EUROPEAN JOURNAL OF PUBLIC HEALTH 2003; 13 (3 SUPPLEMENT): 15–19

Assessing comparability of dressing disability in different countries by response conversion

S. VAN BUUREN, S. EYRES, A. TENNANT, M. HOPMAN-ROCK *

Background: Comparability of health data is a major challenge within the context of the Health Monitoring Programme of the European Commission. A common problem in surveys is that many variations of essentially the same question exist. Methods: Response conversion is a new method for improving comparability by scaling the data onto a common scale. Comparisons between member states can then be made in terms of the common scale. A first step is the construction of a conversion key. This is a relatively complex activity, but needs to be done only once. The second step is the actual data transformation. This is simple, and can be repeatedly done on a routine basis as new information arrives. Construction of the key is only possible if enough overlapping information can be found. Results: The method is illustrated for dressing disability from five European countries. Differences occur between countries, between sexes and between age groups. These were similar in magnitude. Conclusion: Response conversion is a new method for enhancing comparability among existing data. Conversion can only be done if a key is available. More work is needed to establish the technique. Future implications within the Health Monitoring Programme are discussed.

Keywords: dressing disability, comparability, item response theory, response conversion, post-harmonization

The goal of the Health Monitoring Programme (HMP) of the European Commission (EC) is to provide relevant and timely information about the health in each member state.[1] As far as possible, the monitoring system will be fed by existing data collected through health surveys, or other sources that are performed by individual member states. Substantial variations in the actual measurement exist, e.g. in sampling procedures, in the coverage per topic, in the wording of questions and formulation of response categories. Incomparability of information is thus a major problem. See Harkness et al.[2] for a thorough overview of all problems related to the integration of cross-cultural surveys.

This paper addresses the problem created by differences in the formulation of survey questions and response categories. As an example, the UK health survey contains a question *Can you get dressed and undressed on your own?* with response categories 'without difficulty', 'some difficulty', 'only with help'. The Austrian health interview contains the item *Washing and dressing?* with response categories 'yes possible without help', 'yes possible with help' and 'not possible'. Both items obviously intend to measure the ability to take care of oneself, but it is not clear how an answer on the UK item can be compared with one on the Austrian item without making arbitrary recoding assumptions.

* S. van Buuren[1], S. Eyres[2], A. Tennant[2], M. Hopman-Rock[1]
1 TNO Prevention and Health, Leiden, The Netherlands
2 Psychometric Laboratory for Care Outcomes Measurement, University of Leeds, UK
Correspondence: Stef van Buuren, PhD, Department of Statistics, TNO Prevention and Health, P.O. Box 2215, 2301 CE Leiden, The Netherlands, Voice: +31 71 5181802, fax +31 71 5181920, e-mail: S.vanBuuren@pg.tno.nl

Response conversion (RC) is designed to enhance comparability in such cases. The method attempts to scale seemingly incomparable data in the same area onto a common scale. Where this can be done, comparisons between MS can be made in the common units. The goal of this paper is to demonstrate how RC can be applied in the context of post-harmonization of disability information.

METHOD

Conversion to a common unit is a common scientific activity. For example, the distance between two points can be measured in many ways: by a ruler, by the time taken to move between the points (e.g. sonar), by a shift in the electromagnetic spectrum (as in astronomy), by a difference between viewing angles, and so on. The resulting values (cm, seconds, colours, degrees) can be expressed in terms of a common distance unit provided that one knows how the observed data relate to the common unit.

RC applies this idea to survey measurement. Application of RC consists of two main steps. The first step is the construction of a conversion key. The conversion key relates the common scale to the observed data. The construction of the key is relatively complex, but needs to be done only once. The second step involves the conversion of the observed data into the common scale. This step is straightforward and can be repeated if new information arrives. Information on the common scale can be compared, for example across countries.

Table 1 contains data taken from Van Buuren and Hopman-Rock.[3] The rows represent three survey questions, labelled SI01, HAQ8 and GAR9. The columns denote two studies in which these were administered,

ERGOPLUS[4,5] and EURIDISS.[6] Both SI01 and GAR9 measure the ability to walk, but with only these two items, there is no way of comparing the amount of walking disability between the studies. However, both studies also administered the HAQ8 item, another walking disability item. The category frequencies on HAQ8 tell us that the EURIDISS sample is more disabled than the ERGOPLUS sample. Moreover, the link by HAQ8 allows one to relate the answers in SI01 and GAR9.

The data in *table 1* are described by a Partial Credit Model,[7] and the estimated parameters from this model form the conversion key. We will not describe the model in detail here, and the reader is referred to the report.[8] *Figure 1* describes the probability of responding in each category given the value of the common scale as fitted by RUMM 2010.[9] The points at which the curves cross are called thresholds. Knowledge of the location of the threshold is enough to reconstruct the curves. The collection of thresholds of all items forms the conversion key. Suppose that we have two *new* studies on different samples, where the first administers only item SI01 and the second administers only GAR9. Is it possible to compare the level of disability in the two new samples? Provided that an appropriate conversion key is available, the answer is yes. The problem of converting data to a common scale under a given model is known as ability estimation or scoring. We omit the technical details here and refer the reader to Van Buuren et al.[8] for more information.

Table 2 contains the mean disability per category in terms of the common scale. So, a response in category 'no' on item SI01 corresponds to a walking disability of –2.44 on the common scale, while a response 'yes' corresponds to an disability of –0.49. An estimate for the entire sample can be obtained by the sum of the disability levels weighted by the category frequencies. Thus, for ERGOPLUS we obtain ((–2.44 * 276) + (–0.49 * 28)) / 304 = –2.26. The mean disability in EURIDISS based on GAR9 is equal to –2.13, so the difference is 0.13. Thus, in the absence of a bridge item or bridge study, but with the aid of a conversion key, we can infer that the ERGOPLUS sample has on average fewer disabilities. Information contained in different items is now expressed

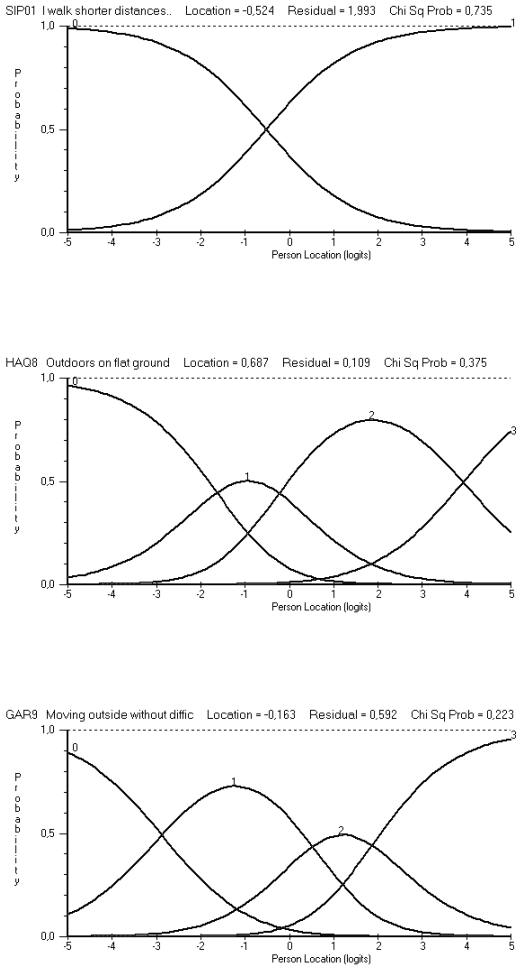

Figure 1 Category characteristic curves: probability of responding in a category in terms of the common scale

Table 2 Mean walking disability per category on the common scale

	Response category (c)			
Item	0	1	2	3
SI01	–2.44	–0.49		
HAQ8	–2.72	–1.71	0.06	2.68
GAR9	–2.89	–1.94	–0.22	2.00

Table 1 SI01 and GAR9 items linked by bridge item HAQ8

			Study	
Item	Description	Response categories	ERGOPLUS n=306	EURIDISS n=292
SI01	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
GAR9	Can you, fully independently, walk outdoors (if necessary, 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

on a common scale that can be used to compare the level of disability of the underlying samples.

RESULTS

Disability is one of the topics of the Health Monitoring Programme of the EC. Our illustration relates to the measurement of dressing disability. Dressing disability is described by ICIDH-D codes 35 and 36.[10] 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.

We obtained responses on dressing disability items that were posed in the health surveys of the five European countries listed in *table 3*. *Table 4* shows the frequency distributions split according to sex and age for ages between 55 and 89 years. All items measure the ability to dress, but do so in different ways. The British and the (translated) Italian items are similar in both question and response category. The Norwegian item uses similar categories, but does not use the phrase 'on your own'. The Austrian and Dutch items differ in both the questions and the response categories. The last column contains the mean disability estimate for each category on the common scale. This was an estimate derived from the conversion key.

The conversion key was constructed using a combined sample from 11 bridge studies. We used data from 4693 persons that answered at least two questions on dressing disability. This allowed us to construct a conversion key of 15 different items on dressing disability. We had not enough data available to separately estimate the difference caused by the phrase 'on your own'. The keys of these items were therefore equated to each other. For more details, we refer the reader to Van Buuren *et al*.[8]

Figures 2 and *3* contain the mean dressing disability for the five countries on the common scale. They display three kinds of effects: 'age effects' within each subplot, 'country effects' between the subplots, and 'sex effects' between the two figures. The Italian sample nicely illustrates the age effect, where disability increases with age, as expected. Somewhat surprisingly, the age effect is almost absent is the British sample, especially in the male group. Note that the age effect is generally stronger for females. The largest sex difference occurs in the Norwegian sample, where women experience more dressing disabilities. The country that deviates most from the others is The Netherlands, for which disability is clearly below the country averages. The interesting observation in this analysis is that the country effect is similar in magnitude to the other two, and thus might not be unreasonable in size. On the other hand, it would certainly be comforting if we had an explanation why the 55–59 year old Austrians males have more trouble getting dressed than Dutch males that are 30 years older.

CONCLUSIONS

Incomparability of data is a key problem in international comparisons. Response conversion is a new method for enhancing comparability of seemingly incomparable data. RC consists of two main steps: construction of a conversion key and application of the key to data. The method systematically exploits any overlap between existing data sources. The statistical methodology is built on well-established psychometric theory, and has been applied before on linked health data.[3,11] The most important asset of the methodology is that it can work with existing data. This aids in setting up a health monitoring system without the need to drastically change established ways of working in the participating countries.

Conversion can only be done if a key is available. A key can only be made if the questions can be linked by data, i.e. by bridge studies and bridge items. The quality of the conversion key varies with the quality of the links (e.g. with sample size, plausibility of equivalent translation, use of an internal versus external frame of reference). Given appropriate resources, one could collect new data with the explicit goal of key construction. Construction of the linkage matrix is then more controllable and less sensitive to historical data. Such a study need not be very large or costly, and can lead to a more compact and workable linkage matrices for the health parameter of interest.

Let us be clear that we do not advocate response conversion as a panacea for all harmonization. It is of course always better to get comparable data in the first place, but our experience is that this is not always possible. There is increasing evidence that 'getting comparable data' is not as easy as it sounds, even when the questions are the same.[2] Cross-cultural validity now goes beyond trans-

Table 3 Five national studies that measure dressing disabilities

Country	Name	Sample	Sample size	Year	Organization
Austria	Microcensus Survey on Disabilities	All, includes institutions	60000	1995	Austrian Central statistical office
Italy	Italian survey on health conditions and recourse to health services	6+ (in households)	75000	1994	ISTAT
Norway	Health interview survey (Helseundersokelsen)	All in households	14000	1995	Statistik sentralbyråå
Netherlands	Health Interview Survey	All, excludes institutions	10000	1995	Statistics Netherlands
United Kingdom	Health Survey for England	>2 years, includes 2000 institutions	20000	1995	SCPR + UCL + Dept of Health

lation and cultural equivalence, and requires scale equivalence through more formal methods like the one discussed in this paper. Getting comparable data can be relatively easy for new topics in new samples. It is already more difficult to attain comparability on old topics in new samples, as harmonization might compromise historic comparability. But it is impossible to compare existing data in the form in which they were collected. It would be wasteful to discard existing data with the argument that they are not comparable. Likewise, it would be inefficient if new community-wide surveys duplicate data collection efforts of the individual member states. Our method is particularly useful to deal with these kinds of inefficiencies.

Further work within the context of the Health Monitoring Programme (HMP) is currently being done. This consists of the evaluation of the suitability of response conversion for projects and data within the HMP, the construction of new conversion keys, and the development of an interactive web site for actual conversion to a common scale. A couple of technical challenges require further attention. We developed methods to check the appropriateness of equivalence assumptions, but more work is needed to investigate these methods in the context of linked data. Also, we need a more complete methodology to verify the assumption of unidimensionality under linked data. In addition, further validation of the method used to estimate ability estimation would be useful. We are currently working on conversion keys for physical activity from data of the EUPASS project[12] and personal care.

The ECHI-indicator list[13] contains a quality indicator, coded as categories *a* through *d*. Response conversion is most useful for indicator of type *b* and *c* indicators, where

Table 4 Response frequencies of samples in five national studies, split according to country, age and sex. The last column given the mean of the disability distribution of the category in units of the common scale

	Male							Female							Mean
	Age (years)							Age (years)							
Country	55–59	60–64	65–69	70–74	75–79	80–84	85–89	55–59	60–64	65–69	70–74	75–79	80–84	85–89	dis-ability
Austria															
Washing and dressing oneself															
0 Independent	282	212	253	131	149	74	50	254	264	277	277	258	171	136	–2.69
1 Only with help	7	5	10	14	6	16	10	3	6	7	13	16	38	43	–1.40
2 Not possible	1	2	0	0	1	3	3	1	1	0	4	1	5	7	1.29
Italy															
Can you dress and undress yourself on your own?															
0 Without difficulty	1814	1573	1398	1017	418	350	93	1858	1623	1607	1270	559	478	165	–2.73
1 Some difficulty	34	36	55	79	55	97	45	49	57	119	143	118	204	100	–1.64
2 Only with help	10	28	31	36	28	43	28	7	20	21	26	28	66	44	0.51
Netherlands															
Can you dress and undress?															
0 Without difficulty	202	180	151	138	85	45	6	215	175	195	169	109	78	31	–2.94
1 With some difficulty	12	9	9	13	11	6	2	10	15	12	16	22	16	15	–1.87
2 With much difficulty	1	1	0	1	0	1	0	1	1	1	3	3	2	1	0.69
3 Only with help	0	1	1	1	0	1	0	0	1	1	3	2	1	3	3.18
Norway															
Can you dress and undress yourself?															
0 With no difficulty	29	36	60	50	39	170	51	35	48	56	78	59	200	93	–2.73
1 With some difficulty	6	1	8	8	10	27	12	8	6	13	20	13	53	29	–1.64
2 Only with the help of others	2	1	0	2	6	11	4	1	3	7	4	9	30	25	0.51
United Kingdom															
Can you dress and undress yourself on your own?															
0 Without difficulty	34	48	45	41	23	18	10	36	48	50	67	53	34	25	–2.73
1 Some difficulty	10	29	16	23	12	5	5	16	10	12	13	14	19	10	–1.64
2 Only with help	10	3	10	4	3	1	3	5	5	11	9	6	10	11	0.51

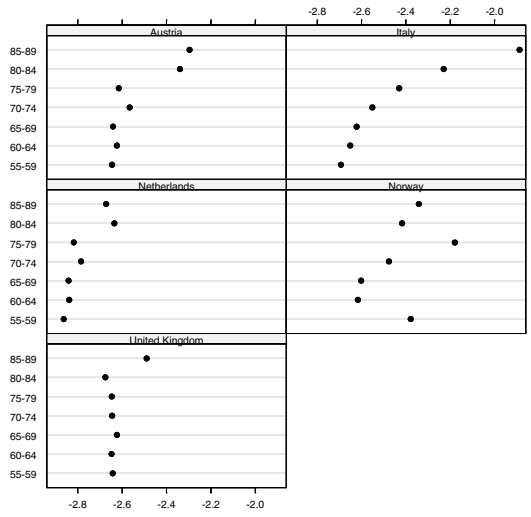

Figure 2 Mean dressing disability for five European countries, expressed on a common scale, by age (males)

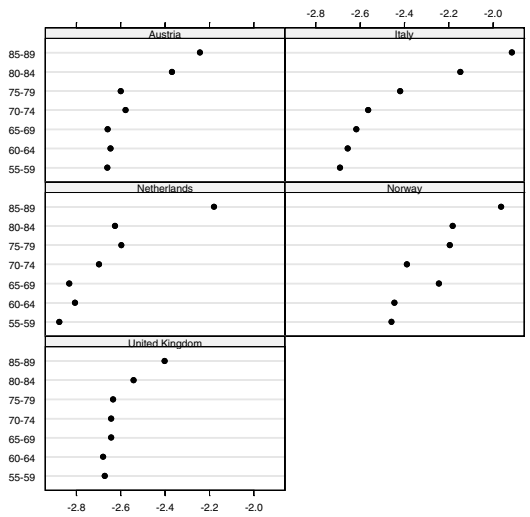

Figure 3 Mean dressing disability for five European countries, expressed on a common scale, by age (females)

comparability is a major issue. Indicators are often defined as percentages. The RC framework applies to percentages if it is possible to conceptualize the percentage as the size of the group falling above a certain cut point on a continuum. Linkages between different definitions can be established if the same sample can be classified according to multiple cut points. Breastfeeding is an example that comes to mind. One could conceptualize 'type of infant nutrition' as a continuum ranging from exclusively breast-fed on one extreme to exclusively non-breastfed on the other extreme, with many shades of grey in between. Given this continuum, different indicators result when a cut point is placed at different locations. Given appropriately linked data and a fitting statistical model, it is possible to estimate the prevalence under each such indicator.

The goal of the Health Monitoring Programme was to develop a viable health information system on a European scale. Much emphasis has been put on the definition of indicators and on the development of co-ordinated networks. Further development and implementation of the information system will require appropriate strategies for dealing with comparability problems of the type discussed here.

This study has been financially supported by the Health Monitoring Programme of the European Commission under agreement number SI2.131854 (99CVF3–510).
The views expressed in this paper are the authors' and are not necessarily held by the European Commission.

REFERENCES

1 European Commission. Work programme of community action on health monitoring 1998-1999. Luxembourg: European Commission, 1998.
2 Harkness JA, van de Vijver FJR, Mohler PP, editors. Cross-cultural survey methods. New York: Wiley, 2002.
3 Van Buuren S, Hopman-Rock M. Revision of the ICIDH Severity of Disabilities Scale by data linking and item response theory. Stat Med 2001;20:1061-76.
4 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.
5 Hopman-Rock M, Odding E, Hofman A, Kraaimaat FW, Bijlsma JWJ. Physical and psychodisability in elderly subjects in relation to pain in the hip or knee. J Rheumatol 1996;23:1037-44.
6 Suurmeijer TPBM, Doeglas DM, Moum T, et al. The Groningen Activity Restriction Scale for measuring disability: its utility in international comparisons. Am J Public Health 1994;84:1270-3.
7 Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149-74.
8 Van Buuren S, Eyres S, Tennant A, Hopman-Rock M. Response conversion: a new technology for comparing existing health information. Leiden: TNO Prevention and Health; 2001. Report No.: 2001.097.
9 RUMM Laboratories. RUMM 2010 Rasch Unidimensional Measurement Models. Perth: RUMM Laboratories, 2001.
10 World Health Organisation (WHO). International Classification of Impairments, Disabilities and Handicaps. Geneva: World Health Organisation, 1980. Reprinted 1993.
11 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. Disabil Rehabil 2000;22:363-71.
12 Rütten A. European Physical Activity Surveillance System (EUPASS). Luxembourg: European Commission; 2001: Final report.
13 ECHI Group. Design for a set of European Community Health Indicators. Luxembourg: European Commission; 2001: Final report.