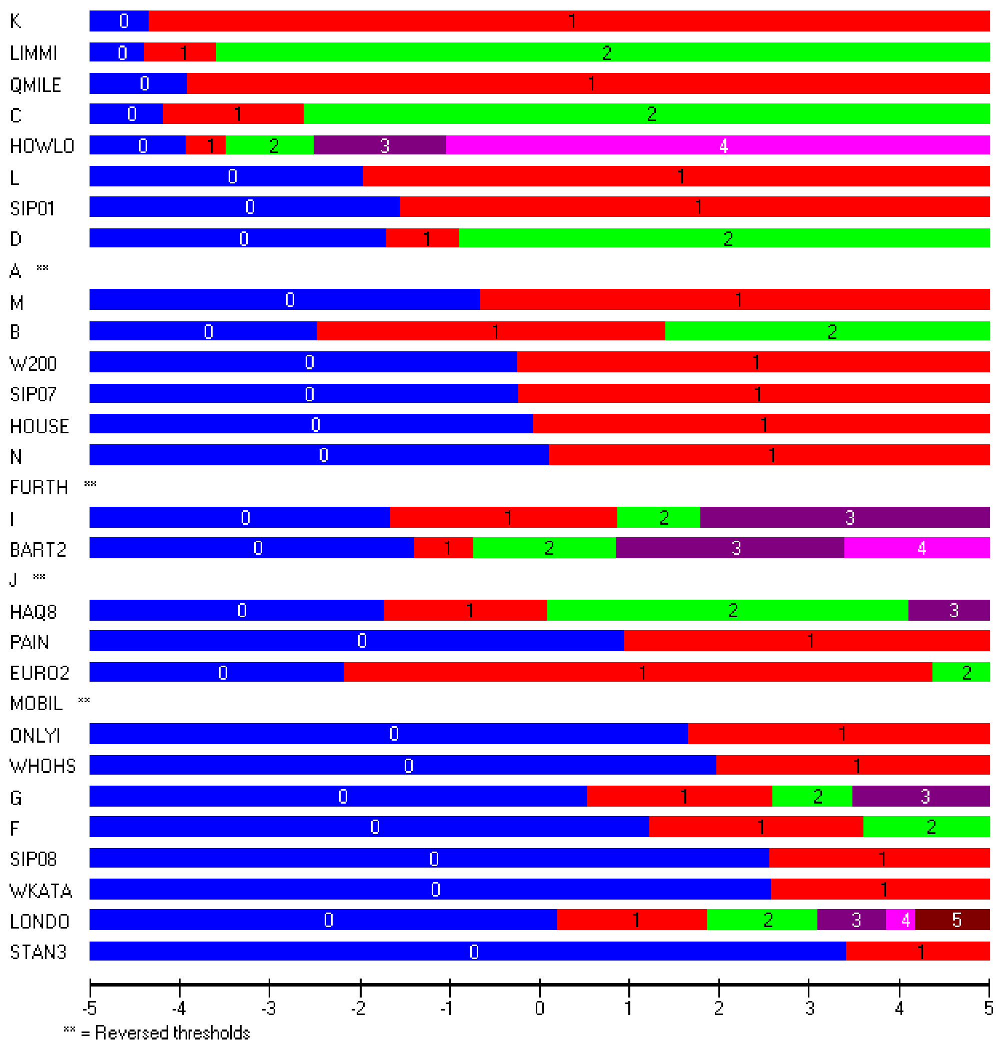


Figure 3.3 Item threshold map of walking disability item, ordered by level of disability.

Some items (e.g. block A) are not depicted in Figure 3.4. This is because the software does not plot the colour bar if thresholds are not strictly ordered. Figure 3.5 gives the Category Characteristic Curve (CCC) of block A. The threshold sequence (0-1, 2-3, 1-2) is not ordered. Note that category 2 is never the preferred category, and is always dominated by its direct neighbours. In classic test construction applications, threshold reversal is a sign that the item is possibly reversibly coded, or that categories should be taken together. In the present application, it is not yet clear what the consequences are. Threshold reversal is often associated with bad fit. On the other hand, there is nothing in the model that says that threshold reversal may not occur. The precise consequence of threshold reversal within the context of response conversion is something that needs further study.

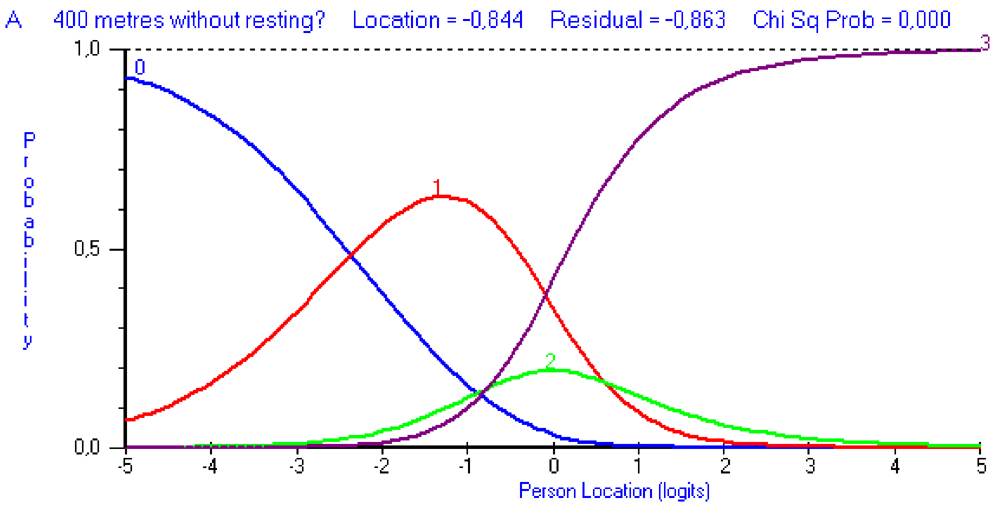


Figure 3.4 Category characteristic curve for item block A with reversed thresholds (Walk 400 meters without resting?), indicating the probability of responding in each category for a given level of disability.

3.9 Checking equivalence assumptions

A number of equivalence assumptions were made in Section 3.5. These assumptions were necessary to get a linked data matrix. These assumptions can be assessed to some extent. Assuming that two items are equivalent implies that the CCC's of both items are the identical. Thus, equivalence means that the relation between the latent trait θ and the response probabilities is the same for both items. The problem of potentially unequal CCC's is known as *differential item functioning* (DIF) (Holland & Wainer, 1982). DIF is a politically sensitive topic since an ability test with DIF-items discriminates on characteristics other than pure ability.

How can we investigate DIF? If we would have a way to see whether the CCC's of the items within an equivalence block are in fact different, then we can investigate the appropriateness of the equivalence assumption. A simple way to do this is to add the original block variables to the analysis, refit the model, and inspect the threshold plot.

Figure 3.6 presents the threshold map for item block B. The first four rows are the original items, while the common item is located at the bottom row. The grey bars indicate the threshold points of the common item. Item FAR1 is more like the common item B than RFAR2 (Note: The RFAR2 item is the FAR2 item, but recoded in an obvious way to get three categories). The maximum difference between the grey bar

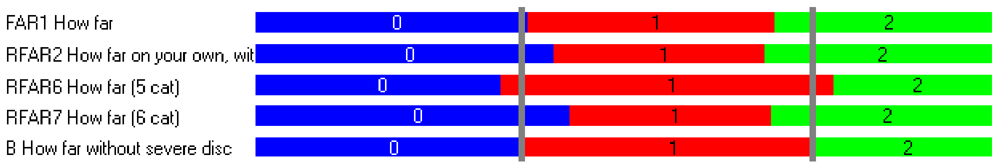


Figure 3.5 Threshold map of block B. The grey bars indicate the difference between the thresholds of the item block B, and the individual items from which block B was constructed assuming equivalence.

Table 3.4 Tension coefficient for assessing the equivalence assumption. Larger values indicate items for which equivalence is more questionable.

Block	Item	Label	Tension
B	FAR1	How far	0,15
B	RFAR2	How far on your own, with aid	0,32
B	RFAR6	How far (5 cat)	0,07
B	RFAR7	How far (6 cat)	0,39
L	RST1	How long standing without seve	0,04
L	RST10	Can you stand for ten minutes?	0,67
K	SLOWL	Do you walk more slowly than a	0,00
K	RSI12	I walk more slowly (SIP)	4,58
C	RLIMH	Does health limit w half a mil	0,46
C	RA1KM	More than 1 km affected by cur	0,21
M	RSI11	Use walking frame, crutches, s	3,65
M	RAI1	Aids either inside or outside	3,64
M	RAI4	Uses walking aids	0,10
M	RAI3	Do you use any of the followin	0,13
J	RWAOD	Can you walk outdoors?	0,09
J	RMANA	Manage to go outdoors & walk d	0,49
N	RHOUT	Need help to walk outside	0,00
N	RAIM5	Unable to walk unless assisted	0,03
G	RADL5	Move towards another room on t	0,10
G	RROOM	Trouble moving from one room t	0,95
I	RADL8	Move along outside the house?	0,14
I	ROUTD	Fully ind walk outdoors (if n	0,03
I	RGAR9	Move Outdoors	1,08
A	W400B	Walk 400metres without resting	0,55
A	W400C	Walk 400m without resting (wit	0,67

and the thresholds of the original item thresholds is located at RFAR7.

We computed a *tension coefficient* for each equivalenced item as the squared difference between the thresholds of the individual and common item, averaged over the number of thresholds. The tension coefficient expresses how bad the item fits the equivalence assumption. Table 3.4 contains all tension coefficients.

Sensible cut-point for tension coefficients are not known. Three items with large tension coefficients clearly stand out from the rest: RSI12 (derived from SIP12) in block K, and

RSI11 (derived from SIP11) and RAI1 (derived from AIDS1) in block M. Since the equivalence assumption is problematic for these items, one could try to refit the model without these making assumptions. It is indeed possible to compute such models, but the solution becomes unstable. Note that the link between the block A (Can you walk 400 meters) and B (How far) passes through blocks K, L and M. Breaking up blocks K and M leaves only block L as the central link. This is now the only connection between the block A and B. This puts a very large weight upon the L-block. The solution with broken K- and M-blocks is more difficult to interpret than the preliminary analysis. Major changes occur in the location of the F-block (Walking indoors). The location of 0-1 threshold shifted to the right from 1.3 to 3.0, and the location of the 1-2 threshold shifts from 3.7 to 7.2. Both are too much to the right to be realistic.

Tension coefficients could not be estimates for some items (DWELLING, W400D). The reason for this is that they are located at a "dead end" of the linkage structure. Consequently, there is no comparative information available to which these items can be linked. It is thus not possible to study the appropriateness of the equivalence assumptions in these cases.

This section has shown in what way the correctness of the equivalence assumptions can be assessed. In the case that the assumption is untenable, the appropriate response is to break the block and refit the model with the individual items. In the present case, it was not possible to do the latter for the most devious items because that would weaken the linkage structure too much. The cure would be worse than the problem. This demonstrates that investigations of the equivalence assumption are limited by the linked data structure, and underwrites the need for many links, as short as possible. In the present case, we would have been considerably helped if we had a study containing items from both the A-and B-blocks.

3.10 A conversion key for walking disability

A conversion key consists of a collection of the threshold values. The RUMM software estimates these threshold values under a given statistical model. The precise model we use is a compromise between the models of Sections 3.8 and 3.9. Items with a tension value of lower than unity are represented by their common item. Thus a conversion key of "How far"-items will be given as the thresholds of the common B-block, and only this block is fitted. Items with tension values of one or more, that is, items for which the equivalence assumption does not seem to work very well, are represented by the threshold values of the item itself. These items are fitted together with their block item.

Table 3.5 is the resulting conversion key of items for measuring walking disability. The number of different items is equal to 48. The conversion key is close to the item map of Figure 3.4. Some items (SIP12, AIDS1, AIDS2, GARS9, MANSHE) were fitted separately to account for questionable equivalence.

Table 3.5 Conversion key for walking disability

Block	Item(s)	Label	Threshold				
			0-1	1-2	2-3	3-4	4-5
K	SLOWLY	Walk more slowly as	-4,40				
	LIMMILE	Does health limit walk	-4,37	-3,68			
	QMILELEV	Difficulty ¼ mile in	-3,95				
C	LIMHMIL, AFFS1KM	Limited to walk 1 km	-4,19	-2,67			
	STAND1	How long remain stand	-3,98	-3,53	-2,57	-1,08	
K	SIP12	I walk more slowly	-2,66				
M	AIDS1	Aids either inside or	-2,48				
L	STAND10M, STAND1	Stand for ten minutes	-2,00				
	SIP01	I walk shorter distanc	-1,46				
D	LIM100Y, AFFS100M	Health limits walk 100	-1,81	-1,01			
A	W400B, W400C	400 metres without res	-2,48	0,53	-0,81		
M	AIDS3, AIDS4	Uses walking aids	-0,71				
B	FAR1, FAR2, FAR6, FAR7	How far without severe	-2,53	1,36			
	W200WS	Walk 200yards without	-0,30				
	HOUSE	Get in and out house u	-0,18				
	SIP07	Walk by self but with	-0,13				
N	HELPOUT, AIMS5	Need help outside?	0,14				
	FURTHEST	On level what is the f	-0,04	-1,08	-0,35	2,15	
I	GARS9	Move Outdoors	-3,00	1,27	2,85		
I	ADL8, OUTDOORS	Moving outside without	-1,78	0,81	2,09		
	BART	BARTHEL ambulation	-1,48	-0,86	0,90	3,78	
J	WALKOUTD, MANAGE	Can you walk outdoors	0,04	0,95	0,81		
	HAQ8	Outdoors on flat groun	-1,78	-0,01	4,49		
M	SIP11	Use walking frame, cru	0,93				
	PAIN	I'm in pain when I wal	1,00				
	EURO	EUROQOL mobility	-2,30	4,84			
	MOBIL	Mobility	0,15	-1,52	0,77	7,03	
	ONLYIND	I can only walk about	1,73				
F	WALKINDO	Walk indoors without h	1,17	3,56			
	WHOHS	Who helps to get aroun	2,47				
G	ADL5, GARS7	Get around in the hous	0,41	2,94	4,16		
	SIP08	Only walk with help	2,61				
	WKATA	Walk at all?	2,66				
	LONDO	Does your health stop	0,24	1,97	3,24	4,05	4,40
G	MANHSE	Manage same floor	1,26	3,77	4,00		
	STAND3	Can you stand at all	3,37				

Assuming a lognormal prior with mean = 0 and log(sd) = 0.5, we can compute the average of the posterior disability distribution for each category. Table 3.6 lists the mean ability per category for some items. SI01, HAQ8 and GARS9 have been included in order to allow comparison to Table 2.4. The thresholds are quite similar to those found earlier, especially at the lower levels of disability.

Table 3.6 Mean disability levels per category on the common scale for some walking items.

	Response category			
	0	1	2	3
SI01	-2.60	-0.94		
HAQ8	-2.73	-1.71	0.18	2.89
GARS9	-2.88	-1.84	0.28	2.64
B	-2.81	-1.68	0.86	
A	-2.85	-2.02	-0.84	0.83
F	-2.22	0.14	2.85	

3.11 Expressing prevalence data on a common scale

Walking disability estimates on a national level can be calculated in two steps. First, replace each score by the average disability estimates per category (from Table 3.6). For the states given in Table 3.1, blocks A, B and F will be needed. Second, take the average over groups of interest (e.g. specific sex and age classes). If desired, survey weights can be incorporated into this calculation.

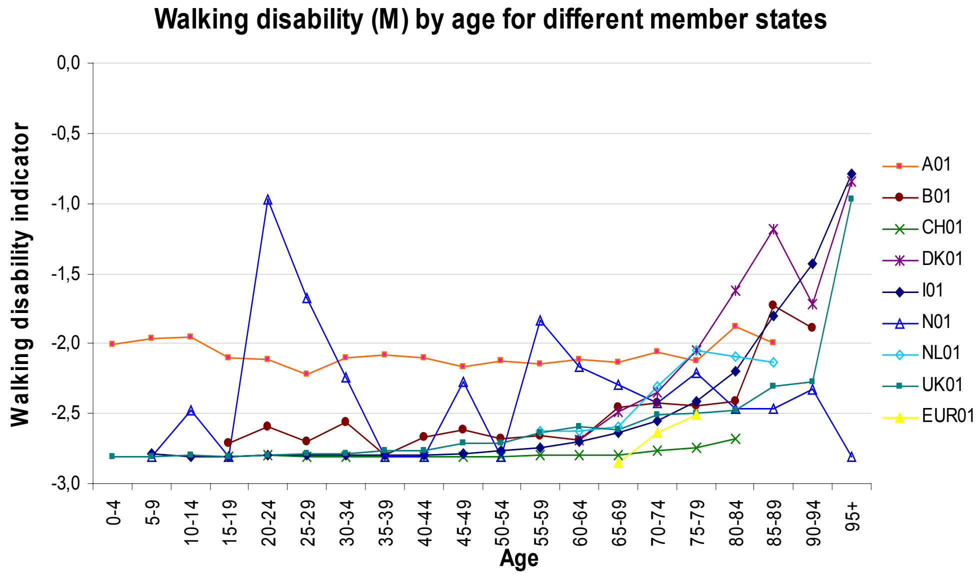
Appendix C gives the number of observations and the mean disability by age and sex for a number of countries. Estimates for Finland are absent from these tables because we did not have the appropriate ages. The mean disability estimates for Finland are -2.51 for males (n=1202) and –2.43 for females (n=1462), which positions Finland near the European average.

Figure 3.7 provides a graphic representation of the outcomes. Not all items were sampled at all ages, and the lengths of the curves vary. As expected, disability generally increases with age in almost all cases. The Norwegian curve (N01) appears very irregular as a result of small samples. The Austrian curve (A01) is peculiar in the sense that it is high and shows little trend. Swiss walking disability (CH01), on the other hand, is very low both for males and females. Both the Danish (DK01) and Dutch (NL01) curves appear somewhat higher than average. As both studies are based on "400 meters" item, this could raise suspicion about any systematic bias in the conversion key. Observe however that the low position of the EUR01 study, which posed a "400 meters" question to a mix of European countries, does not really support this.

3.12 Conclusion

This chapter applied the principles outlined in Chapter 2 to a realistic case for measuring walking disability in different member states of the EC. The technology can be used to put dissimilar items on a common scale using a set of explicit and verifiable steps.

The conversion key can be used to convert new data on a common scale. We feel that this is an important advance over current practice. The current conversion key as in Table 3.5 should be considered as the end point, since its construction relies on data that happened to be available. Much more can be done to tune and validate the new key. We will return to this topic in Chapter 5.



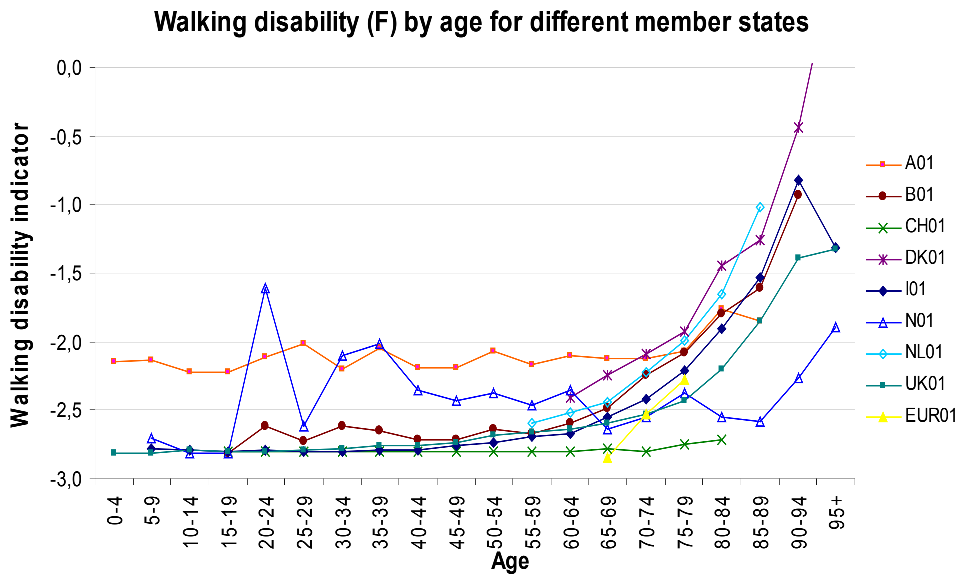


Figure 3.6 Mean walking disability for European countries, expressed on a common scale, by sex and age.

4 Dressing disability

4.1 Type of disability

Chapter 4 deals with instruments for measuring dressing disability, one of the forms of a personal care disability. In terms of the ICIDH-D, personal care disability refers to "an individual's ability to look after himself to basic physiological activities, such as excretion and feeding, and to caring for himself, such as hygiene and dressing" (WHO, 1993). Dressing disability is described by ICIDH-D codes 35 and 36. Code 35 includes all clothing disabilities except footwear. This includes activities like putting on skirts, trousers, jackets, blouses, shirts, night-dresses, overalls, smocks and overcoats, and doing up buttons, hooks and zips. Code 36 includes other dressing disabilities like putting on socks and stockings and shoes, tying shoelaces, putting on gloves, helmets, cosmetics, jewellery, and so on. Though not strictly part of the ICIDH-D classification, items that refer to undressing are also included. Some items were included that refers jointly to bathing (code 33) and dressing.

Ample variation between countries exists with respect to the wording of the questions and the formulation of response categories. Appendix D is the complete list of dressing items, including all items found in bridge studies. The total number of different items is equal to 56.

Table 4.1 Target items for measuring dressing disability in the European Community.

Country	Study	Survey	Dressing
Austria	A01	Microcensus Survey on Disabilities	WASHDRES
Belgium	B01	Health Interview Survey	DIFOWN6
Denmark		- no items found -	
Germany		- no items found -	
Spain	E01	Disability, Impairment & Health Status survey	CAN1
	E02	Spanish Health Interview Survey	DIFOWN4
Finland	FIN01	Health care survey	DIFF11
France	F01	National Disability Interview	DIFF4
	F02	National Health Interview Survey	DIFOWN5
Greece		- no items found -	
Ireland		- no items found -	
Italy	I01	Italian Survey on health conditions	DIFOWN8
Luxembourg		- no items found -	
Netherlands	NL01	Health Interview Survey	ADL4
Portugal	P01	Health Interview Survey	DIFF13
	P02		DIFF8
Sweden		- no items found -	
UK	UK01	Health Survey for England	DIFOWN8
		NON-EC member states	
Norway	N01	Health interview survey (Helseundersokelsen)	DIFF14
Canada	CAN01	Hals	PCAR2, DIFF1
New Zealand	NZ01	1996 Household Disability Survey	PCARE1, DIFF6

	BLOCK	E01	UK08	EURO1	E02	P01	P02	N01	NZ01	UK03	UK01	F01	B01	I01	CAN01	FIN01	UK07	A01	UK06	UK04	UK05	NL02	UK02	NL03	NL06	NL07	NL01	NL04	F02	F03
CAN1	M	Y	L																											
CAN3	M	L	Y																											
DIFF3	B			Y	L	L	L	L	L	L																				
DIFF6	B			L	L	L	L	L	Y	L																				
DIFF7	B			L	Y	L	L	L	L	L																				
DIFF8	B			L	L	L	Y	L	L	L																				
DIFF13	B			L	L	Y	L	L	L	L																				
DIFF14	B			L	L	L	L	Y	L	L																				
DIFF15	B			L	L	L	L	L	L	Y																				
DIFFOWN2	D										L	L	L	L							Y									
DIFFOWN4	D										L	Y	L	L							L									
DIFFOWN6	D										L	L	Y	L							L									
DIFFOWN8	D										Y	L	L	Y							L									
PCARE2	G								L						Y															
PCARE1	G								Y						L															
DIFF1	A														Y	L	L													
DIFF4	A														L	Y	L													
DIFF11	A														L	L	Y													
WASHDRES	H									L							L	Y	L	L										
EUROSC	H									Y							Y	L	Y	L										
LIMWDRES	H									L							L	L	L	Y										
HELP1	F																L		L	L	Y	L	Y							
HELP2	F																L		L	Y	Y	L	L							
HELP3	F																Y		Y	L	L	L	L							
HELP4	F																Y		L	L	L	L	L							
DIFFOWN1																				Y	Y									
AIDDRES																				Y	Y									
AIDSHOES																				Y	Y									
AIDUNDRES																					Y									
HELPSHOES	L																			Y	Y	L								
FSIHP10	L																			L	L	Y								
SIP29	K																			L	L	Y		Y						
DIFFSHOES	K																			Y	Y	L		L						
AIMSDEX3																						Y								
AIMSDEX2	I																					Y								
FSIDIF8	I																					Y								
FSIHP8	J																					Y								
FSIPN8																						Y								
FSIHP7																						Y								
FSIPN7																						Y								
FSIDIF7																						Y								
FSIDIF10																						Y								
FSIPN10																						Y								
SIP34																						Y		Y						
SIP31	J																					Y		Y						
SIP35																						Y		Y						
HAQ01	C																					Y	L	Y	L	L	L	Y		
DIFF10	C																					L	Y	L	L	L	L	L		
OECDDRES	C																					L	L	L	Y	Y	L	L		
ADL4	C																					L	L	L	L	L	Y	L		
GARS1	E																											Y	L	L
DIFFOWN5	E																											L	Y	L
DIFFOWN7	E																											L	L	Y
HELPTIME																			Y											
PPT4																									Y	Y				

Figure 4.1 Linkage diagram of dressing items

4.2 Linkage diagram and equivalence assumptions

Bridge studies and bridge items were identified from the same sources as cited in Chapter 3. Figure 4.1 is the linkage of dressing items. The 'Y'-symbol indicates that the specific item-study combination exists, while the 'L'-symbol indicates that items are linked by equivalence assumptions. The B-block has three categories. Some of the items within the block have four categories, with separate possibilities for "only with help" and "cannot". These most extreme categories were combined during construction of the B-block.

There is more than one way of specifying equivalence assumptions as there are many subtle differences in item formulations. We've chosen to be rather restrictive, so that small variations in wordings lead to allocations to different blocks. The diagram makes clear that under this treatment not all items can be linked. Most problems occur in the

Table 4.2 Number of observations in the linkage data set for dressing disability.

Study	Country	Bridge study	Prevalence data	Frequency
A01	Austria		y	6087
EUR01	Various	y		2585
I01	Italy		y	62461
N01	Norway		y	1726
NL01	Netherlands	y	y	1978
NL02	Netherlands	y		38
NL03	Netherlands	y		306
NL04	Netherlands	y		292
NL06	Netherlands	y		50
NL07	Netherlands	y		30
UK01	UK	y	y	19788
UK02	UK	y		11158
UK03	UK	y		1426
UK04	UK	y		301
UK05	UK	y		966
UK06	UK	y		681
UK07	UK	y		263
TOTAL				110136

upper left corner of the diagram. Block M, with target item CAN1, cannot be linked at all. Block B is linked to other blocks only through blocks G and H, and block D is linked only through studies UK05 and F01.

4.3 Linked data set

Where appropriate, prevalence data and bridging information were requested from the statistical offices of the MS and investigators of bridge studies. Similar data collection problems as for walking disability occurred. We had access to prevalence information of Austria, Italy, Netherlands, New Zealand, Norway, United Kingdom, and to 11 bridge studies. Table 4.2 gives a breakdown of the number of cases in the linked data set. Like before, only records were selected for key construction that contained at least two valid responses. The total number of records for key construction equalled 4693.

4.4 Conversion key

A first round of the preliminary analysis found an extreme threshold estimate for block D (around +8). This is a sign that the solution for that block is unstable due to thin linkage. It was therefore decided to delete this block from further analyses. The disadvantage of this is that no conversion key can be computed for DIFFOWN8 in block D. DIFFOWN8 is the item that is used in the health surveys in Italy and United Kingdom. Prevalence estimates for these MS will therefore be based on the key values of block B, the block most closely related to block D. The linkage map in Figure 4.2 is the solution of the analysis.

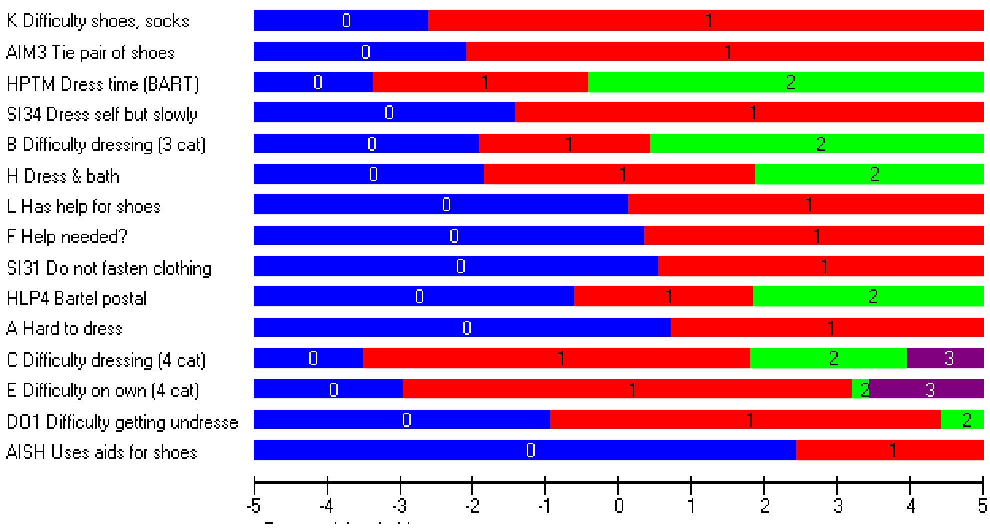


Figure 4.2 – Threshold map of dressing items.

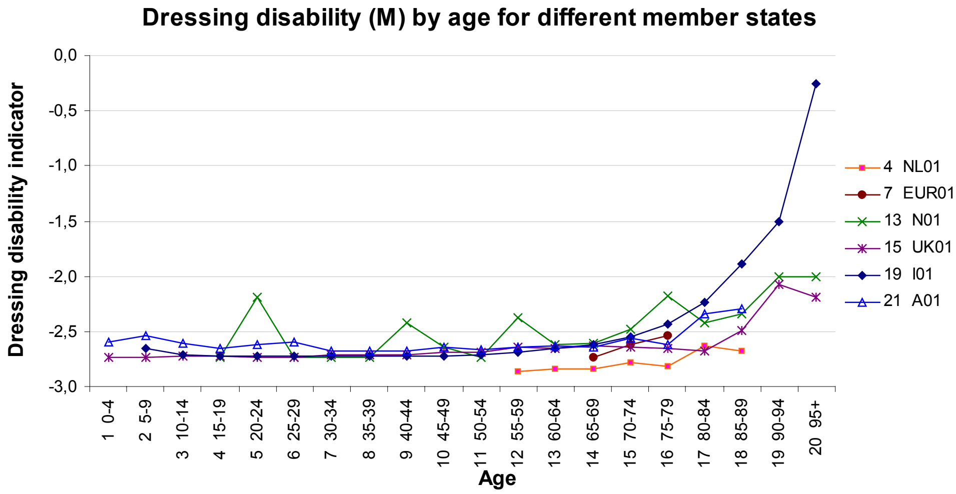
An interesting comparison is that between blocks B and C, which are responses to the same question but having a different number of categories. The peak of category 1 in both items is similar at approximately –0.8, but the category is longer for the C-block. This is somewhat counter-intuitive because category 2 ("much difficulty") in the 4 category item C commences later than category 2 ("can't, only with help") in the 3 category item B. One could expect that category 3 ("only with help") of C would similar to category 2 ("can't/only with help") of B, but it appears that more disability is needed to respond the most extreme category in the 4-category items than in the most extreme category of the 3-category item.

The assessment of equivalence assumptions by tension coefficients was hampered by some analysis problems. The software informed us that combinations of extreme items were found, and that therefore no analyses could be done. Removal of the offending item combination resulted in similar messages for other item-combinations, at which point the equivalence analysis was abandoned. These problems are likely to be related to the thin linkage. It seems that we might be asking just too much from the data.

The conversion key was based on the solution of Figure 4.2. Table 4.3 contains the mean disability estimates that can be used to compute dressing disability prevalence on the common scale.

Table 4.3 Mean dressing disability per response category on the common scale.

Item	Response category			
	0	1	2	3
B	-2.73	-1.64	0.51	
C	-2.94	-1.87	0.69	3.18
H	-2.69	-1.40	1.29	



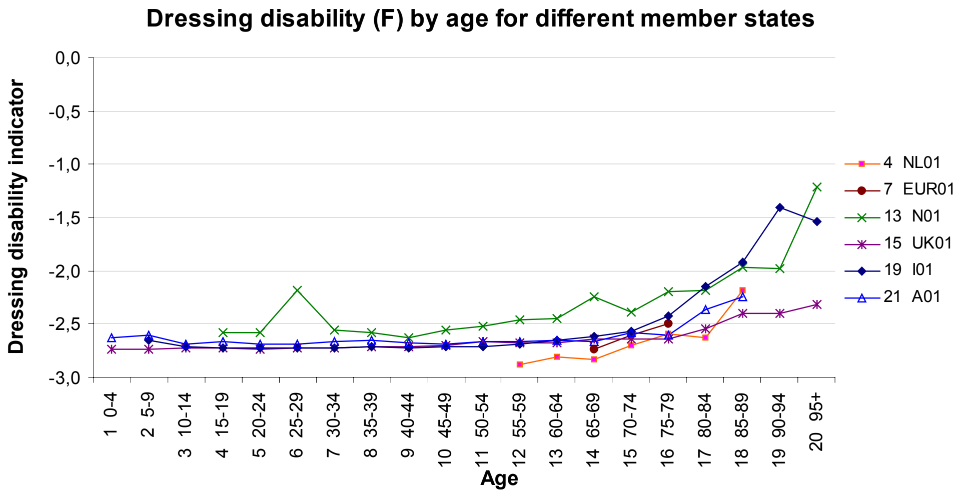


Figure 4.3 – Mean dressing disability for European countries, expressed on a common scale, by sex and age.

As we could not check the equivalence assumptions, the estimates in Table 4.3 and the results in Figure 4.3 have a provisional character. Also, we noted that relatively small changes in the analysis could have substantial effects on the thresholds, and thus on the mean disability estimates per response category. Both phenomena are a direct consequence of thin linkage.

5 Conclusion

This project aims to develop and demonstrate a new post-harmonisation technology, called response conversion. This methodology makes it possible to convert existing health information into community indicators. The goals of the pilot project were:
- to demonstrate the response conversion methodology on a practical problem,
- to identify key problems, if any.

Chapter 2 described the comparability problem in detail, and outlined the principles of the method. Chapter 3 and 4 described its practical application to walking and dressing disability. Chapters 2 to 4 thus cover the first goal. This chapter addresses the second goal, and draws conclusion and recommendations for further application of the method.

5.1 Evaluation

We were not confronted with major problems that made the application of response conversion impossible. We were able to produce comparative values for walking and dressing disability. It became clear that the role of the linkage structure is critical. Some conditions needed to be fulfilled in order to be able to apply the method. Most of these had to do with the linkage structure and the available data. The text below discusses these in more detail.

5.1.1 Linkage

Response conversion only works if the items of interest are linked. Unlinked items cannot be placed on a common scale. This was for example the case with the Swedish walking disability question BRISK. Items can be linked by data (i.e. by bridge studies and bridge items), by assuming equivalence, or by a combination of both. It will be clear that linkage by data is preferable over the use of equivalence assumptions, as the latter are, by definition, not backed up by data.

Walking and dressing disabilities are frequently measured type of disabilities. The problem of constructing a linkage matrix for walking and dressing disability may therefore be somewhat easier than for other types of disabilities. On the other hand, there are probably also more different varieties of walking and dressing items than for most other disabilities, which complicates the linkage. Further application of the method will tell us how well response conversion can be used for other types of health information.

We were confronted with some surprises regarding the documentation on which we had based the initial linkage structure. In some cases, the documentation was in clearly in error. Some items had wrong numbers, some items were not actually sampled (but according the documentation they were), and in one instance an incorrect institute was mentioned. These findings threatened to break the linkage structure, since some items that were central in the linkage never existed. Response conversion is thus sensitive to the quality of the documentation that is used to construct the linkage. The associated risks would probably diminish if the source documentation were subjected to some form of independent quality control.

5.1.2 *Data acquisition*

In order to make our method work, we needed prevalence data, bridge items and bridge studies. One of the side-conditions of the project was to use only existing information. No new information was to be sampled. Obtaining prevalence data turned out not to be so easy because statistical offices of the MS differ in their data sharing policies. Notwithstanding several reminders, some offices did not respond at all to our requests for prevalence data.

Bridge studies and bridge items were primarily taken from studies conducted in the UK and The Netherlands. Apart from some exceptions, obtaining microdata from these studies was relatively easy as most of them were distributed from public data archives, and could be acquired at nominal costs.

It will be clear that access to data is paramount to the success of any statistical method. The lesson to be learned from these experiences is that some organisation is needed that facilitates the exchange of information between member states. Getting prevalence information took a disproportional amount of effort and time. Routine application of our technology is hardly possible if appropriate prevalence information is not already available in a central place, or at least can be obtained in a timely matter.

One option worth studying is the possibility to collect new data. The new data could function as a bridge study. This makes the construction of the linkage matrix more controllable, as its entries will not depend anymore on what items happened to be available from previous studies. Such dedicated bridge studies need not be very large or costly, and will lead to a more compact and workable linkage matrices for any health parameter.

5.1.3 *Unidimensionality*

The selection of items to be taken up in the linkage was guided by specific ICIDH-D categories. Throughout the report, we implicitly made the assumption that walking items measure walking disability, and that dressing items measure dressing disability. The construction of the group of items was based on face validity. All items seem to measure some aspect of walking or dressing disability.

The property that items measure the same trait is formally known as *unidimensionality*. Unidimensionality can be defined in various ways, and there are several approaches to actually check unidimensionality in a given set of data (c.f. Hattie, 1985). Unidimensionality of items is an important property because it is a prerequisite in the model that we used. We only marginally addressed this topic in order not to divert from the main message, and because it is technically complex for linked data. The primary danger of not properly accounting for unidimensionality is that some of the linkage items may not measure walking or dressing ability. Using such items for linking could yield conversion keys that regress towards the middle. The item threshold maps indicate no systematic traces of such a phenomenon, but a more complete analysis would also include steps to verify unidimensionality.

An alternative to unidimensionality is to change the model to a regression type of model, where responses on one item is predicted from those of one or more other items and covariates. The hard part in this approach is to impute (i.e. to fill in) the missing parts in the linkage matrix in such a way that the structure among the items in the completed data is maintained. This can be done by a form of multivariate imputation (Schafer, 1997; Van Buuren & Oudshoorn, 2000). In principle, anything can be part of the linkage matrix, but it is still necessary to have linkage. The unidimensionality requirement is effectively replaced by a conditional independence requirement on the connecting items, which is generally weaker. The downside of this is that the conversion key will become more complex.

5.2 New technical problems

Item response theory has traditionally been applied within the field of psychological testing and education. The objective in those fields is individual measurement of ability. The number of items is typically much larger (say 5 to 50) than in the present application. In response conversion, disability estimates may depend on as few as one item. This introduces some new problems in the estimation of ability. We adopted a Bayesian estimator of ability under an informative prior. More work is needed to verify whether this choice is optimal.

Traditional model fitting procedures are often based on homogeneous ability groups. In the present application, this often leads to empty homogeneous groups, which complicates the interpretation of the fit statistics. Model fitting is a somewhat circular activity because ability estimates are derived from items, and item fit is derived from ability groups. This circularity becomes a problem as the number of items becomes smaller. Also, thin linkage plays a role in assessing fit. We deleted the worst fitting items from the final walking solution. Item fit was measured by the *t*-statistic, with a value of 3 or larger indicating a bad fit (Wright & Masters, 1982). It appeared that three items had to be removed: HOUSE, QMILE and HAQ8. As expected, the resulting solution fitted better, but at the same time, we observed a considerable and implausible upward drift of the block A ("400 meters"). Again, thin linkage puts a limitation on what can be achieved in terms of model fitting.

Chapter 3 introduced a tension statistic for measuring the difference between the common and individual items. This index is a bit crude and the cut-off point is arbitrary. It is quite likely that better alternatives for describing the similarity between items as a function of thresholds exist. Using such properties, better decision rules in assumption tests can be developed.

In Section 2.2.3 we found that the difference as measured by SIP01 and GARS9 was smaller than the difference as measured on the common HAQ8 item (0.13 versus 0.39). This discrepancy is a result of overfitting. The model is fitted on essentially HAQ8, and applied to SIP01 and GARS9. There may thus be a regression-to-the-mean effect, which may dilute real differences. This effect is stronger if the number of categories is small, if measurement error is large, and if linkage becomes thin. It is not yet clear whether the effect would be absent in a complete data matrix.

The linkage structure plays a crucial role in response conversion. It would be useful to have diagnostics that measure the quality of the linkage. For example, shorter paths between items are generally better, having multiple possible connecting paths between two items is better, and a good fit of bridge items is preferable.

In this report, we placed items onto a common scale. Another interesting use of the technique is to express the information collected with one item into the scale of the other item. For example, using the conversion key, one can express the French disability data in the German response system, and back. Such *item-to-item conversion* gives an answer to the question: What would the data have been like if we had used the German item in the French population (after translation of course)? If we require that the item should be translated forward and back, and produce the same result as the original, then we need to account for any translation errors that might occur during the process. We have some limited experience with this, but the method has not yet fully worked out.

It would also be interest to have an idea of the uncertainty in threshold and ability estimates. The statistical framework allows for estimates of uncertainty, for example as 95% confidence intervals. In this way, the effect of any translation and measurement errors could be assessed.

5.3 Conclusion

Response conversion as develop here has some advantages:

- it makes the incomparability a tangible concept;
- it works on existing data, without the need to sample new data;
- assumptions can be tested to some extent;
- it builds on a well-established mathematical framework;
- it yields a common scale with interval scale properties;
- any aspect of the disability distribution can be studied;
- construction of the conversion key can be separated from its application.

The operational work to create and apply conversion keys can be split into a number of logical compartments. Investigations of the linkage structure can be done using only meta-documentation and without access to actual data. After the basic layout of the linkage structure is known, acquisition of microdata and the construction of the linked data can be done by trained staff. The construction of the conversion key is a typical activity for a statistician working from the linked microdata. Key construction is separate from the application of the conversion key, which can be done by anyone. Such division of labour is critical in any large-scale application of our method.

Each new field of health parameters requires a separate conversion key. In general, construction of the conversion key is expensive, but needs to be done only once. Application of the key to new prevalence data is cheap and straightforward.

We distinguished between pre-harmonisation and post-harmonisation. Response conversion is a post-harmonisation technique that is useful for translating existing data. It can however also be useful in pre-harmonisation. Response conversion can be used to repair any trend gaps that are inevitable when a new measure replaces an older one. This could smooth the transition to the new measurement system.

This pilot project demonstrates that the existing disability data collected in different Member States can be placed onto a common scale. This is of value to Health Monitoring Program of DG-SANCO since the method advances the prospects of a working European system for health monitoring. As always, there are still some problems that need to be worked out, but we trust that this pilot will contribute to a better understanding of the validity and usefulness of response conversion.

5.4 Follow-up

Further work within the context of the HMP is being planned. We think that it is useful to disseminate and apply response conversion within the HMP. One could think of the following activities:

1. Evaluation of the suitability of response conversion for projects and data within the HMP;
2. Construction of new conversion keys;
3. Development of an interactive web site for actual conversion to community indicators;
4. Integration of RC into the IDA-HIEMS monitoring system.

New conversion keys can be made for indicators that are of particular interest to the HMP. The ECHI-indicator list (ECHI working group, 2000) contains a quality indicator, coded as categories *a* through *d*. The meaning of each category is as follows:

a indicators based on data regularly available from international sources (e.g. causes of death; European Community Household Panel); the indicators are conceptually clear, valid and reliable; improving comparability may still be needed.

b indicators based on data regularly available from national sources (e.g. national health interview surveys, hospital data); also here, the indicators are conceptually clear, valid and reliable; improving comparability between countries is usually a major issue.

c indicators that have to rely on incidental national sources (e.g. surveys on specific topics or target groups); these indicators may be conceptually clear, valid and reliable, efforts have to be made to make these regularly available within Member States' information systems; clarifying definitions and establishing comparability between countries is a major issue.

d indicators or topics on which data are needed but generally not available; here an R&D trajectory is neede, including concept development, data collection logistics, indicator definition, etc. It is advisable to undertake such activities at the EU level.

Response conversion is probably most useful for indicator in categories *b* and *c*, where comparability is a major issue. Indicators in category *d* are typical candidates for pre-harmonisation. The EUPASS project on physical activity (Rütten, 2001) is an example where response conversion is likely to be straightforward and useful. This project collected data on both old and new indicators, which is very valuable since the study can now act as a bridge study for itself. This eases the construction of the linkage diagram, and thereby the construction of conversion keys.

References

1. ANDRICH D. Rasch models for measurement. Newbury Park: Sage, 1988.
2. BOCK RD, MISLEVY RJ. Adaptive EAP estimation of ability in a microcomputer environment. Applied psychological measurement 1982, 12, 261-280.
3. BOOMSMA A, VAN DUIJN MAJ, SNIJDERS TA (Eds.). Essays on Item Response Theory. Berlin: Springer Verlag, 2000.
4. ECHI working group. Design for a set of European Community Health Indicators. Luxembourg: European Commission, 2000.
5. EUROPEAN COMMISSION, Programme of community action on health monitoring. Work programme 1998-1999. Art. 5.2.b of Decision 1400/97/EC. Luxembourg: European Commission, 1998.
6. FISCHER GH, MOLENAAR IW (Eds.). Rasch models: foundations, recent developments, and applications. Berlin: Springer Verlag, 1995.
7. HAMBLETON RK, SWAMINATHAN H, ROGERS JH. Fundamentals of item response theory. Newbury Park: Sage, 1991.
8. HATTIE J. Methodology review: Assessing unidimensionality of tests and items. Applied Psychological Measurement 1985, 9, 139-164.
9. HOPMAN-ROCK M, VAN BUUREN S, DE KLEIJN-DE VRANKRIJKER M. Polytomous Rasch analysis as a tool in the revision of the severity of disability scale of the ICIDH. Disability and Rehabilitation 2000, 22, 363-371.
10. HUPKENS C. Coverage of health topics by surveys in the European Union. Luxembourg: Eurostat, 1998.
11. KOLEN MJ, BRENNAN RL. Test equating: Methods and practices. New York: Springer, 1995.
12. MURRAY CJL, TANDON A, SALOMON J, MATHERS CD. Enhancing cross-population comparability of survey results. GPE Discussion paper 35, Geneva: World Health Organization, 2000.
13. 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.
14. RASMUSSEN N, GUDEX C, CHRISTENSEN S. Survey data on disability. Report 3/1999/E/20. Luxembourg: Eurostat, 1999.
15. ROBINE J-M, JAGGER C, EGIDI V. Selection of a coherent set of health indicators. A first step towards a user's guide to health expectancies for the European Union. Montpellier: Euro-REVES, 2000.
16. RUMM LABORATORIES. RUMM 2010 Rasch Unidimensional Measurement Models. http://www.rummlab.com.au/. 2001.
17. RÜTTEN A. European Physical Activity Surveillance System (EUPASS). Final report. European Commission, Luxembourg, 2001.
18. SADANA R, MATHERS CD, LOPEZ AD, MURRAY CJL, IBURG KM. Comparative analyses of more than 50 household surveys on health status. GPE Discussion Paper Series: No.15., Geneva: World Health Organization, 2000.
19. SCHAFER JL. Analysis of incomplete multivariate data. Chichester: Chapman & Hall, 1997.
20. 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.
21. VALE CD. Linking item parameters onto a common scale. Applied Psychological Measurement 1986, 10, 333-344.

22. VAN BUUREN S, HOPMAN-ROCK M, MIEDEMA HS. The development of a proposal for revision of the severity of disabilities scale of the ICIDH. Technical report TNO/PG 96.024. Leiden: TNO Prevention and Health,1996.
23. VAN BUUREN S, OUDSHOORN CGM. Flexible multivariate imputation by MICE. TNO-rapport PG 99.054. TNO Prevention and Health, Leiden, 1999.
24. VAN BUUREN S, HOPMAN-ROCK M. Revision of the ICIDH Severity of Disabilities Scale by data linking and item response theory. Statistics in Medicine 2001, 20, 1061-76.
25. VAN DER LINDEN WJ & HAMBLETON RK (Eds). Handbook of modern item response theory. New York: Springer, 1997.
26. WHO. International Classification of Impairments, Disabilities and Handicaps. Geneva: World Health Organisation, 1980. Reprinted 1993.
27. WRIGHT BD, MASTERS GN. Rating scale analysis: Rasch measurement. Chicago, MESA Press, 1982.

A List of studies

Study	Bridge	Name	Sample	Sample size	Country	Year	Organisation
A01		Microcensus Survey on Disabilities	All, includes institutions:few questions	Ca. 60,000	Austria	June 1995 (every 5 years)	Austrian Central statistical office
B01		Belgian Health survey	60 and over and 15 and over screened	10,000 persons (4500 households)	Belgium	1997	IHE = Institut d'Hygiene et d'Epidemiologie
CAN01	Y	Hals			Canada	1986 & 1991 surveys	
CH01		Swiss health survey	ages 15-74 in households	16000 households	Swiss	1997	Bundesamt für Statistik
D01		Health Interview Survey	18-80 years	7500	Germany	1997	
DK01		Danish Health & morbidity survey	>16 yrs	6000	Denmark	1987, 1994, 1999	Danish Institute for clinical epidemiology
E01		Disability, Impairment & Health Status survey	0-99 years	230000	Spain	1999	Nat institute of statistics.
E02		Spanish Health Interview Survey	all	8400	Spain	1995	Ministerio de Sanidad y Consumo
EUR01		Euronut Seneca	60-80		12 European	1988 - 1989	Seneca group
F01		National Disability Interview	All, includes institutions	35000	France	1998-2001	INSEE
F02		National Health Interview Survey	All, excludes institutions	21000	France	1970, 1980, 1991	INSEE
F03		French Health and medical care survey	65 and over	21,500 (8000 households)	France	1991/2	INSEE
FIN01		Health care survey	60+	4000 households	Finland	1995/96	Stakes, KELA, KTL, and Tilastokeskus
I01		Italian Survey on health conditions and recourse to health services	6+ (in households)	75000	Italy	1994 (every 4 years)	ISTAT
N01		Health interview survey (Helseundersokelsen)	all in households	14000	Norway	1995 (every 10 years)	Statistik sentralbyråå
NL01		Health Interview Survey	All, excludes institutions	10000	Netherlands	Annual	Statistics Netherlands
NL02	Y	Liang et al study	50-80	38	US	1990	
NL03	Y	ERGOPLUS	55-75	306	Netherlands	1991, 1993	
NL04	Y	EURODISS	Mean age 53.9	242	Netherlands	1994	
NL06	Y	GOW	75-85	50	Netherlands	1995	

Study	Bridge	Name	Sample	Sample size	Country	Year	Organisation
NL07	Y	DETER	75 and over	30	Netherlands	1994	
NL08	Y	VFSIPH		6000	Belgium	1992	
NL09	Y	CBS-GE Disability Survey	Excludes institutions	15000	Netherlands	1986-1988	Statistics Netherlands
NZ01		1996 Household Disability Survey	Adults	2712752	New Zealand	1996	Statistics New Zealand
P01		Health Interview Survey	Excludes institutions	50000	Portugal	1998/9 (every 3 years)	Ministry of health
P02		Portuguese Health Interview survey	10 and over	50,000 (17,400 households)	Portugal	1995/1996	Ministerio da Saude
S01		Swedish Living Conditions Interview Survey	all	7000	Sweden	1995 (annual)	SCB Statistiska Centralbyrån
UK01		Health Survey for England	>2 years, includes 2000 institutions	20000	England	1995, 2000	SCPR +UCL+Dept of Health
UK02	Y	OPCS Survey of Disabled Adults in Private Households	Adult		UK	1985	
UK03	Y	General Household Survey	>65		UK	1994	
UK04	Y	Major Trauma Study	1-92 years	304	UK	1995	
UK05	Y	Calderdale Health & Disablement Survey	16+ years		UK	1998	
UK06	Y	Continuing Care study	Adults	Approx 400	UK		
UK07	Y	Young Adult Team Project	18-26	Approx 200	UK		
UK08	Y	British Disability Interview Survey	16+	7000	UK	1996	Department of Health and ONS

B List of walking items

Item	Question	Coding/categories	Target	Block
adl5	Can you move towards another room on the same floor? (ADL for 55+)	0=yes no difficulty, 1=yes some difficulty, 3 = yes great difficulty, 4 = only with help.	Y	G
adl8	Can you move along outside the house?(ADL for 55+)	0=yes no difficulty, 1=yes some difficulty, 3 = yes great difficulty, 4 = only with help.	Y	I
aff1100m	Indicate whether (walk 100 metres) affected by current health state	0=not affected at all, 1=a little affected, 2=Very affected.	Y	D
aff1km	Indicate whether (walking more than 1 km) affected by current health state	0=not affected at all, 1=a little affected, 2=Very affected	Y	C
affs100m	Indicate whether (walk several 100 m) affected by current health state	0=not affected at all, 1=a little affected, 2=Very affected	Y	E
aids1	Do you use any walking aids either inside or outside the house?	0=N, 1=Y		M
aids2	Do you use, or do you need any equipment to help you move about, such as a walking stick or wheelchair?	0=N, 1=Y		M
aids3	Do you use any of the following walking aids? (Artificial limbs), (Calipers, surgical shoes or surgical corset, orthesis), (cane, crutch, walking frame, walking aids with wheels), (wheelchair), (invalid cariagge), (adapted bicyle), (adapted car)	0=N, 1=Y		M
aims5	Unable to walk unless assisted (by another person or cane, crutches, artificial limbs or braces)? (AIMS)	0=N, 1=Y		N
aims6	I'm unable to walk at all/Do not walk at all (AIMS)	0=N, 1=Y		-
bart	Ambulation (barthel item)	0=fully independent, 1=minimal help, 2=moderate help needed, 3=heavily dependent, 4=unable		-
brisk	Can you take a short walk, say five minutes, at a fairly brisk pace ?	0=yes, 1=no	Y	-
diff	Do you have any difficulty in walking?	0=N (if in doubt, choose Y), 1=Y.		-
difout	Health problems cause difficulty walking outside?	0=N, 1=Y.	Y	-
dwelling	Walking up & down in the dwelling?	0=yes possible without help, 1=yes possible with help, 2=not possible.	Y	F
euro	Mobility (EUROQoL)	0=confined to bed, 1=some problems walking, 2=no problems walking.		-
far1	How far can you walk without stopping/experiencing severe discomfort?	0=200 m or more. 1=More than a few steps but less than 200m, 2=A few steps only,	Y	B
far2	How far can you walk without stopping/experiencing severe discomfort, on your own, with aid if normally used?	0=200yds or more. 1=more than a few steps but less than 200yds, 2=a few steps only, 3=can't walk,	Y	B

Item	Question	Coding/categories	Target	Block
far3	How far can you walk without stopping/experiencing severe discomfort? (walk with/without aids/uses wheelchair etc)	in meters	Y	B
far4	How far can you walk without stopping/experiencing severe discomfort on your own?	0=200 m or more. 1=More than a few steps but less than 200m, 2=A few steps only.	Y	B
far5	How far can you walk without stopping/experiencing severe discomfort on level ground?	0=can't walk. 1=can't walk but uses wheelchair, 2=200 m or more, 3=More than a few steps but less than 200m, 4=A few steps only,	Y	B
far6	How far can you walk without stopping/experiencing severe discomfort?	0=A few steps only, 1=More than a few steps but less than 50y, 2=more than 50y but less than 200yds, 3=more than 200yds but less than 1/4 mile, 4=1/4 mile or more		B
far7	How far can you walk without stopping/experiencing severe discomfort?	0=1/4 mile or more. 1=more than 200yds but less than 1/4 mile, 2=more than 50y but less than 200yds, 3=More than a few steps but less than 50y, 4=A few steps only, 5=can't walk at all,		B
far8	How far can you walk without stopping/experiencing severe discomfort?	0=more than 200yds. 1=more than 50y but less than 200yds, 2=more than a few steps but less than 50y, 3=A few steps only,		B
far9	What is the furthest he/she can walk on his/her own, without stopping and without getting too much tired ?	0=200 m or more. 1=More than a few steps but less than 200m. 2=A few steps only	Y	B
far10	How far can you walk without stopping and without any discomfort ?	0=200 m or more. 1=More than a few steps but less than 200m. 2=A few steps only, 3=cannot walk unaided,	Y	B
far11	How far can you walk without stopping/experiencing severe discomfort on your own?	0=200 m or more. 1=More than a few steps but less than 200m, 2=A few steps only, 3=cannot walk unaided,	Y	B
fsidiff	Difficulty walking inside? (FSI)	0=no difficulty, 1=mild difficulty, 2=moderate difficulty, 3=severe difficulty, 4=extreme difficulty.		H
fsihelp	Help to walk inside? (FSI)	0=no help, 1=cane, 2=someone elses help, 3=devices and someone elses help, 4=unable.		-
fsipain	Painful walking inside?(FSI)	0=no pain, 1=mild pain, 2=moderate pain, 3=severe pain, 4=extreme pain		-
furthest	On the level, what is the furthest you can walk at all? (even if you have to stop/have discomfort)	0=More than a few steps but less than 50yards, 1=A few steps only, 2= 50-200 yards,4 = 200 yds - 1/4 mile, 5 = 1/4 mile or more		-
gars7	Can you, fully independently, get around in the house (with cane if necessary)?	0=Yes, without any difficulty, 1=Yes, with some difficulty, 2=Yes, with much difficulty, 3=Only with help from others,		G
gars9	Can you fully independently walk outdoors (if necessary, with a cane)?	0=Yes, without any difficulty, 1=Yes, with some difficulty, 2=Yes, with much difficulty, 4 = Only with help from others.		I
haq8	Able to walk outdoors on flat ground?(HAQ)	0=without any difficulty, 1=some difficulty, 2=great difficulty, 3=unable.		-

Item	Question	Coding/categories	Target	Block
helpout	I need help to walk outside (walking aid or someone to support me)	0=N, 1=Y		N
house	Can you get in & out of your house without help?	0=Y, 1=N.		-
howlong	How long can you walk on your own without stopping/experiencing significant discomfort?	0=more than 30 mins. 1=15-30 mins, 2=10-15 mins, 3=5-10 mins, 4=less than 5 mins,		-
lim100y	Does your health limit you in walking 100 yards? If so how much?	0= No not limited at all, 1=Yes limited a little, 2=Yes limited a lot.		D
lim1bl	Walking one block	0= No not limited at all, 1=Yes limited a little, 2=Yes limited a lot.	Y	D
limhmil	Does your health limit you in walking half a mile? If so how much?	0= No not limited at all, 1=Yes limited a little, 2=Yes limited a lot.		C
limmile	Does your health limit you in walking more than a mile? If so how much?	0= No not limited at all, 1=Yes limited a little, 2=Yes limited a lot.		-
limsevb	Walking along several blocks	0= No not limited at all, 1=Yes limited a little, 2=Yes limited a lot	Y	E
london	Does your health stop you from getting around? (London Handicap)	0=Not at all, 1=Very slightly, 2=Quite a lot, 3=Very much, 4=Almost completely, 5=Completely.		-
manage	Do you usually manage to go outdoors & walk down the road…..	0=on your own, 1=only with help, 2=not at all.		J
manhse				G
mobil	Mobility (aroung hospital bay or in home)	0=independent, 1=independent in electric/self prop chair, 2=walk with assistance, 3=attendant operated wheelchair, 4=bedbound,		-
mobind				-
moveins	Difficulty moving around inside your home?	0=N, 1=Y		H
onlyind	I can only walk about indoors	0=N, 1=Y		-
outdoors	Move Outdoors	0=yes no difficulty, 1=yes difficulty no help, 3 = yes only with help, 4 = no		I
pain	I'm in pain when I walk	0=N, 1=Y		-
ppt7	Walk 15 meters (time test)	0=can't walk, 1= >25sec, 2=20.5-25 sec, 2=15.5-20 sec, 3= <=15 sec.		-
qmilehil	Do you have difficulty walking for a quarter of a mile if its uphill and downhill?	0=N, 1=Y,		-
qmilelev	Do you have difficulty walking for a quarter of a mile on the level?	0=N, 1=Y, 2=Can't walk at all		-
qmilews	Walk quarter of a mile on your own without stopping/severe discomfort, with aid if normally used?	0=Y, 1=N, 2=Can't walk at all.		-
room	Do you have any trouble moving from one room to another?	0=no trouble, (if yes are you completely unable?), 1=yes trouble	Y	H
rugby	Can you walk the distance around a rugby field without resting (that is 350 metres or 400 yards)?	0=easily, 1=with difficulty, 2=not at all	Y	-
sip01	I walk shorter distances or often stop for a rest (SIP)	0=N, 1=Y		-
sip07	I walk by myself but with some difficulty (eg; limp, stumble, wobble or have a stiff leg) (SIP)	0=N, 1=Y		-

Item	Question	Coding/categories	Target	Block
sip08	I only walk with help from someone else(SIP)	0=N, 1=Y		-
sip11	I get about only by using a walking frame, crutches, stick, walls or hold onto furniture (SIP)	0=N, 1=Y		M
sip12	I walk more slowly (SIP)	0=N, 1=Y		K
slowly	Do you walk more slowly or at the same pace as somebody else your age who is in good health?	0=same pace. 1=more slowly,		K
stand1	How long can you stay standing without severe discomfort?	0=more than 30 mins.1=15-30 mins, 3=10-15 mins, 3= 5-10 mins, 4=1- 5 mins, 5=less than 1 min,		L
stand10m	Can you stand for ten minutes?	0=Can do so without any problems. 1=Can do so but experiences tiredness or pain. 2=Cannot do so,		L
stand2	How long can you remain standing without having to move around ?	0= 30 minutes or more. 1= 10 but less than 30 minutes, 2=Less than 10 minutes,		-
stand3	If can't walk, can you stand at all?	0=Y. 1=N.		-
stand30m	Can you stand for half an hour?	0=Can do so without any problems 1=Can do so but experiences tiredness or pain, 2=Cannot do so.		-
w200ws	Walk 200 yards or more without stopping/discomfort, with aid if normally used?	0=Y. 1=N		-
w400a	Do you have any trouble walking 400 yards/metres without resting?	0=no trouble, (if yes are you completely unable?), 1=yes trouble.	Y	A
w400b	Can you walk 400metres without resting?	0=yes no difficulty, 1=yes minor difficulty, 2=yes major difficulty, 3=no.		A
w400c	Can you walk 400 yards/metres without resting (with walking stick if necessary)?	0=yes no difficulty, 1=yes minor difficulty, 2=yes major difficulty, 3=no,	Y	A
w400d	Can you walk 400 yards/metres without resting (with walking stick if necessary)?	0=yes no difficulty, 1=yes minor difficulty, 2=yes major difficulty, 3=no.	Y	A
Walk10m	Can you walk for ten minutes without stopping?	0=Without help of walking aids, 1=Without help from others, but with a walking aid, 2=Only with help from others, 3=Not at all		-
Walk30m	Can you walk for half an hour without stopping?	0=Not at all, 1=Only with help from others, 2=Without help from others, but with a walking aid, 3=Without help of walking aids		-
Walkindo	Can you walk indoors?	0=Without help of walking aids, 1=Without help from others, but with a walking aid, 2=Only with help from others, 3=Not at all.		F
Walkoutd	Can you walk outdoors?	0=Without help of walking aids, 1=Without help from others, but with a walking aid, 2=Only with help from others, 3=Not at all.		J
Whohse				-

Item	Question	Coding/categories	Target	Block
Wkatall	Can you walk at all?	0=Y, 1=N		-

C Walking disability on a common scale

Walking disability prevalence estimation, Male, counts

AGE5	NL01	EUR01	B01	N01	UK01	CH01	I01	DK01	A01	Total
0-4					449				58	507
5-9				4	672		1344		68	2088
10-14				11	638		1881		74	2604
15-19			11	6	503	354	2173		65	3112
20-24			17	2	549	331	2385		49	3333
25-29			34	1	632	681	2288		72	3708
30-34			53	2	763	668	2267		144	3897
35-39			66	2	746	579	2153		144	3690
40-44			65	5	640	576	2188		173	3647
45-49			66	11	624	463	2151		191	3506
50-54			93	6	559	432	1906		269	3265
55-59	191		93	11	498	350	1852		267	3262
60-64	176		159	7	502	382	1624	109	249	3208
65-69	146	5	145	16	470	289	1477	130	232	2910
70-74	137	952	166	21	451	298	1118	97	142	3382
75-79	88	305	90	24	276	178	490	74	158	1683
80-84	44		50	82	160	179	471	40	81	1107
85-89	8		21	44	63		157	18	76	387
90-94			4	14	18		37	4		77
95+				3	2		6	1		12
Total	790	1262	1133	272	9215	5760	27968	473	2512	49385

Walking disability prevalence estimation, Female, counts

AGE5	NL01	EUR01	B01	N01	UK01	CH01	I01	DK01	A01	Total
0-4					445				33	478
5-9				11	676		1209		63	1959
10-14				13	616		1650		51	2330
15-19			22	3	563	374	1987		72	3021
20-24			39	4	637	353	2413		50	3496
25-29			67	6	829	733	2424		62	4121
30-34			68	10	909	785	2413		115	4300
35-39			90	6	780	688	2282		147	3993
40-44			100	10	722	666	2182		150	3830
45-49			111	18	730	533	2169		176	3737
50-54			101	29	650	530	1977		271	3558
55-59	209		125	23	583	445	1909		260	3554
60-64	168		166	18	537	506	1688	147	286	3516
65-69	190	3	216	40	562	469	1735	128	268	3611
70-74	182	968	187	45	498	456	1425	133	292	4186
75-79	120	314	122	44	364	352	693	89	278	2376
80-84	84		83	145	283	354	712	67	210	1938
85-89	47		52	90	142		292	32	190	845
90-94			21	30	39		78	11		179
95+				4	8		19	1		32
Total	1000	1285	1570	549	10573	7244	29257	608	2974	55060

Walking disability prevalence estimation, Male, mean disability on common scale

AGE5	NL01	EUR01	B01	N01	UK01	CH01	I01	DK01	A01	Total
0-4					-2,81				-2,00	-2,72
5-9				-2,81	-2,81		-2,79		-1,96	-2,77
10-14				-2,48	-2,80		-2,81		-1,95	-2,78
15-19			-2,71	-2,81	-2,81	-2,81	-2,80		-2,11	-2,79
20-24			-2,59	-0,98	-2,80	-2,80	-2,80		-2,12	-2,79
25-29			-2,70	-1,68	-2,78	-2,81	-2,80		-2,22	-2,79
30-34			-2,56	-2,25	-2,78	-2,81	-2,80		-2,10	-2,77
35-39			-2,79	-2,81	-2,77	-2,81	-2,80		-2,08	-2,77
40-44			-2,67	-2,81	-2,77	-2,80	-2,80		-2,10	-2,76
45-49			-2,61	-2,27	-2,71	-2,81	-2,79		-2,17	-2,74
50-54			-2,68	-2,81	-2,71	-2,81	-2,77		-2,13	-2,71
55-59	-2,63		-2,66	-1,83	-2,64	-2,80	-2,74		-2,14	-2,67
60-64	-2,62		-2,69	-2,16	-2,59	-2,79	-2,70	-2,69	-2,11	-2,64
65-69	-2,59	-2,85	-2,45	-2,30	-2,62	-2,79	-2,64	-2,49	-2,13	-2,59
70-74	-2,31	-2,63	-2,43	-2,42	-2,51	-2,77	-2,55	-2,35	-2,06	-2,54
75-79	-2,05	-2,51	-2,44	-2,21	-2,49	-2,74	-2,41	-2,06	-2,13	-2,41
80-84	-2,10		-2,41	-2,47	-2,48	-2,68	-2,20	-1,63	-1,88	-2,30
85-89	-2,14		-1,73	-2,47	-2,31		-1,81	-1,19	-1,99	-1,97
90-94			-1,89	-2,33	-2,28		-1,43	-1,72		-1,83
95+				-2,81	-0,98		-0,79	-0,84		-1,33
Total	-2,47	-2,60	-2,56	-2,40	-2,72	-2,80	-2,74	-2,31	-2,10	-2,69

Walking disability prevalence estimation, Female, mean disability on common scale

AGE5	NL01	EUR01	B01	N01	UK01	CH01	I01	DK01	A01	Total
0-4					-2,81				-2,15	-2,76
5-9				-2,71	-2,81		-2,78		-2,14	-2,77
10-14				-2,81	-2,79		-2,79		-2,22	-2,78
15-19			-2,81	-2,81	-2,80	-2,80	-2,80		-2,22	-2,79
20-24			-2,62	-1,61	-2,80	-2,80	-2,80		-2,12	-2,78
25-29			-2,72	-2,62	-2,79	-2,80	-2,80		-2,02	-2,79
30-34			-2,62	-2,10	-2,78	-2,81	-2,80		-2,20	-2,78
35-39			-2,65	-2,01	-2,76	-2,81	-2,79		-2,05	-2,75
40-44			-2,72	-2,36	-2,76	-2,80	-2,79		-2,19	-2,76
45-49			-2,72	-2,43	-2,74	-2,80	-2,76		-2,19	-2,73
50-54			-2,64	-2,37	-2,69	-2,80	-2,74		-2,07	-2,68
55-59	-2,59		-2,67	-2,47	-2,66	-2,80	-2,69		-2,17	-2,65
60-64	-2,52		-2,59	-2,36	-2,64	-2,81	-2,67	-2,41	-2,10	-2,62
65-69	-2,44	-2,85	-2,48	-2,64	-2,59	-2,78	-2,56	-2,25	-2,13	-2,54
70-74	-2,22	-2,53	-2,24	-2,55	-2,52	-2,81	-2,42	-2,10	-2,12	-2,45
75-79	-1,99	-2,27	-2,08	-2,38	-2,43	-2,75	-2,21	-1,93	-2,07	-2,29
80-84	-1,65		-1,80	-2,55	-2,20	-2,71	-1,91	-1,44	-1,77	-2,10
85-89	-1,01		-1,61	-2,58	-1,85		-1,53	-1,26	-1,86	-1,73
90-94			-0,93	-2,26	-1,39		-0,83	-0,43		-1,18
95+				-1,89	-1,33		-1,31	0,83		-1,32
Total	-2,26	-2,47	-2,44	-2,49	-2,69	-2,79	-2,69	-2,03	-2,08	-2,64

D List of dressing items

Item	Question	Coding/categories
ADL4	Can you dress and undress? (ADL4)	0=without difficulty, 1=with some difficulty, 2=with much difficulty, 3=only with help
AIDDRES	Do you use any aids to help you get dressed?	0=N, 1=Y, If Y, which ones?
AIDSHOES	Do you use any aids to help you put on shoes, socks or stockings ?	0=N, 1=Y, if Y, various codes
AIDUNDRES	Do you use any aids to help you get undressed?	0=N, 1=Y, If Y, which ones?
AIMSDEX2	Can you button articles of clothing?(AIMS)	0=Y, 1=N
AIMSDEX3	Can you easily tie a pair of shoes?(AIMS)	0=Y, 1=N
CAN1	Can you dress & undress yourself?	0=Y, 1=N
CAN3	Do you usually manage to dress and undress yourself	0=on your own, 1=only with help from someone, 2=not at all
DIFF1	Do you have any trouble dressing or undressing yourself? (HALS)	0=N, 1=Y
DIFF3	Do you have any trouble dressing or undressing yourself?	0=no difficulty, 1=difficulty no help, 3=only with help, 4=can't
DIFF4	Do you have any difficulty dressing or undressing?	0=N, 1=Y
DIFF6	Can you dress & undress yourself?	0=easily, 1= with difficulty, 2=not at all
DIFF7	Can you dress & undress yourself?	0=without difficulty, 1=with difficulty, 2=only with help, 3=no
DIFF8	Can you dress & undress yourself?	0=Without help/withoutdifficulty, 1=without help but with difficulty, 2=with help
DIFF10	How difficult is it for you to dress/undress yourself?	0=not difficult, 1=quite difficult, 2 = very difficult, 3 = impossible
DIFF11	I find it hard to dress myself	0=N, 1=Y
DIFF13	Can you dress & undress yourself?	0=on own no difficulty, 1=on own with difficulty, 2= only with help
DIFF14	Can you dress & undress yourself?	0=with no difficulty, 1=with some difficulty, 2=only with help from others
DIFF15	Do you usually manage to dress/undress yourself	0=on own easily, 1=on own with difficulty, 2=only with help, 3=not at all
DIFFOWN1	Do you have any difficulty getting undressed on your own?	0=N, 1=Y, 2=impossible
DIFFOWN2	Do you have any difficulty getting dresssed on your own?	0=N, 1=Y, 2=impossible
DIFFOWN4	Can you dress & undress entirely without help?	0=yes, no help/ no difficulty, 1=yes, no help/some difficulty, 2 = No, need help for everything
DIFFOWN5	Can you dress & undress without help?	0=yes without difficulty, 1=yes without too much difficulty, 2 = yes great difficulty, 3 = no
DIFFOWN6	Can you dress & undress without help?	0=yes no difficulty, 1=yes some difficulty, 2 = always need help

Item	Question	Coding/categories
DIFFOWN7	Can you get dressed & undressed on your own?	0=yes without difficulty, 1=yes without too much difficulty, 2 = yes great difficulty, 3 = no
DIFFOWN8	Can you get dressed & undressed on your own?	0=without difficulty, 1=some difficulty, 2=only with help
DIFFSHOE	Do you have any difficulty putting on shoes, socks or stockings on your own?	0=N, 1=Y, 2=Impossible
EUROSC	Self care (EuroQol – Washing & dressing together)	0=no self care problems, 1=some problems washing/dressing, 2=unable to dress self,
FSIDIF7	Difficulty putting on underpants? (FSI)	0=no difficulty, 1=mild difficulty, 2=moderate difficulty, 3=severe difficulty, 4=extreme difficulty
FSIDIF8	Difficulty buttoning clothes? (FSI)	0=no difficulty, 1=mild difficulty, 2=moderate difficulty, 3=severe difficulty, 4=extreme difficulty
FSIDIF10	Difficulty putting on shoes/slippers? (FSI)	0=no difficulty, 1=mild difficulty, 2=moderate difficulty, 3=severe difficulty, 4=extreme difficulty
FSIHP7	Help putting on underpants? (FSI)	0=no help, 1=cane,special equipment or other device, 2=someone elses help, 3=devices and someone elses help, 4=unable,
FSIHP8	Help buttoning clothes? (FSI)	0=no help, 1=cane,special equipment or other device, 2=someone elses help, 3=devices and someone elses help, 4=unable,
FSIHP10	Help putting on shoes/slippers? (FSI)	0=no help, 1=cane,special equipment or other device, 2=someone elses help, 3=devices and someone elses help, 4=unable,
FSIPN7	Pain putting on underpants? (FSI)	0=no pain, 1=mild pain, 2=moderate pain, 3=severe pain, 4=extreme pain
FSIPN8	Pain buttoning clothes? (FSI)	0=no pain, 1=mild pain, 2=moderate pain, 3=severe pain, 4=extreme pain
FSIPN10	Pain putting on shoes/slippers? (FSI)	0=no pain, 1=mild pain, 2=moderate pain, 3=severe pain, 4=extreme pain
GARS1	Are you able to dress and undress yourself without help?	0=without difficulty, 1=with some difficulty, 2=with much difficulty, 3=only with help
HAQ01	Are you able to dress yourself, including tying shoelaces /doing buttons/handling of clusures? (HAQ)	0=without difficulty, 1=some difficulty, 2=with much difficulty, 3=unable to do
HELP1	If difficult or impossible, do you need any help to dress/undress?	0=N, 1=Y
HELP2	Do you need any help to undress?(if difficulty/imposs on own)	0=N, 1=Y, if Y, various codes
HELP3	Dressing	0=fully independent, 1=minimal help, 2=moderate help needed, 3=heavily dependent, 4=unable
HELP4	Do you get dressed	0=without help, 1= just help with buttons, 2=with someone helping most of time

Item	Question	Coding/categories
HELPSHOE	Do you need any help to put on shoes, socks or stockings ? (if difficulty/imposs on own)	0=N, 1=Y, if Y, various codes
HELPTIME	Dressing	0=independent, 1=set up help only or incidental help for just one, 2=help from 1, takes <1/2 hr, 3=help from 1, takes>1/2 hr, 4=help from 2, takes <1/2 hr, 5=help from 2, takes >1/2 hr
LIMWDRES	Does your health limit you in bathing or dressing yourself? If so how much?	0=No not limited at all, 1=Yes limited a little, 2= Yes, limited a lot
OECDDRES	Are you able to dress and undress yourself? (OECD)	0=without difficulty, 1=with some difficulty, 2=with much difficulty, 3=only with help
PCARE1	Because of your condition, do you get help with Personal care, such as washing, grooming, dressing & feeding yourself?	0=N, 1=Y
PCARE2	Do you get help with personal care, such as washing and dressing & taking medication?	0=N, 1=Y
PPT4	Put on and off a coat (PPT4)	0=<=10 seconds, 1=10.5-15 seconds, 2=15.5-20 seconds, 3=>20 seconds, 4=can't
SIP29	I have trouble putting on shoes, socks or stockings (SIP)	0=N, 1=Y
SIP31	I do not fasten my clothing (eg I require help with buttons, zips or shoe laces) (SIP)	0=N, 1=Y
SIP34	I dress myself but do so very slowly (SIP)	0=N, 1=Y
SIP35	I only get dressed with someones help (SIP)	0=N, 1=Y
WASHDRES	Washing & dressing? (together)	0=yes possible without help, 1=yes possible with help, 2=not possible

E Dressing disability on a common scale

Dressing disability prevalence estimation, Male, counts

AGE5	NL01	EUR01	N01	UK01	I01	A01	Total
0-4				449		68	517
5-9				672	1343	60	2075
10-14				638	1889	78	2605
15-19			5	503	2175	64	2747
20-24			6	549	2388	52	2995
25-29			1	632	2288	66	2987
30-34			9	763	2274	137	3183
35-39			8	746	2158	152	3064
40-44			14	640	2189	188	3031
45-49			25	624	2151	178	2978
50-54			20	559	1910	248	2737
55-59	215		37	498	1858	290	2898
60-64	191		38	502	1637	219	2587
65-69	161	5	68	470	1484	263	2451
70-74	153	953	60	451	1132	145	2894
75-79	96	305	55	276	501	156	1389
80-84	53		208	160	490	93	1004
85-89	8		67	63	166	63	367
90-94			18	18	40		76
95+			3	2	7		12
Total	877	1263	642	9215	28080	2520	42597

Dressing disability prevalence estimation, Female, counts

AGE5	NL01	EUR01	N01	UK01	I01	A01	Total
0-4				445		61	506
5-9				676	1219	59	1954
10-14				616	1653	51	2320
15-19			7	563	1993	60	2623
20-24			7	637	2416	56	3116
25-29			10	829	2432	43	3314
30-34			19	909	2414	125	3467
35-39			15	780	2283	136	3214
40-44			22	722	2184	153	3081
45-49			50	730	2174	174	3128
50-54			62	650	1980	274	2966
55-59	226		44	583	1914	258	3025
60-64	192		57	537	1700	271	2757
65-69	209	3	76	562	1747	284	2881
70-74	191	968	102	498	1439	294	3492
75-79	136	314	81	364	705	275	1875
80-84	97		283	283	748	214	1625
85-89	50		147	142	309	186	834
90-94			39	39	88		166
95+			5	8	19		32
Total	1101	1285	1026	10573	29417	2974	46376

Dressing prevalence estimation, Male, mean disability on common scale

AGE5	NL01	EUR01	N01	UK01	I01	A01	Total
0-4				-2,73		-2,59	-2,71
5-9				-2,73	-2,65	-2,54	-2,67
10-14				-2,72	-2,72	-2,61	-2,71
15-19			-2,73	-2,72	-2,72	-2,65	-2,72
20-24			-2,19	-2,73	-2,72	-2,61	-2,72
25-29			-2,73	-2,73	-2,72	-2,59	-2,72
30-34			-2,73	-2,71	-2,72	-2,67	-2,72
35-39			-2,73	-2,71	-2,72	-2,67	-2,72
40-44			-2,42	-2,71	-2,72	-2,68	-2,71
45-49			-2,64	-2,69	-2,72	-2,64	-2,71
50-54			-2,73	-2,68	-2,71	-2,67	-2,70
55-59	-2,86		-2,38	-2,64	-2,69	-2,65	-2,69
60-64	-2,84		-2,62	-2,65	-2,65	-2,62	-2,66
65-69	-2,84	-2,73	-2,60	-2,62	-2,62	-2,64	-2,64
70-74	-2,79	-2,62	-2,48	-2,65	-2,55	-2,57	-2,60
75-79	-2,82	-2,54	-2,18	-2,65	-2,43	-2,61	-2,54
80-84	-2,63		-2,42	-2,68	-2,23	-2,34	-2,37
85-89	-2,67		-2,34	-2,49	-1,89	-2,30	-2,16
90-94			-2,01	-2,07	-1,51		-1,76
95+			-2,00	-2,19	-0,26		-1,02
Total	-2,82	-2,60	-2,44	-2,69	-2,68	-2,62	-2,67

Dressing disability prevalence estimation, Female, mean disability on common scale

AGE5	NL01	EUR01	N01	UK01	I01	A01	Total
0-4				-2,73		-2,63	-2,72
5-9				-2,73	-2,66	-2,60	-2,68
10-14				-2,73	-2,71	-2,69	-2,72
15-19			-2,57	-2,72	-2,72	-2,67	-2,72
20-24			-2,57	-2,73	-2,73	-2,69	-2,73
25-29			-2,19	-2,73	-2,73	-2,69	-2,72
30-34			-2,56	-2,73	-2,72	-2,67	-2,72
35-39			-2,58	-2,71	-2,72	-2,65	-2,71
40-44			-2,63	-2,71	-2,73	-2,67	-2,72
45-49			-2,56	-2,70	-2,72	-2,69	-2,71
50-54			-2,52	-2,67	-2,71	-2,67	-2,70
55-59	-2,88		-2,46	-2,67	-2,69	-2,66	-2,69
60-64	-2,81		-2,44	-2,68	-2,66	-2,65	-2,67
65-69	-2,83	-2,73	-2,25	-2,64	-2,62	-2,66	-2,63
70-74	-2,70	-2,60	-2,39	-2,64	-2,56	-2,58	-2,59
75-79	-2,60	-2,50	-2,20	-2,63	-2,42	-2,60	-2,50
80-84	-2,63		-2,18	-2,54	-2,15	-2,37	-2,28
85-89	-2,18		-1,96	-2,40	-1,92	-2,24	-2,10
90-94			-1,98	-2,40	-1,40		-1,77
95+			-1,21	-2,32	-1,53		-1,68
Total	-2,74	-2,57	-2,26	-2,69	-2,66	-2,60	-2,66