Int J Public Health 54 (2009) 55–60
1661-8556/09/020055-6
DOI 10.1007/s00038-009-7104-y
© Birkhäuser Verlag, Basel, 2009

Original article

The reliability of the Minimum European Health Module

Bianca Cox[1], Herman van Oyen[1], Emmanuelle Cambois[2], Carol Jagger[3], Sophie le Roy[4], Jean-Marie Robine[4], Isabelle Romieu[4]

[1] *Unit of Epidemiology, Scientific Institute of Public Health, Brussels, Belgium*
[2] *Institut National d'Etudes Démographiques, Paris, France*
[3] *University of Leicester, Leicester, England*
[4] *Institut National de la Santé et de la Recherche Médicale, Montpellier, France*

Submitted: 10 December 2007; revised: 13 May 2008; accepted: 02 October 2008

Published online first: 30 January 2009

Abstract

Objectives: The Minimum European Health Module (MEHM) consists of 3 global questions concerning 3 health domains: self-perceived health, chronic conditions and long-term activity limitation. The objective of this paper is to evaluate the reliability of the MEHM.

Methods: Participants of the Belgian Food Consumption Survey were interviewed twice: 170 individuals were selected for the MEHM reliability evaluation. For each of the 3 questions Pearson and Kappa coefficients were estimated. Analyses were stratified by gender, age, education, language and time between the interviews.

Results: The Pearson correlations are between 0.73 and 0.81. The Kappa estimates are good or excellent: 0.74 (self-perceived health), 0.77 (chronic conditions) and 0.68 (activity limitation). Also stratified analyses indicated in general an acceptable reliability.

Conclusion: The MEHM has an acceptable reliability.

Keywords: Reliability – Survey instrument – Self-perceived health – Chronic conditions – Activity limitation.

The Euro-REVES 2 project, 'Setting up of a coherent set of health expectancies for the European Union', was begun in 1998 under the European Health Monitoring Programme with the aim of selecting a concise set of instruments to simultaneously monitor mortality and the different facets of health.[1] An important spin-off of this project was the proposal for the use of the Minimum European Health Module (MEHM) in health-related surveys in European countries. The European Statistical Agency, Eurostat, adopted the MEHM as an essential part of the European Health Survey System (EHSS). The EHSS consists among other things of a European Core Health Interview Survey (ECHIS) which has two components:

- The annual MEHM which should be included in all health and health-related surveys such as the social survey on Statistics on Income and Living Conditions (SILC). The inclusion of the MEHM in the SILC is carried out from 2004/2005 onwards on a routine basis;
- A European Health Interview Survey (EHIS) to be held every five years and including the following modules: a module on health status (including the MEHM), a module on health care, a module on health determinants, and a module on background variables. The EHIS is planned to be undertaken in all EU Member States in the period 2008/2009 and is part of the new Regulation of the European Parliament and Council on public health and health and safety at work statistics, which is currently in the process of being adopted.

The objective of the MEHM is to obtain, with a short instrument (3 global questions), information on 3 health domains: self-perceived health, chronic (long-standing) conditions and long-term activity limitation. Appendix 1 shows the 3 questions in English and the corresponding French and Dutch version as used in the Belgian national health survey and SILC.

The self-perceived health (SPH) item is a global measure that encompasses different dimensions of health, i.e. physical, social and emotional function and biomedical signs and symptoms.[2] Self-perceived health appears to be an effective summary of health and has been shown to be a strong predictor of future functional limitations, cognitive impairment and mortality.[3,4,5] Its holistic approach and its generalisation of the

concept of health, as well as the comprehensiveness of the population responses, allows it to be used when comparing different populations.[2] The self-perceived health item is based on the existing recommendations of the WHO.[6]

The next global item of the MEHM is a measure of the prevalence of chronic conditions, developed by the Italian National Institute of Statistics for the WHO EuroHIS project titled 'Developing common methods and instruments for health interview surveys in Europe'.[7] Longstanding diseases or conditions affect the health-related quality of life and are one of the major causes of utilisation of the health services.

The third item, the global activity limitation indicator (GALI), is defined as an item that is able to identify subjects, in both general and/or specific populations, who perceive themselves as having long-standing, health-related restrictions or limitations in their usual activities.[1] In a period of ageing populations, the policy relevance of this item is to have easily obtainable information on the perception of disability that could result in a need for care. Moreover, similar to the concept of perceived health, there was a search for developing a global single-question item to measure these activity restrictions, independently of any specific type of activity life situations, health problems, age groups, gender or other subgroups.[8] In the French health survey run in 2002–2003, the global activity limitation indicator was shown to be highly predictive of functional problems reported later in the questionnaire; it captures more systematically severe levels of disability as it is most sensitive to personal care activity restrictions and less sensitive to functional limitations that are not associated with activity restrictions.[9] Noteworthy, the healthy life years (HLY) item, which has been elected as the new structural indicator of the European community (http://ec.europa.eu/health/ph_information/indicators/lifeyears_en.htm), will be calculated yearly based on the global activity limitation indicator from SILC data.[10]

In spite of the prominent position of the MEHM within the European Health Survey System there are only a few published and unpublished papers[11,12] describing an assessment of the MEHM.

The objective of this paper is to evaluate the reliability of the MEHM by means of a test-retest procedure.[13]

Methods

Data

In the 2004 Belgian Food Consumption Survey information on food intake was collected for a total of 3200 individuals, using two 24h-recalls in combination with a self completion food frequency questionnaire.[14] In both 24h-recall interviews, the types and quantities of all foods and beverages consumed over the preceding full day had to be reproduced by the respondent. The second interview was carried out within 11 to 55 days after the first interview. The methodology provided the opportunity to select a subsample of 170 individuals and to administer the MEHM as part of the self completion questionnaire. The subsample was selected to ensure an equal representation of persons according to factors which might influence the reliability: males (n = 82) versus females (n = 88), age 15 to 64 (n = 103) versus age 65 and over (n = 67), technical secondary education or less (n = 99) versus general secondary or higher education (n = 71; Belgium has 4 types of secondary education. General education is very broad and theoretical, preparing for further education, while technical, art, and vocational education are more technical and/or practical, preparing for the job market or for further education, depending on the direction), Dutch (n = 88) versus French questionnaire (n = 82; Dutch (60 %), French (40 %) and German (< 1 %) are the three official languages in Belgium), and time span between the two interviews 20 days or less (n = 93) versus more than 20 days (n = 77). The age cut-off was chosen arbitrarily at 65 years in order to distinguish the elderly from the other respondents, while the subsample was selected to have the same age composition as the total survey sample (60.1 % of the respondents younger than 65 years). The time span cut-off at 20 days corresponds to the median time between the two interviews in both the total survey sample and the selected subsample (interquartile range: 17–26 days).

The pairwise associations between the dichotomous variables for gender, age, education, language and time span between the interviews are not significant ($P > 0.05$), with the absolute values of the phi coefficients ranging from 0.01 to 0.12, except for the significant association between education and language (phi = –0.17, $P = 0.025$), indicating that the Dutch language is associated with a higher education.

Analyses

The association between test and retest of the 3 health items of the MEHM was assessed by Pearson correlations and Kappa coefficients. For reliability testing, a correlation value of 0.7 to 0.8 is considered to be sufficient.[13] Correlations, however, are only exploratory and not sufficient for reliability testing, because they measure the strength of a relation between two variables, not the agreement between them.[13,15] Kappa represents agreement corrected for chance and is typically used to measure agreement between raters.[13] Landis and Koch suggest that a Kappa less than 0.40 is poor, from 0.40 to 0.74 is good and equal or greater than 0.75 is excellent.[16]

For the health items with more than two response categories, a weighted Kappa was used to account for the greater

© Birkhäuser Verlag, Basel, 2009

Table 1. Minimal European Health Module at interview 1 and interview 2.

		Interview 2					
Interview 1	***Self-perceived health: How is your health in general?***						
		Very Good	Good	Fair	Bad	Very bad	Total
	Very good	38	11	0	0	0	49
	Good	7	63	6	1	0	77
	Fair	1	6	29	1	0	37
	Bad	0	0	2	3	0	5
	Very bad	0	0	1	0	1	2
	Total	46	80	38	5	1	170
	Chronic conditions: Do you suffer from any long-standing illness or condition (health problem)?						
		Yes	No	Total			
	Yes	40	12	52			
	No	4	114	118			
	Total	44	126	170			
	Global activity limitation indicator: For the past 6 months or more have you been limited in activities people usually do because of health problem?						
		Yes, severely	Yes, moderately	No	Total		
	Yes, severely	5	1	1	7		
	Yes, moderately	2	17	10	29		
	No	0	7	127	134		
	Total	7	25	138	170		

disagreement between categories that are further apart than those that are close together on an ordinal scale. Relatively conservative linear weights were calculated as $w_i = 1-[i/(c-1)]$, where w_i is the weight given for a distance of i categories between the two measurements of a health item with c categories. For the five-category item self-perceived health, the weights applied were therefore 1.00, 0.75, 0.50, 0.25, or 0.00 when there was a difference of 0, 1, 2, 3, or 4 categories between the two measures, respectively. For the global activity limitation indicator with 3 categories, the weights were 1.00, 0.50 and 0.00. Standard errors and confidence intervals for Kappa coefficients were calculated using bootstrap resampling.[17,18]

In addition to values for the total population, results are also presented by gender, age group (15–64 versus 65+), education (technical secondary or less versus general secondary or higher), language (Dutch versus French) and the time span between the two interviews (20 days or less versus more than 20 days). The statistical significance of the difference between two independent kappa coefficients was assessed by means of a z-test, using the standard errors estimated by bootstrap resampling.[19] Results were considered to be statistically significant at $P < 0.05$ (two-tailed). All statistical analyses were computed using Stata software (version 9) (Statacorp, College station, TX).

Results

The raw data from the two interviews are presented in Tab. 1. The Pearson correlation coefficients are 0.81, 0.77 and 0.73, while the Kappa estimates are 0.74, 0.77 and 0.68 for self-perceived health, chronic conditions and the global activity limitation indicator respectively (Tab. 2). According to the definitions given above, these values are good to excellent. The Kappa coefficients are consistently lower than the Pearson correlations, because Kappa is a more conservative estimate of reliability, as it takes into account the agreement occurring by chance.

For the stratified analyses the correlation coefficients are at least 0.70 in all strata, except for the global activity limitation indicator. For this item three strata show a value less than 0.70, but still above 0.5: females, age 15 to 64, time less than or equal than 20 days. Most of the Kappa coefficients are above 0.60 (indeed many are above 0.75), except for the global activity limitation indicator within 2 strata: females (Kappa = 0.54) and age 15 to 64 (Kappa = 0.53).

Comparing the Kappa values between stratification levels, significant differences have only been found between Dutch and French language for self-perceived health (difference in Kappa = 0.21, P = 0.008) and between males and females for the global activity limitation indicator (difference in Kappa = 0.28, P = 0.035).

© Birkhäuser Verlag, Basel, 2009

Table 2. Pearson correlations and Kappa coefficients (with 95 % confidence interval: LL and UL) for self-perceived health, chronic conditions and the global activity limitation indicator.

		Pearson correlation	Kappa Coefficient				
			Obs[a]	Exp[b]	Kappa	LL	UL
Self-perceived health							
Gender	Male	0.89	0.96	0.78	0.82	0.71	0.91
	Female	0.74	0.93	0.78	0.67	0.53	0.79
Age	15-64	0.81	0.94	0.76	0.73	0.62	0.85
	65+	0.77	0.93	0.76	0.70	0.56	0.84
Education	Technical secondary or less	0.80	0.92	0.71	0.72	0.63	0.82
	General secondary or higher	0.81	0.95	0.79	0.75	0.62	0.86
Language*	Dutch	0.90	0.96	0.73	0.85	0.76	0.93
	French	0.74	0.91	0.77	0.64	0.50	0.75
Time span	<= 20 days	0.79	0.92	0.74	0.69	0.57	0.80
	> 20 days	0.83	0.95	0.76	0.79	0.66	0.89
Total		0.81	0.94	0.78	0.74	0.66	0.82
Chronic conditions							
Gender	Male	0.86	0.94	0.60	0.85	0.71	0.97
	Female	0.70	0.88	0.59	0.70	0.51	0.86
Age	15-64	0.80	0.94	0.72	0.79	0.60	0.93
	65+	0.71	0.85	0.50	0.70	0.50	0.85
Education	Technical secondary or less	0.81	0.92	0.57	0.81	0.67	0.93
	General secondary or higher	0.71	0.89	0.63	0.69	0.45	0.86
Language	Dutch	0.76	0.91	0.63	0.75	0.57	0.89
	French	0.78	0.90	0.56	0.78	0.63	0.92
Time span	<= 20 days	0.73	0.89	0.62	0.72	0.55	0.88
	> 20 days	0.82	0.92	0.57	0.82	0.65	0.94
Total		0.77	0.91	0.59	0.77	0.66	0.87
Global activity limitation indicator							
Gender*	Male	0.87	0.96	0.80	0.82	0.63	0.93
	Female	0.61	0.91	0.80	0.54	0.30	0.72
Age	15-64	0.57	0.96	0.91	0.53	0.24	0.74
	65+	0.75	0.90	0.68	0.70	0.51	0.86
Education	Technical secondary or less	0.71	0.92	0.76	0.67	0.48	0.81
	General secondary or higher	0.76	0.96	0.87	0.67	0.24	0.89
Language	Dutch	0.71	0.94	0.81	0.67	0.41	0.81
	French	0.76	0.93	0.79	0.68	0.49	0.85
Time span	<= 20 days	0.67	0.92	0.80	0.62	0.41	0.79
	> 20 days	0.82	0.95	0.80	0.75	0.54	0.90
Total		0.73	0.94	0.80	0.68	0.54	0.80

[a] Obs = percentage agreement observed (weighted)
[b] Exp = percentage agreement by chance (weighted)
* Significant difference in kappa between stratification levels ($P < 0.05$)

Discussion

Both the Pearson and Kappa coefficients indicate an acceptable reliability on the 3 global health items of the MEHM. According to the Pearson correlations, the highest reliability seems to be obtained for the item on subjective health, whereas the highest Kappa is observed for the chronic conditions item. The global activity limitation indicator has the lowest values for both types of measures, but the values are still high enough to conclude good agreement. A possible explanation for this finding is that the response categories for the global activity limitation indicator are less precise than for the other two health items.

For all three health items, the agreement is higher for males than for females, although the difference in Kappa values is only significant for the global activity limitation indicator.

© Birkhäuser Verlag, Basel, 2009

Analyses based on the French data showed that women are more inclined to report activity limitation than men when controlling for age, socioeconomic variables and the other disability measures.[9] This might indicate that women have the propensity to also include mild levels of activity limitation, while men only report more severe levels. This could result in a greater variability among women, because some will report such mild levels and some will not. For the self-perceived health item, subjects questioned in Dutch show a significantly higher agreement than French speaking people, while the opposite can be observed for chronic conditions and the global activity limitation indicator (not significant). The reliability coefficients are consistently higher for a time span of more than 20 days between interviews compared to 20 days or less. The differences in Kappa values however are not significant. This finding may be unexpected. Although we do not have an obvious interpretation, one possible reason could be that when the second interview is close to the first one, the interviewee may be less spontaneous which may result in a different way of replying.

This small study suggests that the MEHM has an acceptable reliability. Given the importance that the MEHM has within the European Health Survey System and the important cultural and language differences within Europe, it is important that these and other indicators of the quality of the MEHM are documented within other populations.

Acknowledgements

This study was supported by the European Union (DG SANCO) as part of the EHEMU project 2004–2007 (EU Grant agreement n° 2003116).

Résumé

Objectifs: Le Mini-module santé européen (MEHM) comprend trois questions couvrant trois dimensions de la santé: santé perçue, morbidité chronique et limitations d'activité de long terme. Cette étude vise à tester la fiabilité du MEHM.

Méthodes: L'enquête de consommation alimentaire belge comporte deux passages: l'étude de fiabilité porte sur un échantillon de 170 personnes. Des coefficients de Pearson et Kappa sont estimés pour chaque question, en stratifiant selon le sexe, l'âge, le niveau d'étude, la langue et la durée entre les deux passages.

Résultats: Les coefficients de Pearson vont de 0,73 à 0,81. Les coefficients de Kappa sont bons ou excellents: 0,74 (santé perçue), 0,77 (morbidité), 0,68 (limitations d'activité). Les analyses stratifiées confirment la fiabilité du MEHM.

Conclusion: Le MEHM présente une fiabilité acceptable.

References

1. Robine JM, Jagger C. Creating a coherent set of indicators to monitor health across Europe: the Euro-REVES 2 project. Eur J Public Health 2003; 13(3 Suppl):6–14.

2. Robine JM, Jagger C, Romieu I. Selection of a Coherent Set of Health Indicators for the European Union, Phase II. Montpellier: Euro-Reves, 2002.

3. Idler EL, Russell LB, Davis D. Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey. Am J Epidemiol 2000; 152(9):874–883.

4. De Salvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. A meta-analysis. J Gen Intern Med 2006; 21:267–275.

5. Bond J, Dickinson HO, Matthews F, Jagger C, Brayne C. Self-rated health status as a predictor of death, functional and cognitive impairment: a longitudinal cohort study. Eur J Ageing 2006; 3:193–206.

6. de Bruin A, Picavet HS, Nossikov A. Health interview surveys. Towards international harmonization of methods and instruments. WHO Reg Publ Eur Ser 1996; 58:1–161.

7. Burata V, Frova L, Gargiulo L, Gianicolo E, Prati S, Quattrociocchi L. Development of a common instrument for chronic physical conditions. In: Nossikov A, Gudex C, eds. Eurohis: Developing Common Instruments for Health Surveys. Amsterdam: IOS Press, 2003:21–34.

8. Verbrugge LM. A global disability indicator. J Aging Stud 1997; 11(4):337–362.

9. Cambois E, Robine JM, Mormiche P. Une forte baisse de l'incapacité en France dans les années 1990? Discussion autour des questions de l'enquête santé. Population 2007; 62(2):361–386.

10. Oortwijn W, Mathijssen J, Lankhuizen M, Cave J. Evaluating the uptake of the Healthy Life Years Indicator: Final report. Rand Europe, 2007.

11. Van Oyen H, Van der Heyden J, Perenboom R, Jagger C. Monitoring population disability: evaluation of a new Global Activity Limitation Indicator (GALI). Social and Preventive Medicine 2006; 51(3):153–161.

12. Dewar A, Woodgate-Jones G, Wilmot A, Betts P, Hand C. Evaluation of Modules for the European Health Interview Survey (EHIS). Workpackage 4: Formal Qualitative Field Testing. London: Office for National Statistics, 2005

13. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006; 119:166.7–16.

14. De Vriese S, Debacker G, de Henauw S, Huybrechts I, Kornitzer M, Leveque A, et al. The Belgian food consumption survey: aims, design and methods. Arch Public Health 2005; 63:1–16.

© Birkhäuser Verlag, Basel, 2009

15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1996; 1:307–310.

16. Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.

17. Lee J, Fung KP. Confidence interval of the kappa coefficient by bootstrap resampling [letter]. Psychiat Res 1993; 49(1):97–98.

18. Reichenheim ME. Confidence intervals for the kappa statistic. Stata Journal 2004; 4(4):421–428.

19. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC, 2003.

Address for correspondence

Bianca Cox
Unit of Epidemiology
Scientific Institute of Public Health
J Wytsmanstraat 14
1050 Brussels
Belgium
Tel.: +32 2 642 5752
Fax: +32 2 642 5410
E-mail: b.cox@iph.fgov.be